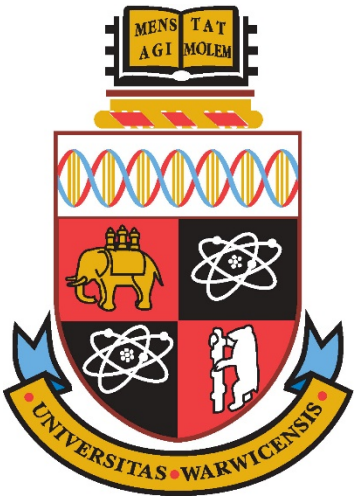
WARWICK



Early Assessment of Medical Technologies – addressing uncertainty to inform good decisions

by

Carlo Baldassarre Federici

Thesis

Submitted at the University of Warwick

For the degree of

Doctor of Philosophy

School of Engineering

December 2021

To Yelda,
for glittering up my life,
and everything else around
(literally)

Table of Contents

List of figures

List of tables

Acknowledgments

My first thanks go to my supervisor, Prof. Leandro Pecchia, not only for the inspiring talks (with or without beers), continuous support and friendly advice, but also for the various misunderstandings, losses in translations, and (gentlemen*)* disagreements. The economists-to-engineers dictionary we have created should rightly be considered as one of the highest achievements of this PhD.

I would like to express my sincere gratitude to Prof. Aleksandra Torbica, for many, many reasons, but above all for her human and professional qualities and the example she sets every day. This journey without her discrete *safeguard* would have been far more difficult than it actually was.

I also owe the deepest gratitude for the encouragement, support, and friendship to Alessia Melegaro, Oriana Ciani, Giuditta Callea, Rosanna Tarricone and Mike Drummond, the other colleagues from the COMED projects, and all the fellows at the University of Warwick. Working together has been and continues to be a source of continuous inspiration.

A special thought goes out to my mum because this PhD is just the last step in a journey of redemption that began with the pain for loosing her and the desire to make her proud. So, thank you again mum, also for this.

Oh, and of course I thank all the rest of my family and friends for their unconditional love and comprehension, despite probably not having a clue about what I was studying and why I insist on doing it at this *venerable* age.

Declarations

The work contained in this thesis is based on the author's own work. This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy, and has not been submitted in any previous application for any degree.

Most of the work presented here has been produced as an output for the Horizon2020 research project "COMED: Pushing the boundaries of Cost and Outcome analysis of Medical Technologies" (Grant Agreement #779306). As such, some of the sections in this work are the results of a joint work with other researchers participating in the project. Their main contributions are listed here:

1. Section 2.4. Prof. Mike Drummond contributed to the literature review and discussion on value of information analysis presented in this section;
2. Section 2.5. The discussion on the regulatory approval processes for medical devices and pharmaceuticals is based on the joint work with Prof. Rudolf Blankart;
3. Section 2.6. Dr. Vivian Reckers-Droog contributed to the systematic literature review presented in this section, including the collection and analysis of relevant literature and the drafting of the manuscript on which this section is based;
4. Section. 4.2. Dr. Rowan Iskandar was the main responsible for the application of the Probability Bound analysis to the case study developed by the author, as well as for the writing of the manuscript on which the sections that refer to this analysis are based;
5. Chapter 5. The author was supported by the COMED researchers in the interviews to the European policy maker, namely: Dr. Vivian Reckers-Droog, Dr. Oriana Ciani, Mr. Florian Dams, Dr- Bogdan

Grigore, Dr. Zoltán Kaló, Dr. Sándor Kovács, and Dr. Kosta Shatrov.

More generally, the work presented here has been subject to extensive reviews and comments by the leaders of two working packages in the project, namely Prof Mike Drummond, Prof. Werner Brower and Dr Rudolf Blankart.

Beyond the COMED project, for the case study presented in section 4.1 Dr. Eva Pirc and Prof. Damijan Miklavčič collected the data used in the model developed by the author and contributed to the drafting of the manuscript on which this section is based.

Parts of this thesis are based on journal articles that have already been published (1–10) or are currently submitted to scientific journals (11–13).

List of abbreviations

AHA	Active Health Aging
CBA	Cost-Benefit Analysis
BiVAD	Biventricular Assist Device
CDF	Cumulative Distribution Function
CE	Conformitè Européenne
CEA	Cost-Effectiveness Analysis
CEAC	Cost-Effectiveness Acceptability Curves
CED	Coverage With Evidence Development
CNEDiMTS	French Medical Device And Health Technology Evaluation Committee
CRGs	Clinical Reference Groups (England)
CrI	Bayesian Credible Intervals
CTIIMH	Belgium Implant And Invasive Medical Device Reimbursement Committee
DAM	Decision Analytical Model
DF-HTA	Developer Focused Health Technology Assessment
DNA	Deoxyribonucleic Acid
DOC	Declaration Of Conformity
DRG	Diagnosis Related Groups
DSA	Deterministic Sensitivity Analysis
EBM	Evidence Based Medicine
ECT	Electrochemotherapy
EIP	European Innovation Programme
EMA	European Medicine Agency
EVPI	Expected Value Of Perfect Information

EVPPI	Expected Value Of Perfect Parameter Information
EVSI	Expected Value Of Sampling Information
FDA	Food And Drug Administration
FI	Forfait Innovation Scheme
FOPH	Swiss Federal Office Of Public Health
G-BA	German Federal Joint Committee
GDP	Gross Domestic Product
GRP	Good Research Practice
HF	Heart Failure
HTA	Health Technology Assessment
ICER	Incremental Cost-Effectiveness Ratio
InEK	German Institute For The Hospital Remuneration System
INMB	Incremental Net-Monetary Benefit
IQWiG	German Institute For The Hospital Remuneration System
IRA	Irreversible Electroporation
ISPOR	International Society For Pharmacoeconomics And Outcomes Research
IU	International Units
LBF	Lower Bounding Function
LPPR	French Positive List Of Reimbursable Products And Services
LVAD	Left Ventricular Assist Device
MBI	Maximum Budget Impact
MCD	Minimum Clinical Difference
MCMC	Markov Chain Monte Carlo
MD	Medical Device
MDR	Medical Device Regulation

NHB	Net Health Benefit
NICE	National Institute For Health And Care Excellence
NICE	UK National Institute For Health And Care Excellence
NMB	Net Monetary Benefit
NPV	Net Present Value
OIR	Only In Research
OWR	Only With Research
PBA	Probabilistic Boundary Analysis
PBRSA	Performance-Based Risk Sharing Arrangements
PRS	Post Registration Studies
PSA	Probabilistic Sensitivity Analysis
QALY	Quality Adjusted Life Years
RCT	Randomized Controlled Trials
RIZIV	Belgian Medicines Verification Organisation
ROA	Real Options Analysis
ROI	Return On Investment
RV	Right Ventricular
SEE	Structured Expert Elicitation
SoC	Standard Of Care
TAH	Total Artificial Heart
UBF	Upper Bounding Function
VBP	Value Based Pricing
VOI	Value Of Information
WTP	Willingness To Pay
ZIN	Netherlands National Health Care Institute

ZonMW	Netherlands Organisation For Health Research And Development, Implant And Invasive Medical Device Reimbursement Committee

Abstract

Innovation in medical technologies has the potential to strongly improve the quantity and quality of life of sick people suffering from a broad range of conditions. However, the true value of medical technologies at the time of taking adoption and reimbursement decisions is often highly uncertain due to limited clinical and economic evidence.

The overarching objective of this thesis was to explore methods and policy tools to characterize, quantify and address the uncertainty over the effectiveness and cost-effectiveness of medical technologies early on in their life cycle. The problem of uncertainty was addressed both under a methodological aspect and a policy aspect.

Two early cost-effectiveness models were built and used as case-studies to apply a series of methodological approaches. In a model on electrochemotherapy in patients with skin melanoma a probabilistic headroom approach was applied to estimate the maximum achievable price for a technology given the existing early evidence on its effectiveness. A model on a novel total artificial heart was then used to explore the use of Probability Bound Analysis (PBA), a new method which was recently proposed as an alternative way to characterize uncertainty when the lack of evidence hampers the use of more established methods such as probabilistic sensitivity analysis. The same case study was used to explore the potential misalignments which may exist about the most appropriate level of evidence between manufacturers and payers. By defining different utility functions and decisions rules for payers and technology developers respectively, value of information analysis was used to quantify the value of further evidence from both perspectives and the potential gains of reducing it through further research. From a policy perspective, this thesis explored the use of coverage with evidence development schemes as a policy tool for medical devices in Europe. Particularly, through a review of the available evidence and the collection of interviews to European decision-makers the main characteristics and challenges of such schemes were identified.

Uncertainty is an unavoidable fact of life and particularly in healthcare decision making. The recently enforced medical device regulation (MDR) and the approved proposal for a regulation on cooperation in HTA among member states seem to go in the direction of a lifecycle approach for the evaluation of healthcare technologies. According to this new regulatory approach, Technologies will likely be assessed moreiteratively (and differently) as new evidence accumulates over their life cycle. Greater interaction between the stakeholders involved, and in particular technology developers and payers is also envisaged.This paradigm shift will require the development of shared methods to assess the uncertainty surrounding a technology. This thesis aimed to contribute to filling this need, so to allow more appropriate, fair and transparent policy decisions in everyone's interest.

Chapter 1
Introduction

Universal health coverage is rightly considered as a milestone achievement for human rights since it allows patients to have equitable access to health services regardless of their ability to pay. Several countries in the world, both high- and low-income, have now introduced forms of universal coverage, either tax-funded or based on some form of social insurance, with evident benefits on equity of access and overall population health. Nevertheless, the universal provision of health services at no or low cost to those who need them must be confronted with the availability of resources raised for this purpose, which are of course limited. In fact, any time healthcare payers (e.g., a government or an insurance fund) considers the inclusion of a new technology in their benefit package, they need to consider questions like: does this technology provide an added clinical benefit compared to other alternatives already available to patients? Will the technology generate additional costs for the healthcare systems? And if so, does the additional health benefits justify these additional costs?

In economics, these questions are strictly related to the concept of the opportunity cost of the budget constraint. In simple terms, the limitedness of available resources requires that any potential way in which resources can be used must be compared to all other competing alternatives, with the aim to maximize the technical efficiency of the budget, given a certain social objective (14).

According to the prevalent evidence-based decision-making paradigm, solid clinical and economic evidence is usually required to answering these questions and therefore to inform both clinical decisions as well as adoption and reimbursement decisions in healthcare. However, especially at the time when these decisions have to be made, i.e., generally at the beginning of a

technology's life cycle, the level of available evidence fall short in providing an unambiguous answer as to what is the best course of action. This uncertainty makes decision-making particularly challenging and requires a thorough analysis of the possible consequences of taking sub-optimal decisions should this uncertainty resolve not as expected. As a matter of fact, the negative consequences of uncertainty may be realised either if a technology is mistakenly given to patients based on false, premature claims about its efficacy or cost-effectiveness; or if access to a new technology which would yield positive health gains is denied because of insufficient available evidence. Therefore, given that it is generally not possible to resolve all existing uncertainty around a decision, decision-makers are often required to express judgments on the appropriateness of the available evidence; or alternatively on whether more evidence should be warranted and whether the value gained by generating it is worth its costs.

Answering these questions may prove difficult, leading to sub-optimal decisions and negative consequences on the health of patients.

Contributing to the efficient and informed decision-making in healthcare is what I am most passionate about and therefore, the overarching objective of this thesis focuses on exploring methods and policy tools to characterize, quantify and address the uncertainty over the effectiveness and cost-effectiveness of medical devices (MDs) early on in their life cycle.

While universally valid for any type of healthcare technology, the methods and policy tools explored here are mainly focused on MDs. MDs are, in fact, an interesting case when addressing uncertainty in healthcare decision making: First, their pace of innovation has increased sharply in recent years, with a growing number of patents filed and approved in different applications (diagnostics, implantable devices, wearables etc.), and a promise for providing huge impact in terms of patients' health and well-being. Second, uncertainty over medical devices at the time of making adoption and reimbursement decisions is usually considerable, mainly because devices have been traditionally subject to a less degree of scrutiny before getting to the market, and because evidence generation through standard high-quality

research (e.g., randomized controlled trials) may be more challenging as opposed to other technologies such as pharmaceuticals.

In this thesis, the issue of decision making under uncertainty will be approached from two different standpoints: 1) a methodological standpoint, related to the quantitative exploration of the consequences of uncertainty on decision-making and the assessment of the value of further evidence; and 2) a policy standpoint, mainly related to the exploration of available policy tools that allow healthcare payers to minimize the risks of taking wrong adoption decisions.

Particularly, by using both qualitative and quantitative approaches the research reported in this thesis has aimed at:

- Exploring new and existing methods to characterize uncertainty over the effectiveness and cost-effectiveness of novel medical technologies, especially in the context of early-stage economic evaluations based on decision-analytical models
- Extending the application of validated methods to estimate the value of further evidence to explore potential misalignment between healthcare payers and technology developers about the optimal perceived level of evidence from each individual perspective
- Explore the actual perceptions and adoption in Europe of existing policy tools, such as performance-based risk-sharing agreement and coverage with evidence development schemes that allow to minimize the risk of uncertainty.

During these years of my PhD course, I submitted and published a number of papers to scientific academic journals on topics related to this thesis (Table 1-1), and contributed to two introductory chapters in the second edition of the Clinical Engineering Handbook (Elsevier, Edited by Prof. Ernesto Iadanza) (Table 1-2).

Table 1-1 Papers submitted and published in scientific journals

Journal articles							
#	**Authors**	**Title**	**PY/ Status**	**Journal**	**Scopus Cite Score**	**Highest perc.**	**Ref.**
1	Reckers-Droog V, Federici C, Brouwer W, Drummond M.	Challenges with coverage with evidence development schemes for medical devices: A systematic review	2020	Health policy and technology	2.7	60%	(1)
2	Pirc E, Federici C, Bošnjak M, Perić B, Reberšek M, Pecchia L, et al.	Early cost-effectiveness analysis of electrochemotherapy as a prospect treatment modality for skin melanoma.	2020	Clinical Therapeutics	5.1	75%	(2)
3	Federici C, Armeni P, Callea G.	A Value-based Revolution in Health Care: Perspectives, Challenges, and Emerging Approaches to Defining and Measuring the Value of Health Care Technologies.	2021	Clinical Therapeutics	5.1	75%	(3)
4	Fornaro G, Federici C, Rognoni C, Ciani O.	Broadening the Concept of Value: A Scoping Review on the Option Value of Medical Technologies.	2021	Value in Health	6.7	95%	(4)
5	Federici C, Reckers-Droog V, Ciani O, Dams F, Grigore B, Kaló Z, et al.	Coverage with evidence development schemes for medical devices in Europe: characteristics and challenges.	2021	European Journal of Health Economics	4.3	93%	(5)
6	Federici C, Pecchia L.	Early health technology assessment using the MAFEIP tool. A case study on a wearable device for fall prediction in elderly patients	2021	Health and Technology	2.5	40%	(6)
7	Federici C, Torbica A.	Expanding the role of early health economic modelling in evaluation of health technologies. Comment on 'Problems and promises of health technologies: the role of early health economic modeling'.	2021	International Journal of health policy and management	5	98%	(7)

Journal articles							
#	**Authors**	**Title**	**PY/ Status**	**Journal**	**Scopus Cite Score**	**Highest perc.**	**Ref.**
8	Blankart CR, Dams F, Penton H, Kaló Z, Zemplényi A, Shatrov K, and Federici C.	Regulatory and HTA Early Dialogues in Medical Devices.	2021	Health Policy	4.3	82%	(8)
9	Iskandar R, Federici C, Berns C, Blankart R.	An approach to quantify parameter uncertainty in early assessment of novel health technologies.	2022	Health economics	4.6	86%	(13)
10	Drummond M, Federici C, Reckers-Droog V, Brouwer W.	Coverage with evidence development schemes for medical devices: A policy guide.	2022	Health economics	4.6	86%	(11)
11	Federici C, Pecchia L.	Exploring the misalignment on the value of further research between payers and manufacturers. A case study on a novel total artificial heart.	2022	Health economics	4.6	86%	(12)

RR, Revise and re-submit

Table 1-2 Chapters published in books

Chapters in Books					
#	**Authors**	**Title**	**PY**	**Book**	**Ref.**
1	Castaldo R, Federici C. and Pecchia L.	Early Stage Healthcare technology assessment	2020	Clinical Engineering Handbook - 2nd Edition. Editor: Ernesto Iadanza	(10)
2	Ciani O., Federici C.	Introduction to Health Economics and HTA	2020	Clinical Engineering Handbook - 2nd Edition. Editor: Ernesto Iadanza	(9)

In addition during these years I participated to national and international conferences presenting my work in panel sessions, oral presentations or poster presentations. The conferences I attended to include:

- The International Society for pharmacoeconomics outcomes research (ISPOR) (Barcellona, Spain, 2018; New Orleans, USA, 2019; Copenhagen, Denmark, 2019; ISPOR Virtual meeting 2021)

- Italian association of health economics (AIES) (Naples, Italy, 2018; Pisa, Italy, 2019)
- Fourth WHO Global Forum on Medical Devices (Visakhapatnam, India, 2019)
- International Conference on Medical and Biological Engineering (CMBEBIH) (Banja Luka, Bosnia and Herzegovina, 2019)
- International Health Economics Association (iHEA) (Basel, Switzerland, 2019)

Chapter 2
Background and review of the literature

This chapter provides a brief background on the types of economic evaluations and their theoretical foundations (§2.1) and introduces decision analytical models as modelling tools that have been largely used for cost-effectiveness analyses of healthcare technologies (§2.2).
Subsequently, the chapter introduces the role, objectives, and main methodological approaches for uncertainty analysis in decision analytical models. In section 2.3, the role of uncertainty analysis is discussed according to two different perspectives: the perspective of healthcare payers, and the perspective of technology developers. Section 2.4 provides a more in-depth introduction of the Value of Information framework, a formal approach that has been proposed to evaluate further investments in evidence generation. Lastly, section 2.5 addresses the issue of decision uncertainty from a policy viewpoint by introducing coverage with evidence development schemes, a policy tool for healthcare payers to address the risk of taking uncertain decisions at the time of market access of newly developed medical devices. A systematic literature review of the main challenges associated with these schemes is finally reported in section 2.6.

2.1. Economic evaluation of health technologies

Economic evaluation is usually conducted during a health technology assessment (HTA). The goal of such evaluations is that of informing whether a specific policy change represents an efficient spending of limited resources, as opposed to other alternative configurations of spending. To provide such information, economic evaluations have two main characteristics (15). First, they require a comparison of two or more alternative options (e.g., alternative

treatments for the same condition in the same patient population). Second, economic evaluations need to consider both the costs and consequences of each intervention so that some judgments of its *value for money* can be made.

There exist different types of economic evaluations which are rooted in as many theoretical paradigms. Broadly speaking, two main types of analyses can be distinguished: cost-benefit analyses (CBA), which have their foundations in welfare economic theory (16), and cost-effectiveness analyses (CEA) which trace back their disciplinary origin in social decision making theory (17), but are also consistent with the constrained optimization approach of operations research and management science. A detailed descriptions of the characteristics of each type of economic evaluation and their theoretical background goes beyond the purpose of this thesis, and good introductory texts are available elsewhere (18–20). Here, only a brief overview of the main purpose behind each type of economic evaluation is provided together with a clarification of the type of analysis which will be considered for the rest of the thesis.

In essence, consistently with their theoretical foundations the two approaches differ in which effects of an intervention should count to inform social choice and how these effects are measured and valued. Indeed, the adoption of healthcare interventions may produce effects that goes beyond just health, including consequences on other public sectors (e.g., education or criminal justice) and the society at large (e.g., individuals out of pocket expenditures, productivity losses, or the time spent by informal caregivers). Adopting a welfarist approach implies that the purpose of health care interventions is that of improving the welfare of society and therefore prescriptions based on such paradigm should consider all effects that are socially valuable, including those beyond health. Accordingly, CBA seeks to value the full range of health and other consequences of a policy change, and it does so by valuing all effects in monetary terms. Because CBA takes such a broad perspective it enables strong prescriptions about what would improve social welfare (and should be done) and therefore it also enables judgments on whether the healthcare budget should be expanded to make

room for a new intervention. While this approach is rooted into welfare economics it implies strong value judgments on what is social optimum and presents several technical challenges. Eventually, the use of CBA encountered resistance in the health care field mainly because of concerns about an individual's cognitive skills in expressing his or her willingness to pay for health-related goods and services, and equity considerations due to the effects of income on willingness-to-pay values (3).

The extra-welfarists paradigms adopt a narrower approach by just considering the perspective of health care decision makers who have been endowed with a budget (e.g., through the democratic process of allocation of funds across different areas of public intervention) to pursue a pre-defined societal objective. According to these perspectives, the societal objective(s) and the budget constraint are exogenously fixed, and the aim of evaluations is that to inform how to maximize the former given the latter (21).

CEA is the form of economic evaluation that has generally been used in health care to apply these principles of resource allocation, adopting as the main objective that of maximizing health in the population. Therefore, in CEA, the costs to be measured in an evaluation are limited to the ones which fall on the healthcare budget; for example, doctors and nurses, capital equipment and buildings, and consumables such as drugs or medical devices. Consequences generally focus on changes in individuals' health, which are usually expressed either in terms of physical units (e.g., reduced incidence of strokes, or life years gained) or in terms of health-related quality of life (HRQL).

Eventually, the output of CEA provides a measure of the additional cost to the healthcare system at which a new health technology will deliver additional health benefits, as compared to the other alternatives identified.

The vast majority of the theoretical and applied literature on CEA has adopted the perspective of public healthcare funders and aimed at informing how to optimally allocate limited public resources. Nonetheless the use of CEA has been proposed also to orient manufacturers' choices about product development strategies, research prioritization and portfolio management.

The use of CEA to inform manufacturers' decisions is usually justified by the fact that early consideration of the potential cost-effectiveness of the technologies under development may help manufacturers to identify technologies that are most likely to succeed and how to best prioritize limited research and development resources (22–27).

Regardless of the perspective used, CEA has been the predominant approach for economic evaluations in health care in the last decades and therefore this approach has been adopted also as the basis for the methodological and applied research conducted in this thesis.

2.2. Decision analytical modelling for economic evaluation

Decision analysis can be defined as a systematic approach to decision making under uncertainty (28).

In health care, the use of decision analysis has been extendedly used to inform clinical decision making as clinical decisions are often hampered by high uncertainty on which is the optimal course of action (29–31). For example, such uncertainty has shown clearly during the recent COVID-19 pandemics about the risk/benefit profile of vaccines in different population sub-groups, or the utility of individual protection devices, in reducing COVID-19 related infections and deaths (32). More generally, medical therapies may have undesired side effects, surgical interventions may lead to complications, whereas diagnostic tests can produce either false positive or false negative results which may result in inappropriate or even harmful treatments. In addition to clinical effects, any clinical decision will have effects on patients preferences and service costs. Also, there may be trade-offs among competing objectives that decision makers need consider, such as maximizing quality of life vs maximizing life expectancy, vs minimizing the resources required.

Given the complexity of decision-making in health care, decision analytical models (DAM) are often used to help decision makers to correctly represent a decision problem and estimate the expected consequences of any possible action available to them. DAM structures and synthetizes the available information on a decision problem through a series of mathematical relationships to define possible (and uncertain) consequences originating from a set of alternative options being evaluated.

In the context of economic evaluation, DAM estimates both costs and relevant health outcomes of each option available to decision makers. By doing so, DAM provides insights to a series of resource allocation questions. Examples of such questions include: should a nationally funded healthcare system reimburse a novel total artificial heart for patients with advanced biventricular heart failure? Are the additional benefits of adding an artificial intelligence algorithm to detect intraoperative hypotension during surgical procedures worth its costs? Is it worth investing private or public limited research resources to reduce the uncertainty over the true performance of a new health technology?

In CEA based on DAM each possible consequence following a choice or decision is attributed a cost and an outcome. In addition, since decision makers do not know in advance how the uncertainty around a specific action will resolve (e.g., will a patient survive to a surgical operation which carries a small risk of perioperative death?) probabilities are used to express the likelihood of occurrence for each consequence following a decision. It is thus possible to calculate the expected cost and expected outcome of each option under evaluation. The calculation of expected values for costs and outcomes is a key feature in decision analysis. Mathematically, the expected value of an uncertain quantity (e.g., costs, life years or utilities) is simply calculated as the average of each possible value weighted by its probability of occurrence. This concept is analogous to the concept of the sample mean of an observed population and reflects the fact the consequences of options are variable. For example, apparently identical patients will respond differently to a given intervention, may or may not experience an adverse event or may remain in

hospital after surgery for different times. The variability in the possible outcomes between identical patients has also been defined as stochastic uncertainty.

In the early 90s, when the results of economic evaluations for new treatments started to be formally required by many governments to decide on pharmaceuticals reimbursement, there was a lively debate on the use of DAM for economic evaluation and CEA (33–35). Indeed, concerns were raised that, since DAM incorporated many assumptions and judgments at the discretion of the analyst, bias could be introduced either intentionally (e.g., due to conflict-of-interest issues, such as financial arrangements between analysts and sponsors) or unintentionally. Most of the debate around the use of DAM for economic evaluation could be traced back to the clash between two empirical paradigms (35). In fact, in line with the paradigm of evidence-based medicine (EBM) which was emerging in those years (36) biomedical scientists were increasingly focusing on studies based on experimentation to inform clinical decision making, thus giving higher priority to the internal validity of a study rather than the generalisability of its results. On the other hand, social scientists were traditionally more used to make use of observational data which are more likely to offer higher external validity and potential to generalise findings validity (at the costs of less robust estimates). At the time, in line with the theoretical principles of the EBM paradigm, many guidelines on economic evaluation expressed concerns about the methodology of economic evaluation based on modelling, arguing that claims about cost-effectiveness should be substantiated by adequate and well-controlled studies. This argument was consistent with the belief that a well conducted clinical trial should provide the sole source of evidence on resource use and health effects that, together with external valuation data (in the form of unit costs and utilities), forms the basis of the estimate of cost-effectiveness (37).

Nonetheless the limitations of this approach to express judgments on cost-effectiveness were soon recognized and DAM as a vehicle for economic

evaluation was increasingly used to overcome them. The main limitations of trial-based economic evaluation can be synthetized as follows (37).

Synthesis. A key concept of decision analysis is that all available information should be explicitly considered and used to inform decisions. For the relative clinical effectiveness of alternative treatments this principle is consistent with the cornerstones of EBM which puts systematic literature reviews and meta-analysis of randomized clinical studies at the top of the hierarchy of evidence. However, these principles extend also to the other information required for economic evaluation including data on resource use and consumption, and quality of life weights (utilities). In addition, given the breadth of data that is required for economic evaluations (clinical and epidemiological data, economic data, patients' health related preferences etc.) it is highly unlikely that the information need for economic evaluation will be satisfied by a single clinical trial. So, DAM provides an explicit framework where all the disparate sources of information can be synthetized and used consistently to inform the identified decision problem.

Most of the times, analysts may also need to synthetize different sources of evidence to derive a measure of *effectiveness* of the technologies of interest from the *efficacy* demonstrated in a clinical trial. This issue can be framed also in terms of the mentioned trade-offs between internal and external validity (38). In fact, the *safety and efficacy* estimate of a health technology from a well-conducted clinical trial will have high internal validity (absence of estimation bias) but may be poorly informative of how the same technology will perform outside the RCT in routine clinical practice where the environment of the experiment no longer holds. Therefore, adjustments to clinical trial estimates may be necessary by integrating trial data with other sources (e.g., real-world registries or epidemiological data) that consider the differences between the experimental setting and the real-world.

Lastly, due to the necessary limitations in the follow up of patients, many clinical trials have relied on intermediate or surrogate clinical measures of outcomes to demonstrate the safety or efficacy of a technology. Some examples are the use of progression-free survival for oncologic drugs as a

proxy for overall survival, or the rate of restenosis in patients' coronary arteries as a proxy of the effectiveness of coronary stents. While this evidence has been generally accepted to grant regulatory approval, it is not sufficient for CEA. This requires to use different sources of evidence to extrapolate a measure of patient-relevant health outcomes from the surrogate outcomes reported in the studies (39,40).

Analysis of all relevant comparators. The principles of economic evaluation prescribe that CEA should consider all relevant comparators to the technology under assessment, including those already used in clinical practice and those that for any reason are not. This axiom draws justification by the fact that since CEA is an incremental analysis, the non-inclusion of a relevant comparator will inevitably introduce a bias in the cost-effectiveness estimates. In contrast, most studies designed for regulatory approval of new drugs or medical devices often report evidence of efficacy compared to a placebo or a limited number of comparators. To obviate this limitation, over the years methods have been developed to indirectly extrapolate measures of relative efficacy from different sources for example through indirect and mixed network meta-analysis which can then be easily included within a DAM (41).

Uncertainty. The estimates of the model parameters drawn from clinical trials or other studies are inevitably subject to variable degrees of precision depending on the limited sample size of the studies. That is, there exist an uncertainty as to their true value which is known as parameter uncertainty.

The concept of parameter uncertainty is different from the concept of stochastic uncertainty mentioned before. Indeed, while stochastic uncertainty refers to the unexplained variation between patients (e.g., in responding to a treatment), parameter uncertainty refers as mentioned to the precision with which parameters have been estimated and therefore it expresses how confident we are about using the estimated values to inform decisions.

The correct characterization and propagation of uncertainty from the input parameters of the model to the output of the model (e.g., a measure of the

cost-effectiveness of the technology under assessment) is a key feature of DAM. Indeed, an explicit and transparent approach to characterize uncertainty is of primary importance in decision-theory. The following section describes in more detail the main types of uncertainty analyses that are conducted in healthcare.

2.3. Uncertainty analysis in decision analytical models – two perspectives

The societal costs of uncertainty over the cost-effectiveness of a technology are high. From the perspective of a public healthcare payer willing to maximize the technical efficiency of limited resources, improper, or insufficient evidence at the time of making adoption and reimbursement decisions may result in healthcare opportunity costs. These costs arise either when a promising technology is not reimbursed because of insufficient evidence (due to the missed opportunity for patients to benefit from a more cost-effective technology), or when technologies fail to confirm their expected value after early adoption in clinical practice (due to direct harms caused to patients and/or indirect consequences of an inefficient displacement of limited healthcare resources) (42,43).

Similarly, for manufacturers, uncertainty over cost-effectiveness will affect the likelihood of having a satisfactory level of return on investments in those jurisdictions where value for money is used as a criterion to prioritize healthcare spending. Since different products in pipeline compete for the same limited developmental resources, development opportunity costs are linked to the existing uncertainty because of the risk of taking wrong decisions about what technology to prioritize.

Given the potentially serious consequences of uncertainty, the value of decision models for CEA lies not only in their ability to provide a point estimate of the cost-effectiveness of a technology, but also in their capacity to systematically evaluate and transparently report the uncertainty around

this estimate and the decision being addressed. Indeed, this process of propagating parameter uncertainty into a measure of uncertainty around the decision being addressed is a key feature of decision models and is generally defined as uncertainty analysis (44).

Different approaches and methods exist to uncertainty analysis which mainly depend on the decision that the model seeks to support. This in turn depend on who is the decision-maker that will make use of the model results and the stage of development of the technology being evaluated. In the following sections the issue of uncertainty is described from two different perspectives: the perspective of healthcare payers, and the perspective of technology developers

The perspective of public healthcare payers

For decision-makers that are responsible for taking resource allocation decisions (e.g., the ministry of health in many publicly funded healthcare systems) analysis of uncertainty serves two main scopes: quantify and assess the uncertainty over a certain decision (e.g., how likely one intervention is the most cost-effective alternative) and ascertain the value of reducing such uncertainty by collecting more information before taking any decision.

If the decision that is informed by CEA can't be postponed, and/or it is unlikely that further evidence on the uncertain parameters will be collected, then, according to the principles of decisions theory, decisions should be based only on expected values of cost-effectiveness regardless of whether statistical significance has been reached (45). This is the case for example of an health system where reimbursement decisions are fundamentally binary i.e., purchasers of healthcare, decide whether or not to pay for the product on the basis of the price set by the manufacturer and the evidence available at launch, and have no role in commissioning or mandating further research (46). In a similar scenario, where decisions can't be deferred, decision makers will have to choose which of the mutually exclusive alternatives under evaluation should be selected, only based on the available evidence. In this case, decisions should not be based on standard statistical significance

because, as stated by Claxton (1999) in a seminal article, "*The opportunity costs of failing to make the correct decision based on the mean are symmetrical and the historical accident that dictates which of the alternatives is regarded as current practice is irrelevant.*" (45). In this case, the role of uncertainty analysis is limited. Notwithstanding, decision makers may still want to know how uncertain decisions are and what parameters contribute the most to the uncertainty in the decision being addressed.

However, if decisions can be deferred, uncertainty analysis assume a more relevant role in decision making. The economic value of deferring an uncertain decision is rooted in the in financial option pricing theory and its application to real-world analogies defined as real options analysis (47,48). The relevance of real-options for the evaluation of healthcare decisions has been firstly introduced by Palmer and Smith (49). The authors argued that, in presence of uncertain, irreversible decisions, the possibility of deferring the decision until some later time, when better information regarding costs and benefits may become available, has an economic value which is not accounted for in traditional economic evaluations based on net expected values (4).

The recognition of this value has led to the development of the Value of Information (VOI) framework that allows to explicitly estimate how reducing decision uncertainty by conducting further research would improve expected net benefits by reducing the chances of making the wrong decision. By doing so, VOI allows payers and HTA bodies to weigh the benefits of immediate, unconditional adoption against other policy options that limit or temporarily withhold access to a technology while mandating or commissioning additional studies (50,51). These types of policy options are often defined as coverage with evidence schemes (CED) and have been widely applied in Europe and other countries worldwide especially for pharmaceuticals products. The underlying theoretical justification and main literature on VOI are described in more detail below in section 2.4, whereas section 3.6 provides a more technical description on how to calculate VOI alongside CEA. The foundations of CED schemes as alternative to binary

accept/reject decisions for policy makers and the challenges relative to their design and implementation for medical devices are described in sections 2.5 and 2.6.

The perspective of developers

HTA aimed at informing developers' decisions, of which CEA constitutes a fundamental part, is usually conducted earlier in time when the design of the products and the business case are not yet defined. Due to this characteristic, the set of methods and approaches proposed for this scope has often been grouped under the general term of early-HTA. However, the more appropriate term of developer-focused HTA (DF-HTA) has been recently proposed by Boutell et al. (52). The authors argue that there are fundamental differences between the DF-HTA and the HTA conducted to inform usage decisions (use-focused HTA). These differences are argued to be more relevant than the timing of the analysis, and mainly arise as a consequence of the differences in the target audience and the decisions that the analysis is intended to inform. In fact, contrarily to public healthcare payers, whose aim is to maximize societal gains via improvements in population health or other goals (e.g., equity), the purpose of commercial developers is that of maximizing their long-term financial return on investments. In addition, technology developers mainly use CEA to inform strategic decisions on further product development, and how to prioritize technologies which are most likely to make it through the market. A number of reviews has explored the role of CEA for technology developers (22,53,54). For example, Hartz and John argue that CEA during product development may support technology developers by providing relevant insights on strategic R&D decision making, pre-clinical preliminary market assessments, go/no-go decisions, development of future trial design, assessment of future reimbursement and pricing scenarios and price determination (53). The use of CEA and uncertainty analysis for DF-HTA has been explored in several academic contributions and projects. For example, the Multidisciplinary Assessment of Technology Centre for Healthcare (MATCH) collaboration

in the UK has proposed a series of decision frameworks and tools focused on the iterative use of CEA throughout the whole product lifecycle (23,24,26). Similarly, other groups focusing on translational research, such as the ProHTA project and the Center for Translational Molecular Agency, built upon the MATCH experience and proposed other methodological guidance and applied studies (22,27,55–57).

In most of these contributions one predominant concept is that at early stages of development, when mature clinical and economic evidence is still lacking, uncertainty is considerable. Therefore, CEA is often proposed as a framework to conduct scenario analysis to draw very preliminary insights on the commercial viability of technologies (23,25). For example, using the "headroom" approach, developers can estimate what would be the maximum acceptable price for their technology given a prior belief on its relative effectiveness compared to existing alternatives and an exogenously fixed cost-effectiveness threshold representing payer's maximum willingness to pay for a unit improvement in health (23,25,26). This approach method is described in more detail in section 3.3.

Overall, uncertainty analysis in CEA for DF-HTA also hinges on the same methodological approaches to characterize parameter uncertainty of more-mature CEA for use-focused HTA. However as mentioned, because CEA are built at earlier stages of product development, the data to inform such analysis is usually sparser requiring stronger assumptions. For example, standard probabilistic methods to propagate uncertainty in DAM require that all uncertain parameters are defined by a probability distribution function. The distributional shape (e.g., the location and scale parameters) are usually informed from either primary or secondary (aggregated) data from experimental or observational studies. However, when evidence is still immature the precise form of probability distributions is usually informed by fewer data points or even elicited from expert opinions, and requires assumptions that are fundamentally unverifiable. The standard way to perform probabilistic sensitivity analyses in CEA is discussed in more detail in section 3.3, whereas section 3.4 introduces the Probability Bound analysis

(PBA) framework, a newly proposed method to characterize parameter uncertainty when there is total or partial ignorance on model parameters.

2.4. Do we need more evidence? Value of information analysis

The logic behind VOI analysis is that acquiring the necessary information through research could reduce uncertainty in the evidence base and the associated risk of making wrong decisions. However, generating scientific evidence is often costly and time-consuming. VOI analysis provides a formal assessment of the value of research, based on the extent to which the new information improves the expected payoff associated with the decision by reducing uncertainty (50). This can then be compared with the cost of conducting the research (58).

VOI can be used to prioritise research, by estimating the value of acquiring perfect information about all aspects of the decision (i.e., the Expected Value of Perfect Information, EVPI), or by estimating the value of perfect information about a specific (group of) parameter(s) in the decision (i.e., the Expected Value of Perfect Parameter Information, EVPPI). It can also inform research design, by estimating the expected value associated with a given sample size and a particular design, which results in a reduction of decision uncertainty (i.e., the Expected Value of Sampling Information, EVSI) (42).

The vast majority of theoretical and applied studies on VOI implicitly or explicitly adopted the healthcare payer perspective, i.e., they derived the value of conducting further research according to the objective of maximizing health gains in the population.

In this regard, VOI analysis has been extended to simultaneously evaluate the payoffs of alternative reimbursement and research design decisions and to estimate how different configurations impact population health benefits over time (51). In addition, VOI can be used iteratively over the life cycle of

a technology and can be reassessed whenever an element of the decision changes, such as publication of new research, the emergence of a new comparator technology, or changes in the relative prices of technologies. As mentioned, these approaches fit very well with CED and the dynamic characteristics of medical devices (59).

Rothery et al (60) have developed a formal approach for considering choices in research and technology adoption in the context of medical devices, based on VOI analysis, which can be used in situations where the assessment of the device has included the development of a cost-effectiveness model. They point out that there are four main coverage decisions possible (approve, reject, only in research, and only with research), the last two of which could be the basis for a CED scheme. They show that the choice between these options depends on several factors, including the likely cost-effectiveness of the device based on current evidence, the level of irrecoverable costs should use of the device need to be restricted or withdrawn (e.g., capital costs in equipment and facilities), the value of additional evidence, the incentives for the manufacturer and other parties to conduct research, the likelihood of changes in the price of the device or its comparators, and the value to patients (and more broadly to society) of early access.

The practical challenges of implementing VOI analysis have also been acknowledged in this literature. Fenwick et al (42), for example, note that 'Although VOI analyses are being increasingly published in academic journals, uptake in real-world decision-making remains limited. This is partially due to perceptions that a VOI analysis is complex to perform, difficult to interpret, requires substantial computational time and does not reflect key relevant uncertainties. Recent methodological developments have partly addressed these challenges by easing the computational burden of VOI using regression-based approaches to estimate EVPPI and EVSI (61).

As mentioned, the vast majority of theoretical and applied work on VOI has adopted a perspective which is broadly consistent with the policy aims of healthcare payers. Nonetheless, the same perspective and normative framework were also maintained when VOI was used within a DF-HTA

(22,23,54,62,63) to inform decisions of technology developers. However, this approach may lead to sub-optimal decisions for technology developers given the inherent differences in the objectives between these two groups.

To allow optimal decisions, the utility function used to estimate VOI for developers should reflect their own objectives rather than the societal objectives of healthcare payers based on net population health gains. This utility function may consider for example a positive return on investments (ROI) rather than population net health benefit as the developers' maximization objective. VOI analysis could then estimate whether further evidence generation would improve the expected likelihood of achieving a satisfactory ROI e.g., by increasing the likelihood of approval and/or affecting market shares after approval. For example, Willan (64), and Breeze and Brennan (65) used VOI to calculate the optimal sample size of a trial according to utility functions that reflect the developers perspective rather than the societal one. Specifically, Willan estimated how further research would increase expected profits by increasing the likelihood of regulatory approval, whereas Breeze and Brennan modelled both prices and expected financial return on investments as a function of trial results.

The recognition of different perspectives, reflected in the use of different utility functions, may also highlight the existence of misalignments in the perceived value of further research between developers and payers. These misalignments may be relevant because they can lead to inappropriate evidence at the time of reimbursement and can also affect the success of conditional reimbursement schemes like CED. Using a case study on a total artificial heart, in section 4.2 we explore the extent through which considering different utility functions may affect the value of further information according to the perspectives of developers and healthcare funders respectively.

2.5. Policy decisions under uncertainty. Coverage with evidence development for medical devices

As mentioned, uncertainty on the true performance of MD often exists in their early stages of diffusion (66,67,8,68). The underlying reasons for such uncertainty are various and include less stringent regulatory requirements for market access as opposed to medicinal products (69,70); difficulties with generating solid evidence from randomized controlled studies (RCT) (68,71) and the fact that specific characteristics of MD, such as their organizational impact, the existence of a learning effect and iterative product modifications make the results of available studies poorly generalizable in other settings or in the real-world (68,70,72).

Unlike pharmaceuticals, where evidence on efficacy and safety is legally required before marketing authorisation is obtained, devices usually only need to demonstrate performance and safety, with the CE mark acquired close to the point of market entry (68,73).

From a regulatory perspective, the regulatory pathway for MDs is specified in the Medical Device Regulation (MDR, EU 2017/745) (74) while the regulatory pathway for medicinal products is detailed in Community Code Relating to Medicinal Products for Human Use (Directive 2001/83/EC) (75) and Regulation (EC) No 726/2004 (76). The different regulatory regimens result in differences in the degree of involvement of the responsible authorities and the evidence requirements to obtain market access for MDs and medicinal products respectively. Regarding the degree of involvement of responsible authorities, manufacturers affirm that their MD complies with all applicable legislation by signing the Declaration of Conformity (DoC) and affixing the Conformitè Européenne (CE) mark (77), without any active involvement from responsible authorities. Following the application of the CE-mark a device can be placed on the market in the 32 countries, i.e., the countries within the European Economic Area, Switzerland, and Turkey (CE-region). Manufacturers self-certify MDs of risk class I, while conformity assessment of higher risk class devices requires the involvement of a private

Notified Body (78). The conformity assessment and the self-declaration differ from an assessment by a competent authority which, instead, is required for medicinal products. For the latter, competent authorities, which are the nationally designated regulatory authorities, actively review the evidence provided by a manufacturer and produce a summary opinion to the national authority or the European Commission (EC). Then, the responsible authority forms a legally binding contract with the manufacturer, i.e., a marketing authorization, which regulates the production, marketing, effective period, and product label. These differences imply that MDs are usually subject to a less degree of scrutiny by competent authorities as compared to manufacturers.

In addition, evidence requirements for regulatory approval are also different. In fact, while for medicinal product evidence on the efficacy and safety of products is a mandatory legal requirement, medical devices only need to demonstrate performance and safety. This difference implies that, for MD, the endpoints reported in pivotal studies used to obtain the CE-mark are rarely sufficient to inform about the clinical value of the technology. Despite the new regulation enforced in May 2021 has reinforced the rules on clinical evidence, as opposed to the previous Directive 2007/47/EC (75), MD manufacturers are not required to generate clinical evidence through RCT not even for higher risk devices.

In addition to not being legally required, well conducted RCT for MD may also be ethically or practically difficult (71). For example, one important issue is that MD often undergo product modifications which are informed by their use in clinical practice. This aspect makes the timing of the assessment particularly relevant, in that, by the time the trial reports its results the device may be considered already outdated in favour of newer generations with substantial design changes. Another issue for which timing is essential is related to the experience of the surgeon and trial personnel with the device. In fact, differently from medicinal products, the existence of a device-user interaction implies a learning effect which may confound the results of a trial. Many studies showed that procedures performed while the operators are in

a learning phase are associated with greater risks and adverse events compared to procedures performed when the learning curve has reached a plateau (79,80). If trials are initiated too early, when the experience of the surgical team with the new device is still immature, the risk is that trials will measure the inexperience with the new procedure compared to the experience with the old one rather than the true efficacy of the device. Conversely, an assessment by an RCT conducted too late could also be problematic if doctors/ care providers no longer adhere to a study protocol, or if the recruitment of patients declines dramatically, for example because of the more widespread diffusion of the "experimental" device. In fact, recruitment to the study may become difficult if clinicians and/or patients think that one treatment is better than another and therefore have clear preferences as to which one they would like to choose. A similar issue of acceptability may also affect the possibility of including one or more comparator groups which are either more invasive (for example laparoscopic versus traditional surgical procedures) or already widely known (81). From an ethical point of view, it may be poorly justifiable to offer patients a sham procedure. In addition, blinding in clinical trials for RCTs may be more difficult for either ethical or practical reasons and this may cause issues with measurement bias due to doctors or assessors' subjectivity.

In addition to the clinical performance of medical devices, at the time of an evaluation, uncertainty also exists on the real impact of a technology on healthcare costs. For example, the extent to which a ventricular assist device (VAD) will reduce hospitalization costs by allowing the discharge of patients after implant will depend on how good the technology is in stabilizing patients' heart conditions and improving their capacity to perform tasks in autonomy. Uncertainty over the medium- and long-term performance of devices may affect both clinical and economic outcomes. For example, uncertainty over the true revision rates of hip or knee implants in the target population, or the time before battery depletion of active pacemakers will inevitably result in uncertainty over the true costs of the technology in the medium and long term.

For all the above mentioned reasons, decisions about the coverage and reimbursement of new medical devices are inherently uncertain as, at the time of market launch, only limited information is available about their real-world cost-effectiveness and value for money (82).

Traditionally, payers have borne the risk of making wrong coverage and reimbursement decisions in the presence of uncertainty regarding the real-world performance of health technologies. A wrong decision may occur when a health technology is reimbursed that later does not confirm its original claims on safety, efficacy, and/or (cost-) effectiveness (type I error), or when a technology is not reimbursed that later proves to be more safe and (cost-) effective than relevant comparators used in clinical practice (type II error) (83–85). Regardless of the type of error, any wrong decision is undesirable as it will likely always lead to loss of benefits to patients and an inefficient use of available resources. The risk of making a wrong decision and the evidence gap between requirements for regulatory and market approval on the one hand and coverage and reimbursement decisions on the other hand have led to the introduction and increased use of CED' schemes (82,86,87).

CED schemes aim to reduce uncertainties associated with the safety, efficacy, and (cost-) effectiveness of health technologies and share the risk of making a wrong coverage or reimbursement decision between payers and manufacturers (82,87,88). These schemes go under different names in different countries, for example, 'coverage with evidence development schemes' in the USA, 'conditionally funded field evaluations' in Canada (Ontario), interim funding schemes' in Australia, 'only in research (OIR)' and 'only with research (OWR)' in the UK (England/Wales), and 'conditional reimbursement schemes' in Belgium and the Netherlands (87,89). However, these schemes typically are all defined as performance-based risk sharing agreements (PBRSA), i.e. "any plan by which the performance of the product is tracked in a defined patient population over a specified period of time and the level or continuation of reimbursement is based on the health and economic outcomes achieved" (82). Following this definition, CED schemes

cover schemes that manage utilization in the real world and link reimbursement to the performance of a health technology as well as schemes that provide additional evidence with the aim to reduce decision uncertainty (82).

Despite the growing interest in CED schemes, they are often costly, complex, and challenging (90,91). In response to these challenges, a good-practice task force of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) formulated four good practice questions that need to be addressed when applying a CED scheme. These questions concern: 1) the desirability of the scheme (as opposed to some other reimbursement or research arrangement), 2) the choice of research design, 3) the approach to implementation, and 4) the method used for evaluating the scheme (82). The following section explore more in detail the several challenges that concern all these phases of a CED scheme.

2.6. Systematic literature review on the challenges of CED for devices

A systematic literature review was conducted as introductory work for this thesis. A full description of the methods and selection process of the studies included in this work are reported elsewhere (1). Here the main results are summarised. In the review, 17 challenges were identified as associated with all phases of CED schemes for MD. These challenges concern: 1) deciding on whether a CED scheme is desirable, 2) understanding the relevant uncertainties and risks, 3) lengthy and complex negotiations, 4) defining the decision problem, 5) data requirements, 6) identifying meaningful outcomes, 7) defining an adequate duration for a scheme, 8) market entry of new technologies during a scheme, 9) obtaining funding, 10) obtaining informed consent, 11) quality of the data, 12) deciding on when a scheme is considered successful, 13) withdrawing a technology, 14) lack of transparency, 15) lack of governance, 16) stakeholder involvement, and 17) ethical issues. Table 2-1

presents a description of each of these challenges. A full overview of the challenges, including those extracted from studies discussing the challenges in the context of CED schemes for pharmaceuticals, is available as Supplementary Material S2.1.

These challenges apply to some extent to all CED schemes for different types of technologies; however, five relate directly to the characteristics of MDs, and hence are specific to MDs. Most of these specific challenges were discussed by Rothery et al. (60) and relate to deciding on whether a CED scheme is required, understanding the relevant uncertainties and risks, identifying meaningful outcomes, defining an adequate duration for a scheme, and market entry of new technologies. For example, it may be particularly challenging to decide whether a CED scheme is required for a MD as the prices of MDs are likely to change over time due to rapid incremental MD innovations and market entry of new MDs. This may directly impact the uncertainty associated with making a wrong coverage or reimbursement decision and the value of further research into MDs' cost-effectiveness. Incremental innovations, high upfront irrecoverable costs, and market entry of new MDs may also result in a disincentive for (individual) manufacturers to invest in research that would reduce uncertainty about MDs' efficacy. This may lead to a situation in which research costs shift to public fund holders. Furthermore, the changes over time in prices of MDs and due to gradual innovations can complicate defining an adequate duration for a scheme and this, in turn, may impact the identification of and data collection on meaningful outcomes. The identification of meaningful outcomes may also be particularly challenging for MDs as the outcomes of interest are typically not only influenced by the MD, but also by the subsequent treatment, e.g., as is the case for diagnostic MDs such as positron emission tomography (PET) scans. Furthermore, MDs' effectiveness is not only influenced by characteristics of the MD itself but may to a large extent be influenced by the learning or training of physicians that is required to achieve the optimal effect. The associated learning curve of physicians may result in a more modest impact on patient outcomes or higher costs during

the early use of MDs, resulting in a lower cost-effectiveness of MDs when assessed in the short run or early in the development phase.

A qualitative assessment of the similarities and differences between challenges associated with CED schemes for MDs and other types of technologies, e.g., pharmaceuticals, did not reveal any challenges with CED schemes for other types of technologies that do not also apply to MDs. However, we identified three challenges that were discussed in the context of CED schemes for pharmaceuticals that were not found in the included studies, yet are also considered to be applicable to MDs. The first challenge concerns the information asymmetry between payers and manufacturers about the potential real-world performance of a technology and the impact this may have on CED-scheme agreements (92). The second challenge concerns the ex-ante definition of a final decision rule based on the gathered information and 'exit strategy'. It needs to be defined when the (cost-) effectiveness and/or safety of a technology is deemed to be below expectations or some relevant threshold, leading to its withdrawal or a premature termination of the CED scheme (86). Ideally, this would also entail a withdrawal implementation plan. The third challenge concerns the economies of scale in the management of CED schemes and the difficulties small countries may have in applying CED schemes because of the associated costs and monitoring mechanisms (92).

Table 2-1 Identified challenges of coverage with evidence development schemes for devices

Domain	#	Challenge	Description
Desirability	1	Deciding on whether a CED scheme is required	Whether CED schemes are recommended depends on both the characteristics of the technology (whether it is expected to have a positive net benefit, whether evidence can be generated following reimbursement, and whether there would be a cost in reversing the decision at a later data) and the range of authority of the purchaser (whether they can delay a decision or review it at later date, whether they can negotiate price, and whether they can ensure that research is actually conducted) (93); Generally, there is a lack of criteria and formal guidelines that can help decide whether a CED scheme can help reduce uncertainty and should be initiated (87,94,95); The question of whether or not further research is worthwhile requires some assessment of how uncertain a decision based on expected cost-effectiveness might be, what the consequences are likely to be if an incorrect decision is made, and what a technology that has been subjected to a CED scheme is displacing (96); There is a close link between the value of a MD, the value of further research to reduce uncertainty and the price of the MD. These links can offer incentives for manufacturers to price accordingly and decide whether there is sufficient value from further evaluative research. The value of additional research can be informed through VOI analysis (60). However, VOI analysis may be difficult to apply in specific cases and a formal guideline may help decide whether research in a particular area is practical and likely to reduce uncertainty (94). This should be enhanced, in particular by clearly stating the selection criteria for MDs that may benefit from such approaches (95); There is a concern that CED schemes could stifle or slow innovation by creating a disincentive to develop new products for conditions for which the evidence base is not well developed or by raising the evidentiary standards (97,98);

Domain	#	Challenge	Description
			CED schemes may have the unintended effect of lowering industry investment in evidence development and shifting research costs to public fund holders (99); There is a concern that manufacturers use CED schemes as a mechanism to secure beneficial formulary placement, gain market share, and increase patient compliance. Manufacturers may be reluctant to take on the risk of a CED scheme when they cannot predict how their product will be used in the real-world population (100).
	2	Understanding the relevant uncertainties and risks	There are challenges in assessing risks upfront due to uncertainties in the real-world performance of a technology and further research is unlikely to be able to resolve all uncertainty (60); Some uncertainties cannot be reduced by further research and may resolve by other changes occurring over time. For example, the effective price of the technology and/or its comparators may change. The price plays a key role in determining the value of the technology, but it also affects the level of uncertainty by changing the likelihood of making an incorrect decision and the value of further research (60). Other uncertainties that cannot easily be resolved by further research may concern previously unrecognised adverse effects that emerge in the long term and changes in market conditions that might cause the price of the technology to drop in the future (88); One of the complexities associated with the evaluation of MDs is the fact that any decision about the adoption of the MD will interact with the ability to gather more evidence and may affect future commercial developments of the technology (60); There is a group of concerns relating to the introduction of additional uncertainty for manufacturers in terms of expected returns, which may have the opposite effect of dis-incentivising additional data collection, the advantage competitors may take of data collected by the manufacturer, and related to this is the problem of free-riding (101).
	3	Lengthy and complex negotiations	Defining the study design is often lengthy and complex, and it may be difficult to reach contractual agreement (100,102); Deciding on the point in the product life cycle at which a technology should be assessed is a contentious issue and various stakeholders may have different views on the technologies that

Domain	#	Challenge	Description
			require further study, the questions that need to be answered, and the necessary methods for answering those questions. It requires the creation of working groups made up of key stakeholders and opinion leaders who are involved in designing the study questions and methods from the beginning of the process (103,104); There is protocol development, sample size and site determination, case report form development, contracts with sites and investigators and dealing with multiple ethics boards submissions, therefore, study initiation is often subject to contractual and legal delays (102).
Research design	4	Defining the decision problem	The decision problem is rarely stated explicitly and this creates the risk that the study design does not address the decision problem or is not designed to feasibly address that problem (105); The research design that is most appropriate depends on the nature and type of the uncertainty that the CED scheme is trying to address, e.g. uncertainty about whether the medical product or service will be used in the right patients or uncertainty at launch about clinical or economic outcomes (106).
	5	Data requirements	A formal guideline for CED schemes should be accompanied by a clear statement regarding what study design and data are required to reduce uncertainty (94,98). Requirements are often not specified and laws can be unclear at this point (95); The study design that is required to answer questions of evidence development is often not clearly defined, especially concerning the need for RCTs or observational/not experimental designs (98); For the establishment of registries, there are generally no guidelines available (85);
	6	Identifying meaningful outcomes	Outcome measures should be clear, measurable, objective, realistically achievable (in relation to the duration of the scheme), and relevant (82); It is important to be certain that the outcomes of interest are largely influenced by the technology concerned (88); Manufacturers and payers may shy away from agreements in disease areas where there are many different treatment paradigms or the relevant outcome is an intermediary outcome,

Domain	#	Challenge	Description
			because it can be challenging to attribute the outcome to the product in question. There is also a risk for manufacturers with being responsible for outcomes when they cannot control the way a technology is used (100); When questionnaires for data collection are designed by physicians, they may not be ideal for use in an economic analysis (85); In some cases, the 'right' outcomes may not be identified until the scheme is implemented, resulting in failure to capture the data needed to reduce the decision-making uncertainty (90).
	7	Defining an adequate duration for a scheme	Designing the necessary clinical research, getting funding, and implementing a scheme in a time frame that is consistent with the needs of clinicians, patients, and other decision makers is challenging (104); With a typical three- to five-year political cycle, there is often a tension between research and political needs (107); Short-term schemes are not desirable given the considerable investment in evidence development, while long-term deals are also not desirable given the costs and risks involved (100); The unique characteristics associated with MDs, such as rapid incremental innovation, learning effects, and upfront irrecoverable costs all present a challenge for the timing of reimbursement decisions and the value of waiting until additional evidence is conducted to support the technology (60); In view of the pace of technological changes in healthcare, a CED scheme of more than three years may be of limited relevance for MDs (103,108). This is because the kinds of policy questions that such studies inform have the habit of changing, for example, as other technologies become available for the same patient group (88).
	8	Market entry of new technologies	The information generated by research will not be valuable indefinitely as new and more effective interventions may become available and make the information no longer relevant for future clinical practice (60); Market access of incremental MD innovations and new technologies may make existing ones obsolete or change their cost-effectiveness (96);

Domain	#	Challenge	Description
			Rapid approval of new entrants may result in a disincentive for manufacturers to invest in further research that would reduce uncertainties about MDs efficacy (60); MDs that enter the market during a CED scheme may not be included in the scheme, and hence reports may be based on evidence from only one MD (85).
Implementation	9	Financial arrangements and costs	The costs associated with CED schemes can be substantial and a barrier to establishing a viable and cost-effective scheme (82,94,109,107,100,85); CED schemes are perceived to have high transaction costs and be difficult to execute, particularly given the fragmented payer system with patient movement across plans, as well as the current lack of data infrastructure that limits feasibility and, to some extent, interest in measuring long-term outcomes (100); It may take years before funding is ensured and then there may still not be sufficient funding to generate the evidence needed to reduce uncertainties and meet the HTA agency and decision makers requirements (96,104,89); Lack of experience with CED schemes, staff turnover, billing requirement complexity, and inconsistency of non research and research requirements may add up to significantly more time and effort than anticipated, at times for studies with no funding for administration (110); Some have suggested establishing public-private partnerships between payers and manufacturers, while others have stressed the importance of locating publicly funded research organizations who may be perceived as neutral and, therefore, better able to provide control over research design and data, and manage vested interests (90,98).
	10	Obtaining informed consent	Identifying and counselling potential participants and obtaining informed consent requires considerable effort and patients may decline to participate or may prematurely withdraw. The need for frequent reconsent, e.g. when regulations change mid study, should be taken into account (110);
Evaluation	11	Quality of the data	It may be difficult to obtain consensus amongst stakeholders about what is considered an acceptable quality of evidence (96);

Domain	#	Challenge	Description
			After coverage or reimbursement is obtained for a technology, there may be a lack of incentive, e.g. for manufacturers, to collect the data (104); There is the risk that research may not happen, does not answer the initial questions, does not feed back into decision making and the technology is funded anyway, or does not deliver the evidence while the funding cannot be stopped (87); Evidence generated may not meet the quality criteria or be sufficient for making coverage decisions, e.g. when relying on observational data alone (87,90,89); The accuracy, reliability, and completeness of the data, e.g. when submitted to a registry, often depends on physicians and they may not always have the necessary time to complete the forms accurately (85,111); Physicians are not (always) paid for data collection and reporting, which may affect the quality of the data, and there may substantial missing data, e.g. due to loss to follow-up in registries, which may lead to bias (85); There is an additional burden to monitoring a CED scheme and of collecting and analysing the data collected as part of a CED scheme. This may affect the quality of the data (103); For the success of a scheme, it is imperative that payers and manufacturers trust the data and clear agreements on data validation and analysis are important to create this trust (100).
	12	Deciding on when a CED is considered successful	There still is little evidence to support the claimed benefits of CED schemes and the extent to which some of the challenges involved in CED scheme implementation impact on the final outcome (101); It is not clear if CED schemes succeed in limiting reimbursement to specific patient subgroups and payers are sceptical that CED schemes will reduce costs in the long run (97,101); As it may not be possible to assess the VOI generated by a CED scheme directly post hoc, there is a need to rely on process indicators for assessing a scheme's success(82); Whether the CED scheme has achieved its objectives and can be considered good value from a health system perspective is linked to the desirability of the scheme and can be addressed

Domain	#	Challenge	Description
			from multiple perspectives: manufacturer, patient, payer, provider, and society. A comprehensive evaluation will therefore need to consider multiple perspectives (82); The success of CED schemes when manufacturers are asked to conduct the research will depend on whether the authorities are able to establish contractual arrangements as part of a CED scheme , that is, arrangements that can be monitored and enforced with credible penalties to ensure that agreed research is conducted and in the way intended (96).
	13	Withdrawing a technology	Once a technology is used in practice—even if formally temporary—ending reimbursement may be less feasible than initially not reimbursing it, especially when the technology proves to be effective, but not cost-effective (103,82,91); Decisions to withdraw may cause heated discussions with doctors, patients, and politicians and be followed by a public debate in the media (87); Patients may also be more motivated to exert political pressure to secure or maintain coverage of last-line treatment for life-threatening illnesses than for preventative or 'me-too' interventions. Inertia in clinical practice may be a barrier to delisting, particularly for interventions with a long-standing place in both formularies and clinical practice. Payers may adopt a passive role and rely upon clinicians to modify their prescribing practice to replace inferior interventions with more effective or better-tolerated alternatives as and when they become available. Evidence development may be delayed if the default position is to extend funding until the data become available (99).
Other	14	Lack of transparency	There is a general lack of information on CED schemes in the public domain that is attributed to 'commercial confidentiality'. Consequently, payers who consider CED schemes as a potential policy option have little information upon which to base a decision (90); There is little information available on the agreements implemented, their objectives, and evaluation of their impact. This prevents cross-country learning and limits the ability of patients to engage with CED processes (101); Disclosure of the results of previous schemes related to a technology of interest may reduce duplication of efforts. Mechanisms for increasing transparency around key components, e.g. objectives, conflicts of interest, data collection management, and oversight, of the scheme

Domain	#	Challenge	Description
			that respect commercial interests are required to build on previous good research practices for specific types of studies (82,90).
	15	Lack of governance	Lack of project management and coordination can be an obstacle for CED schemes and can make it difficult to ensure an update of the recommendation following the production of new evidence (87,89,96); The independence of a scheme from any party with a vested interest in its outcomes should be ensured (105); Stakeholders may take contradictory positions (also amongst themselves) around where the leadership should rest and which stakeholders should be involved in a CED scheme (109); Supervision of the research may create a conflict of interest for a HTA body as they need to keep the image of being a helper for a better quality healthcare system (87).
	16	Stakeholder involvement	The various stakeholders can affect political decisions around the initiation of a CED scheme. For example, manufacturers may pressure the initiation of a scheme and conflicts of interest may arise when manufacturers play a role in the funding, data collection, and evaluation of a scheme (109); Patients, generally, have limited opportunities to engage in the development of a scheme and not all patient groups are aware of what CED schemes entail (101); Patient advocacy groups may be unwilling to accept this policy especially if the assessed treatment is considered to be safe and efficacious (96). They may distrust the motives of payers in their efforts to support evidence development through coverage, and may assume that the primary objective is cost containment, rather than a genuine effort to support early access to innovations and clinical research (97); There may be significant opposition from the clinical community and compliance with data collection by physicians may be weak, e.g. because of lack of staff (87); There is the risk that CED schemes are perceived as a tool for monitoring or controlling physicians, particularly in the context of registers on interventional procedures with or without the use of MDs (100);

Domain	#	Challenge	Description
			Compliance with data collection by physicians may be weak and the monitoring of the study poor because of lack of clinical staff (87); The translation of evidence into policy is riddled with political and economic considerations, both the overt political process involved in CED and the role of the pharmaceutical industry. The most explicit evidence of relations of power comes from the hierarchy of roles in the decision-making process. Political influences play a role in determining where the money for CED will come from and where the ultimate decision-making comes from (109).
	17	Ethical issues	CED schemes may be beneficial for future patients, but they can impose significant opportunity costs on current patients. Some individuals in the present population may benefit from the research condition because they will also be members of the future population. However, this will not be true for all and so the issue of balancing the interests of some individuals in the present population against some individuals in the future population remains (96); Various stakeholders, e.g. policy makers and patient groups, have questioned whether it is ethical to restrict access to technologies to patients participating in registries and clinical trials, and to withhold a potentially beneficial innovation from a subset of patients who cannot, or will not, participate while providing it to another (87,96–98,104,112). It is also questioned whether study participation concerns coercion and whether patients' informed consent is valid in this context (97,104,112); Patient advocacy groups may be unwilling to accept a CED scheme, especially if the treatment has demonstrated safety and efficacy (98); Furthermore, CED schemes may result in inequities as participants in the treatment arm may receive better treatment than those in the other arm and those not participating, and treatments may not be available in all geographical areas (96,112).

AWR, approval with research; CE, Conformité Européenne; CED, coverage with evidence development; FDA, Food and Drug Administration; GRP, good research practice; NICE, National Institute for Health and Care Excellence; OIR, only in research; RCT, randomised controlled trial; USA, United States of America; VOI, value of information; a For reasons of clarity, we used the term CED scheme in this table, where the author(s) sometimes used the terms performance-based risk sharing agreement or access with evidence development scheme. The classification of challenges into 'Desirability',

'Research design', 'Implementation', and 'Evaluation' of CED schemes relate to the four good practices questions that were formulated by ISPOR's 'Good Practices for PBRSA Task Force' (82). 'Other' relates to challenges that fall outside the scope of these questions.

Chapter 3
Cost-effectiveness models and characterization of uncertainty

In this chapter the main methods used in the rest of the thesis are presented and described in detail. Because of the multidisciplinary nature of health technology assessment and economic evaluations in healthcare, this methods chapter has included a general introduction to the most common and widespread approaches to conduct model-based CEA, as well as a more in-depth description of novel methods or approaches proposed. This was done with the objective to facilitate comprehension to the broadest possible audience which may be interested in the topics covered by this thesis. Therefore, section 2.1 introduces the main decision rules used in CEA to take decisions about coverage and reimbursement of health technologies, whereas section 2.2 provides a general description of Markov models, which are one of the most common modelling approaches to CEA. Section 2.3 focuses on the methods used to characterize existing uncertainty in the model parameters and the standard methods to propagate it throughout the model to provide a measure of decision-uncertainty. The issue with conducting uncertainty analysis are addressed by the dual perspective of healthcare payers and technology developers. Sections 2.4 and 2.5 introduce a novel non-parametric method to characterize uncertainty and provide a step-by-step guide to implement it. Finally, section 2.6 describes in detail the methods to conduct value of information analysis and proposes a novel application to verify the existence of misalignment in the incentives to conduct further research from the different perspectives of technology developers and healthcare payers.

3.1. Decision rules for cost-effectiveness

In CEA, both health benefits and resource consumption in the care process of two or more alternatives are estimated and combined to provide a measure of cost-effectiveness. Cost-effectiveness may be expressed in terms of incremental cost-effectiveness ratio (ICER), that is the ratio between the incremental costs of the new technology compared to a relevant alternative, and the incremental health benefits expressed in units of health (e.g., life years gained) or Quality Adjusted Life Years (QALYs). For example, the ICER of a new technology (technology B) compared to its best (and cheaper) alternative (technology A) is calculated as:

$$ICER_{BA} = \frac{Costs_B - Costs_A}{Health_B - Health_A}$$

Intuitively, this ratio estimates the additional cost that is required to obtain an additional unit of effect (e.g., an additional QALY) when using technology B as opposed to technology A. Note that the ICER is by construction a pairwise comparison and therefore a specific process needs to be followed in presence of more than two comparators to rank and estimate the ICER for each pair of alternatives (113,114)

When the alternative being evaluated is both more costly and more effective than its best alternative, cost-effectiveness is assessed by looking at whether the ICER is inferior to a pre-defined cost-effectiveness threshold k. When adopting a decision-maker perspective, the underlying assumption is that the cost-effectiveness threshold represents the opportunity costs (in terms of health forgone) that will occur after displacing resources to fund the new intervention from other parts of the healthcare system (115,116). According to this perspective the objective of CEA is that of supporting decisions that aim to maximize health given a limited and exogenously fixed budget constraint.

Alternatively, to the ICER, other summary measures of cost-effectiveness based on the concept of "net benefit" are broadly used when doing CEA. While these measures require the same data to be estimated and serve the same objective of indicating which alternative is the most cost-effective, they differ from ICER in some important aspects (113). First, while the ICER is always a comparison between two alternative strategies, net-benefit measures can be calculated individually for each strategy included in the CEA, and this facilitates comparisons and identification of the most-cost-effective in presence of multiple mutually exclusive alternatives. Second ICERs are non-linear (as they are expressed as a ratio), whereas net-benefits are linear. This difference implies that net benefit measures have nicer statistical properties. For example, they are monotonically increasing with higher levels of "effects" and monotonically decreasing with higher costs, and therefore ranking of alternatives is much simpler using net benefit as opposed to ICER. In addition, net-benefits tend to be normally distributed, thus facilitating the estimation of measure of precision such as confidence intervals and more advanced sensitivity analyses such as probabilistic sensitivity analysis and cost-effectiveness acceptability curves (CEAC) (117) (§3.3).

There exist two standards measures of net benefit, the net health benefit (NHB) and the net monetary benefit (NMB).

For each strategy *a* included in CEA the NHB is calculated as:

$$NHB_a = Health_a - \frac{Costs_a}{k}$$

The NHB consists of two parts, the first representing the health benefits of patients receiving a strategy, and the second representing the health forgone by other patients as a consequence of the displacement of resources from other part of the healthcare system. This health loss is also defined as the health opportunity cost of the budget constraint and is calculated by dividing the costs required to fund the new strategy by the cost-effectiveness threshold *k*.

The net monetary benefit (NMB) is strictly related to the NHB and expresses the "monetary" value of the gains in benefit. This can be calculated easily by multiplying the NHB for the cost-effectiveness threshold *k* or alternatively using the following formula:

$$NHB_a = Health_a * k - Costs_a$$

Note while the ICER does not require a cost-effectiveness threshold for its calculation (although it requires it to be evaluated), both the NHB and NMB incorporate the threshold in their formula, requiring it to be defined prior to calculation. This means that the threshold should be exogenously determined, for example by the national authority responsible for health technology appraisals and adoption decision. One example of using an explicit threshold is the UK, where the threshold value used for technology appraisals of pharmaceuticals is established in a range between £20,000 - £30,000 per QALY gained.

3.2. Markov models

As mentioned in §2.2, DAM are often used conduct CEA. These models identify the key possible courses of a medical condition and attach probability values to each of them which are specific to the treatment alternatives being evaluated. Expected costs and consequences (e.g., in terms of survival, or quality of life) for each treatment alternative are then calculated.

Several types of DAM exist that can be used depending on the decision problem, the condition (e.g., acute or chronic condition) and the type of technology under assessment (e.g., an implantable medical device, a diagnostic test or a screening programme). Among DAM, Markov models, are widely used for long-term or chronic health conditions (118). These models assume that a patient is always in one of a finite number *s* of discrete health states called Markov states. Events are represented as transitions from one state to another, and occur with a given probability *p*.

When building Markov Models, the following steps are usually performed:

1. Each state *s* is associated to a cost (C_s), which represents the total value of the healthcare resources consumed per unit of time while the patient is in that state (e.g., drugs, hospitalization costs and outpatient visits); and a consequence, which may be measured in terms of clinical effectiveness (E_s), (e.g., life years gained, new cases diagnosed or reduced cardiovascular events), or be an expression of health-related preferences for a specific health state (e.g., QALY).
2. The velocity and directions of the transitions of patients between states are determined using probability values. The values of these probabilities may be dependent on each treatment under evaluation. For example, the probability of transiting from a progression free state to a recurrence state will depend on the oncologic drug received by the patient.
3. The model is evaluated for a certain amount of discrete time units, usually corresponding to the *lifetime* of the target cohort of patients.
4. The total costs and consequences are estimated by considering the values assigned to each state and the time spent by patients in each state.
5. Sensitivity analysis on parameter uncertainty is performed through a set of the deterministic and probabilistic analyses

More detailed description of how to build and evaluate Markov models is provided elsewhere (29,119). In the next section a more detailed description of the methods traditionally used to perform sensitivity analysis and characterize parameter uncertainty is provided.

3.3. Sensitivity analysis and characterization of uncertainty

Deterministic sensitivity analysis (DSA) is the simplest way of exploring parameter uncertainty. DSA generally consists of varying one or two input

parameters at a time and see how such variation impact on the ICER or net benefit. While not enough to correctly characterize the overall decision uncertainty, DSA can be useful especially in early models to understand which one among the input parameters is the most likely to affect the results of the CE model. In fact, DSA's main advantage is that it is easy to implement and can provide rapid insights on the relevance of individual uncertain parameters. Its simplicity of use and easiness of interpretation make DSA particularly suitable to be included in ready-to-use tools to have a rapid glance over the potential cost-effectiveness of a technology in the very beginning phases of product development (6).

However, several limitations regarding DSA have been outlined: (i) the range of variation for the values of the parameters is often chosen arbitrarily, and no insight is provided on the effects of input values at the margin, (ii) non linearities between input parameters model and the output of interest are not visible, (iii) DSA does not provide any insights on the likelihood that input parameters take a certain value in the considered range and (iv) correlations between parameters are not taken into account (120–125).

Probabilistic Sensitivity Analysis (PSA) is the standard approach to uncertainty analysis that overcomes most of the limitations of DSA and allows a full characterization of decision uncertainty. Briefly, PSA is performed by assigning probability distributions to all uncertain parameters in the model. Information on the location and ancillary parameters of the probability distributions are often estimated either from statistical models of primary data (e.g., using individual patient level data from the clinical studies) or from secondary published data using the methods of moments (126).

Once the uncertainty over each single parameter is characterized, the overall uncertainty is propagated from the model input to the model output using Monte Carlo simulations. The process for MC simulations is straightforward. First, a single value for each of the uncertain parameters in the model is sampled from the assigned distributions. Second the output of the model, for example in terms of the NHB of the different alternatives, is calculated for the sampled set of values from all parameters. Third, the described

procedure is repeated many times (e.g., >1000 simulations), and for each of the simulations, the output of the model is stored. Forth, decisions are then informed by the resulting mean of the output value of the model (e.g., the NHB) obtained by the averaging out the MC simulations (119,127).

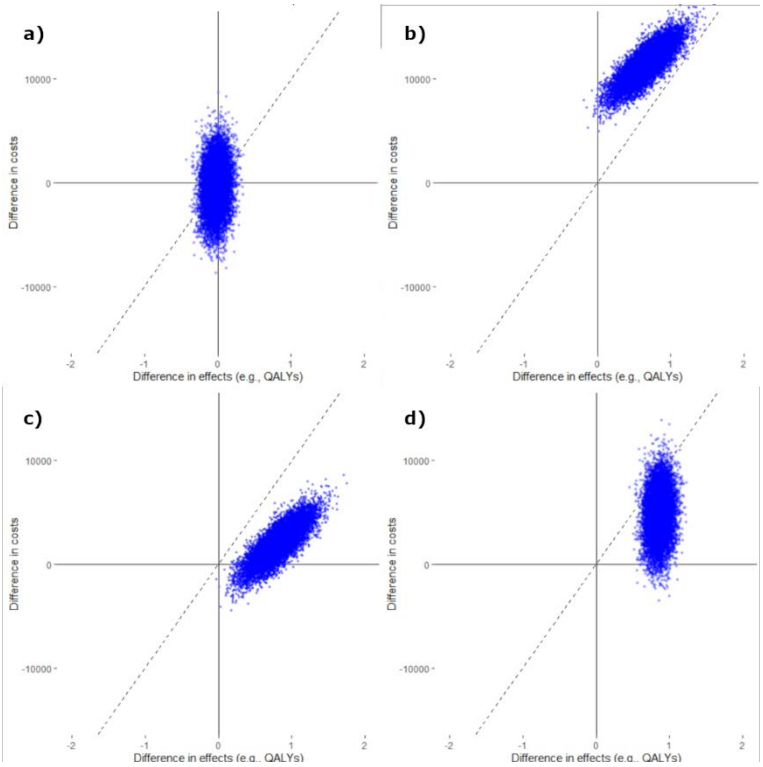
Many different guidelines exist that provide indication of how to choose parameter distributions for PSA In general, due to the central limit theory, normal distributions are always potential candidates. Nonetheless, depending on the sample size of the data used to inform the parameters of the distributions, sampling from a normal distribution may lead to unplausible or biased values. For example, with small sample sizes, the confidence interval of the mean of a probability value when using a normal distribution may be negative, which is of course impossible. Because of this issue, other distributions are usually employed which have convenient characteristics for different groups of parameters. For example, beta and Dirichlet distributions are used for binomial and polynomial probabilities, gamma distribution for costs, and log-normal distributions for relative risk parameters.

The choice of the distribution should also consider any potential correlation across parameters, for which multivariate distributions should be used. For example, the hyper parameters of a parametric survival model should be sampled by the same multivariate distribution.

The simulations obtained through MC simulations can be represented in a standard cost-consequence plane, representing the difference in the effects between two or more alternatives on the x-axis (ΔE or ΔU), and the difference in costs (ΔC) on the y-axis. Figure 3.2 provides 4 different examples of a comparison between two alternative strategies. The position of the simulated points across the 4 quadrants of the plane immediately provides an idea about the cost-effectiveness of the alternative being evaluated. For example, if the points of the MC simulation lay in the first quadrant (panel b in Figure 3.2) the technology is both more costly and more effective compared to its alternative. Note that in the cost-consequence plane, the cost-effectiveness threshold k is represented by the slope of a

straight line passing through the origin of the axes. Any point of the simulations laying above such line is not cost-effective (i.e., the ICER will be higher than the threshold *k*, and the NHB will be negative).

Figure 3-1 Results of a Monte Carlo simulations.



Each dot in the clouds represents an individual MC simulation with random value sets for input parameters. The four panels represent the following situations: a) equivalence between the efficacy of two technologies (symmetry with respect to the vertical axis); b) the new technology is not cost-effective compared to the benchmark; c) the new technology is cost-effective compared to the benchmark as all points of the simulation are below the line of willingness to pay; d) the new technology is cost-effective compared to the alternative with a probability of approximately 0.8.

The analysis of the proportion of simulations below the cost-effectiveness line, or equivalently the number of simulations with NHB>0 over the total number of simulations can be used to express probability statements about the likelihood that the technology will be cost-effective. For example, in Figure 3.2 panel b) the totality of points lies above the cost-effectiveness line, meaning that the probability of the technology being cost-effective are null,

whereas in panel c) and d) the probability that the technology is cost-effective is respectively close to 1 (panel c) or approximately 0.8 (panel d).

For technologies which are at early stages of development, the methods and approaches used to develop CEA are in every way similar to the ones used for CEA of more mature technologies. However, as mentioned earlier, their production is challenged by a systemic lack of solid clinical and economic evidence. For example, cost-effectiveness models evaluating mature technologies, with a high rate of uptake in clinical practice can usually rely on a wealth of data including one or more individual randomized controlled studies, systematic reviews and meta-analyses, real-world observational studies providing long-term data on both clinical and economic outcomes or even previous HTA reports and CEA from other agencies. On the contrary, early CEA models, especially when undertaken in a pre-market stage can usually rely only on limited data, such as pivotal studies, with limited sample sizes, often single arm, and using surrogate clinical endpoints to demonstrate performance (but not effectiveness). In addition, there exists a structural uncertainty that relates to unpredictable future product and market developments (e.g., market entrance of new competing products, unforeseen failure of the device at further stages and need for iterative changes during first-in-man testing, interaction with the environment and learning curve etc.) which further challenge the work of CEA modellers.

The sparsity of data in early models poses additional challenges for PSA as stronger parametric assumptions are needed to characterise the uncertainty of parameters where little or no evidence is available. In the absence of data to inform parameter distributions, formal elicitation of expert judgement, also known as structured expert elicitation (SEE) can be used to characterise parameter uncertainty. Several methods and approaches have been developed to elicit distributions for parameters and aggregating the opinions from multiple experts into a unique functional form (128,129).

However, this exercise may not always be possible. Especially in very early models, experts may not yet have formed an opinion for some parameters or be able to quantify it. In some cases, elicitation might even be

inappropriate, e.g., in the case of the relative effectiveness of a device that has yet to be developed.

Eventually, poorly informed parametrization of probability distributions for unknown parameters may even create pseudo-certainty in model outputs (130) thus leading decision-makers to underestimate the consequences of uncertainty.

Section 3.4 and 3.5 introduce PBA as an alternative way of dealing with parameter uncertainty when evidence is sparse. A practical application of PBA in an early CEA model for a novel total artificial heart is then reported in section 4.2.

The lack of evidence in the early stages of development of a technology is unavoidable. Therefore, as discussed, the primary purpose of early CEA models is not so much to provide an estimate of the cost-effectiveness of the technology under consideration, but to provide information on its potential for marketability and to inform the next steps in the process of generating evidence. Therefore, extensive scenario analyses are often used for this scope.

For example, one common type of analysis which is often conducted in early CEA is the headroom analysis (54,131) exploring the maximum reimbursable price for a technology given its potential for improving patients' health as compared to the existing alternatives in clinical practice, and an exogenously fixed payers willingness to pay.

The method consists of the following steps (Figure 3.3):

1. assume that the device actually improves health (e.g., QALYs gained compared to available alternatives) (Step 1);
2. calculate the maximum cost that the NHS is willing to spend on such improvement. This is usually done by using an explicit cost-effectiveness threshold (for example in terms of cost per QALY gained) (Step 2);
3. subtract the cost of the device production (Step 3);
4. subtract any other costs falling on the national healthcare system or add any potential savings generated by the technology. These are the

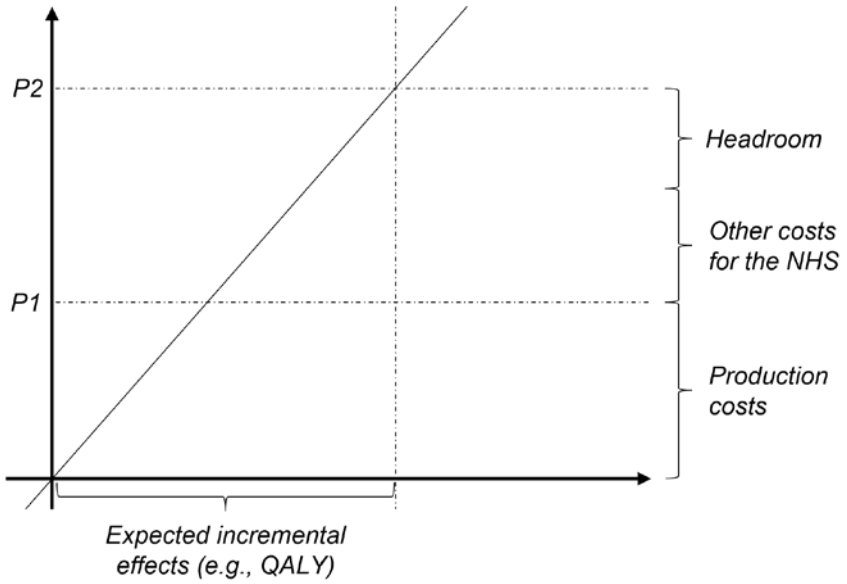
costs (savings) that would be generally considered in a CEA from the perspective of healthcare payers, besides the cost of the device cost (step4);

5. calculate the maximum profit margin for the product (step 5).

From the developer perspective, if the final gain presents a sufficient margin, to allow for example positive return on investments, then the development of the product could reasonably continue. Otherwise, further research investment would be better spent elsewhere.

From the healthcare payer perspective, setting the maximum price of an healthcare technology, so that the health benefits offered are at least equal to health displaced elsewhere in the NHS, is also called as Value-based pricing (VBP)(132). In fact, VBP prescribes that the maximum price for a technology is the one for which the NHB=0 (or the ICER is equal to the cost-effectiveness threshold).

Figure 3-2 Headroom approach



The figure represents a cost-effectiveness plane, and the line passing through the origin represents the cost-effectiveness threshold used by a decision maker to decide on adoption and reimbursement for a technology. The incremental costs of adopting the new technology are divided into the per-patient production costs for the manufacturer and the other costs accruing to the NHS. The difference between the maximum incremental cost for the

technology at the defined cost-effectiveness threshold and the estimated costs represents the maximum mark-up in the price of the technology that would be acceptable by the decision-maker.

In section 4.1 an application of the headroom approach has been used to estimate the maximum achievable price of the electrodes used in the procedure of electrochemotherapy as a treatment modality for stage IIIc and IV skin melanoma in Slovenia.

3.4. Probability Bound Analysis

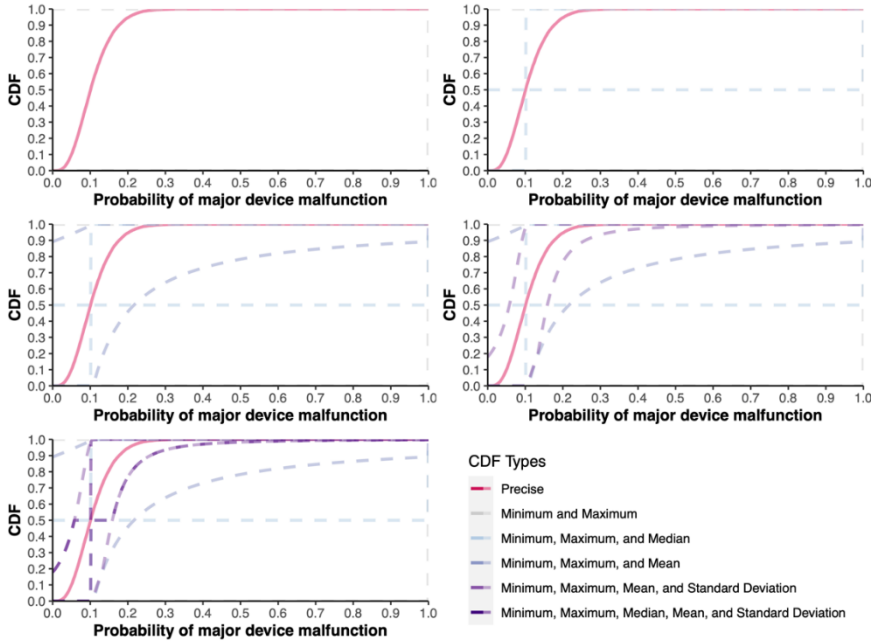
As mentioned, a full characterization of uncertainty through the definition of probability distributions on their value may not always be possible or appropriate.

The starting point of PBA is akin to that of PSA, i.e., model parameters are treated as random variables. However, PBA imposes a weaker restriction, i.e., the relative plausibility of parameter values cannot be defined in terms of precise probability distributions (133). PBA treats the cumulative distribution function (CDF) of a parameter as unknown and provides lower and upper bounds to the CDF instead. We define a p-box as a set of CDFs of a parameter θ ($F(\theta)$, i.e., $\mathcal{P}_{\mathcal{D}} = \left[\underline{F}(\theta), \overline{F}(\theta)\right]$, where each CDF belonging to the set is 1) defined on the support of θ, 2) bounded below and above by a lower-bounding function (LBF) $\underline{F}(\theta)$, and an upper-bounding

function (UBF) $\overline{F}(\theta)$, respectively, and 3) consistent with a given minimal data $\mathcal{D}$ (134,135),

We say that a CDF of θ is consistent with $\mathcal{D}$ if each element in $\mathcal{D}$ can be

Figure 3-3 P-boxes for a parameter under different minimal data and the true unknown cumulative distribution function of the parameter.



equated to a statistic that can be computed from the CDF. For each realization of θ, e.g., $\theta = \widehat{\theta}$, the epistemic uncertainty is represented by the interval $\left[\underline{F}(\widehat{\theta}), \overline{F}(\widehat{\theta})\right]$. In other words, we can only say that the value of $F(\widehat{\theta})$ is known up to an interval due our lack of knowledge about the precise form of the CDF. As a corollary, the more information about the CDF we have, the tighter the interval enclosing the unknown CDF becomes (Figure 3.3).

In fact, if the CDF is fully known, the interval degenerates to a single point, i.e., the p-box coincides with the CDF. Based on this observation, the p-box is dependent on the available data.. Details on the derivations of the formulas, are provided elsewhere (13,133)

After specifying the p-box for each parameter, the parameter uncertainties represented by the p-boxes are propagated into a CEA model.. Since each realization of a parameter corresponds to an interval, the realization of all parameters will be in the form of a Cartesian product of all intervals. The uncertainty propagation becomes a propagation of intervals of the model inputs into the model to derive the corresponding interval of the model output. The propagation is implemented computationally as an optimization problem: find the minimum and maximum of the model over the model parameter space defined by a particular product of intervals. The minimum and maximum values of the model define the interval of the model outcome. The optimization is then repeated over different spaces defined by other products of intervals to generate a p-box of the model outcome, i.e., $\left[\underline{F}(Y), \overline{F}(Y)\right]$.

Note that, for the interpretation of uncertainty in outcomes, the calculation of the expected value of model outcome Y over its p-box results in an interval of expected values. The interval includes all expected values that correspond to CDFs of enclosed by the p-box. This is true because the p-box of Y is guaranteed to enclose all CDFs of Y. Given what is known and assumed about the uncertainty of the model parameters, the expected outcomes cannot be larger (smaller) than expected value of Y over its upper (lower) interval bound. In contrast to assuming precise CDFs for all parameters where the expected value of a decision-relevant outcome is a single value, alternative decision rules are needed as expected value maximization on the basis of the von Neumann–Morgenstern (VNM) utility theorem is no longer amenable. One compatible decision rule with PBA is the Hurwicz decision criterion (136). Briefly, Hurwicz argued that the decision-maker can take uncertain decisions by ranking alternatives according to a weighted average between the worst- and best-case scenario of a choice.

$$O_j = \rho * O_{worst} + (1 - \rho) * O_{best}$$

where O_j is the Hurwicz's criterion for alternative *j*, $\rho \in [0,1]$ is the coefficient of pessimism, O_{worst} represent the worst possible consequence under alternative *j,* and O_{best} represents the maximum payoffs that can be achieved. The Hurwicz's criterion is eventually a weighted average of the maximin criterion, for which decisions should be taken with the highest security, by choosing on the basis of lowest possible payoffs for each alternative; and the maximax criterion, which chooses on the basis of the highest possible payoffs.

In the Hurwicz's criterion, the pessimism coefficient ρ is close to 0 for extreme optimists, risk-prone decision makers, whereas it tends to 1 for pessimists, risk-averse decision makers. It can be noted that Hurwicz's decision criterion can be used alongside PBA where the estimated lower and upper boundaries output of the model correspond to the best- and worst-case scenarios, respectively.

3.5. Implementing PBA

This section describes the steps for implementing PBA in a health economic model and closely follows the steps outlined in Iskandar et al. (133).

S1: The model parameters that are suitable for PBA are identified, comprising the set θ_p.

S2: For each PBA parameter in θ_p ($\theta_i \in \theta_p$) the type of data available $\mathcal{D}_i$ is determined (minimal data) where i indexes the model parameter.

S3: For each θ_i, interval $[0,1]$ (the probability space) is portioned into a finite number of intervals (sub-intervals), where each parameter may have different number of sub-intervals, n_i. The finer the discretization is (higher n_i), the higher the accuracy of the approximation of the PBA in the model outcome will be. The optimal choice of n_i is problem dependent. Practitioners should test and compare different choices of n_i in terms of the approximation accuracy (smoothness) of the resulting p-box. We assign a

weight to each sub-interval w_{i_k}, where i_k indexes the k-th sub-interval of θ_i . A typical weight is a uniform weight for all sub-intervals, i.e., $w_{i_k} = n_i^{-1}$.

S4: For each sub-interval of a parameter, it is then necessary to determine the lower and upper endpoints and calculate their corresponding upper and lower endpoints in the parameter space (which is similar to the inverse transform sampling approach) using the quasi-inverse of the LBF and UBF, conditional on $\mathcal{D}_i$. This step generates the sub-intervals in the parameter space.

S5: Given the sub-intervals, all possible combinations are enumerated, i.e., each combination is formed by intersecting a particular sub-interval for each parameter. The number of combinations is determined by the number of sub-intervals of each parameter.

S6: For each combination of sub-intervals, the minimum and maximum of the model outcome is computed using an optimization method over the parameter space defined by the combination. This step produces intervals constructed from the minimum and maximum values of Y and their associated weights. The weights are calculated by multiplying the weights of the sub-intervals included in the combination. The optimal choice of the optimization method is problem-dependent and typically determined by, for example, the model characteristics such as non-linearity and convexity of the model and the gradient information, or the trade-off between computational budget and accuracy.

S7: To derive the p-box of Y, two empirical CDFs are then calculated by cumulating the weights for the values of the minimum and the maximum are less than some arbitrary values in the support of Y.

In some cases, there may be information that is sufficient for specifying probability distributions of some parameters, comprising the set θ_c. To propagate uncertainty from both parameter sets θ_c and θ_b into $\mathcal{M}$, the following steps are required.

S0: *N* repeated samplings from the precise CDFs of parameters in θ_c are drawn to generate a sequence of *N* samples. For each sample of θ_c, steps **S2** to **S6** are repeated.

S8: The average p-box of Y is calculated by averaging p-boxes over the *N* samples.

3.6. Value of information analysis

As discussed in section 2.5, healthcare decisions about the adoption and reimbursement of a technology are made with uncertainty. In fact, due to imperfect knowledge there is a non-negligible risk of taking sub-optimal decisions. Suboptimal decisions can have negative consequences both in terms of direct adverse health effects to patients and in terms of population health losses should the health opportunity costs of the budget constraint be higher than the health gains generated by the new intervention. Acquiring more information may reduce the risks associated with uncertain decisions and the associated consequences on population health. In the context of model-based CEA, VOI can be directly calculated from the output of the PSA.

The initial assumption for VOI is that decision-makers need to take a decision over a set of *n* mutually exclusive alternatives $\mathcal{D} = \{d_1, d_2, \dots, d_n\}$. The output of a CE model, indicated as $\mathcal{U}\{d, \theta\}$ expresses the utility of each decision d given a set of uncertain model parameters *θ,* with joint parameter distribution $p(\theta)$. The utility functions are usually expressed as the net-benefits of the different interventions.

The expected value of reducing the uncertainty over all model parameters θ has been defined as the Expected Value of Perfect Information (EVPI). The EVPI represents the maximum benefit, in terms of reduced uncertainty that can be achieved by doing further research. In fact, no matter the number and sample size of the additional studies that are (planned to be) conducted, it is highly unlikely that *all* existing parameter uncertainty can be resolved.

Therefore, the EVPI represents a theoretical limit for the benefit of generating further evidence. Nonetheless, because of its easiness of computation, EVPI is often used to gain quick insights on whether further research should be warranted.

Once the PSA of a CEA is conducted through MCMC sampling, the EVPI can be calculated as follows:

$$EVPI(\theta) = \mathbb{E}_{\theta}\{\max_{d\in\mathcal{D}} \mathcal{U}(d,\theta)\} - \max_{d\in\mathcal{D}} \mathbb{E}_{\theta}\{\mathcal{U}(d,\theta)\}$$

The second argument in the formula represents the expected payoffs of the decision on the basis of the current knowledge (and uncertainty). The first argument represents the expected value of the decision in the case of perfect information, i.e., in the case where, for each Monte Carlo simulation, the alternative with the highest payoff was always chosen. Since it is not known in advance how the existing uncertainty will resolve, it is then necessary to calculate the expected value of the max payoffs achievable in each MC simulation.

The calculation of the Expected Value of Partial Perfect Information (EVPPI) and Expected Value of Sampling Information (EVSI) are simple extension to the EVPI formula which account for the fact that only part of the existing uncertainty can be reduced.

Similar to the EVPI, the EVPPI is calculated by comparing the expected value of the utility function with reduced uncertainty, to the expected value of the utility function with the actual knowledge. However, in this case it is assumed that the uncertainty is fully resolved over a subset $\theta_i \in \theta$ of the parameter of interests. The formula to estimate the EVPPI is therefore:

$$EVPPI(\theta_i) = \mathbb{E}_{\theta_i}[\max_{d\in\mathcal{D}} \mathbb{E}_{\theta_c|\theta_i}\{\mathcal{U}(d,\theta)\}] - \max_{d\in\mathcal{D}} \mathbb{E}_{\theta}\{\mathcal{U}(d,\theta)\}$$

As can be noted in the formula above, in the left argument of the EVPPI there is a double nested expectation. The inner expectation calculates the expected value of the utility function for different values of the complementary (uncertain) parameters θ_c conditional on a given set of

values for the parameters θ_i (i.e., assuming perfect information for these parameters). Once the first expectation is done, in a similar way as for the EVPI, the alternative with the maximum payoff value is chosen. However, since the true value of θ_i is unknown, a second expectation is required over the sample of possible values of θ_i.

Because of this two-level expectation, calculation of the EVPPI is much more computationally demanding compared to the EVPI as it requires a nested MCMC simulation.

The logic behind the EVSI is analogous to the EVPI and EVPPI, however, the EVSI considers the fact that a new study can only partially resolve the uncertainty over one or a set of parameters.

Nonetheless, a new study may provide data that can be used to improve decision making. In fact, the EVSI assumes that, should new data X become available, the decision maker would choose the option that maximizes the utility, conditional on knowing X. The formula for EVSI is in all respect like the one for EVPPI but in this case, all parameters remain uncertain to a certain degree and the inner expectation is calculated after updating parameter information with the data X provided by a new (future) study.

$$EVSI = \mathbb{E}_X[\max_{d\in\mathcal{D}} \mathbb{E}_{\theta|X}\{\mathcal{U}(d,\theta)\}] - \max_{d\in\mathcal{D}} \mathbb{E}_\theta\{\mathcal{U}(d,\theta)\}$$

The calculation of the EVSI is also highly computationally demanding, and therefore its use has been very limited to inform policy making on healthcare. However, as mentioned in section 2.4, recent methodological developments have partly addressed these challenges by easing the computational burden of VOI using regression-based approaches to estimate EVPPI and EVSI (61)

All measures of VOI, can be multiplied to the population that will benefit from the reduced uncertainty to calculate the maximum societal benefit. This is calculated as follow:

$$population\ VOI = Individual\ VOI * \sum_{t=0}^{T} \frac{I_t}{(1+d)^t}$$

Where I_t is the incident population in each period *t* of time, *T* is the maximum time for which the benefit of information will be relevant (e.g., until the technology under assessment becomes obsolete, and the results of the underlying CEA are not relevant anymore), and *d* is a discount rate for each time period.

To date, most of VOI studies has adopted the perspective of healthcare payers to decide whether to invest in research. As argued by Breeze and Brennan (137), this perspective assumes that "*research will be financed by public resources. The decision to adopt an intervention now (i.e. before the research study) is uncertain because it is possible that the new intervention is less cost-effective than the alternative, and this is described as the expected opportunity cost. The opportunity cost is typically quantified as the additional investment needed and the expected quality-adjusted life years (QALYs) forgone, valued at the health care provider's willingness to pay for a QALY.*"

This perspective is not directly applicable to technology developers who have different objectives compared to healthcare payers. In fact, the potential role and value of further evidence for manufacturers is the results of more complex interactions and dynamics that need to be considered. When entering a market, manufacturers face the uncertainty that, even after obtaining market access, decision-makers, such as payers or healthcare providers may hamper or even prevent adoption and uptake of their technology if the evidence around it is not considered robust enough. In addition, at different decisional levels, both the assessment criteria and the decision rules may vary. Therefore, what constitutes sufficient evidence for some, may be less relevant for others. For example, payers and HTA bodies may decide based on the cost-effectiveness of the technology, using the wider perspective of the healthcare system and considering long-term costs and consequences of adopting it. Conversely, providers may be more interested in knowing whether a technology reaches a minimum clinical difference and whether the short time budget impact of adopting it remains under a certain threshold, at least in the short-term. Therefore, soon after market approval, manufacturers usually start building their value dossier,

planning what evidence should be generated to maximize the company's objectives. In fact, while deciding their post-market plan for evidence generation, manufacturers need to weigh the cost of conducting further studies with their expected returns. In doing so, they need to estimate the expected payoffs with and without further evidence according to a specified utility function. Generally speaking, the value of conducting further research for manufacturers will be affected both by the behaviours of healthcare decision-makers following a reduction in the uncertainty, and other aspects of their utility function such as the company operating margin, or the market dynamics, like for example the time before manufacturers expect that their technology will become obsolete.

Therefore, since both the utility function and the factors affecting its maximization are different compared to payers, manufacturers may have difference perceptions of what evidence constitutes good value for money. This in turn may lead to potential misalignment in the value of further evidence among healthcare payers and technology developers.

The relevance and magnitude of these misalignment are explored in Section 4.2 by means of a case study on a total artificial heart for patients with advanced biventricular heart failure.

Chapter 4
Case studies on cost-effectiveness models

This chapter presents a series of case studies in which different aspects of uncertainty analysis are addressed. Section 4.1 reports on an early CEA model on electrochemotherapy for patients with Stage IIIc/IV skin melanoma where the headroom approach has been applied to identify the maximum achievable price of the devises for electroporation. Section 4.2 present a CEA model on a novel artificial heart for patients with advanced hear failure. This case study was used to apply Probability Bound analysis and to explore the misalignment of the perceived value of information between healthcare payers and technology developers using Value of Information analysis.

4.1. Early cost-effectiveness of electrochemotherapy to treat patients with stage IIIc and IV skin melanoma

Motivation

This case study focuses on electrochemotherapy (ECT) for treating patients with stage IIIc and IV skin melanoma. The case is relevant as ECT is in the early stages of diffusion with little evidence available on its cost-effectiveness. At this stage of a technology lifecycle, CEA can inform the optimal or maximum achievable price for a technology given the available evidence on the economic and clinical consequences of the technology and the payers willingness to pay. As described in Section 3.3, from a developer perspective this type of analysis is called headroom analysis. Usually, the headroom analysis is performed in a deterministic fashion using point estimates on the expected cost-effectiveness of the device under assessment. Here the whole uncertainty over the input parameters in the model was characterized

through the PSA and results of the headroom approach are provided in terms of probabilistic statements on the likelihood of the technology being cost-effective given a fixed cost-effectiveness threshold. From both the developer and healthcare payer perspectives, especially when evidence is still sparse, the information provided from the headroom approach is highly relevant to inform the development of the value dossier (for the former) and decisions on clinical guidelines and reimbursability (for the latter). From the healthcare payer perspective, this way of setting prices based on the maximum achievable price for the technology to be cost-effective is also known as value-based pricing. Having insights on the value-based price is particularly relevant for devices where the pricing mechanisms are often more complex and linked with the way devices are procured for example at the hospital levels through public tenders, managed often at the subnational level.

The case study is structured as follows: i) first the disease problem, and the technology under assessment are presented, ii) second the methods and data sources are presented followed by a description of iii) materials and methods, iv) results and v) discussion.

Disease problem and the technology under assessment

Electroporation is a phenomenon by which the transport of otherwise impermeant molecules through the cell membrane is facilitated (138–140). Electroporation is becoming increasingly recognized in medicine (141–143) but also in food technology and biotechnology (144–146). In medicine, electroporation is used for the treatment of solid tumors, either in combination with chemotherapy (electrochemotherapy) or alone (irreversible electroporation). Electrochemotherapy (ECT) is a local antitumor therapy that increases the toxicity of chemotherapeutic drugs bleomycin or cisplatin (141). Nonthermal irreversible electroporation (IRE) enables the ablation of undesirable (malignant) tissue, with minimal damage to blood vessels and nerves (147). Additionally, electroporation is also a promising delivery method for the introduction of genetic material, and for

DNA vaccination (148). Published studies demonstrated that the electric field established by applied high voltage, short-duration electric pulses, increases the plasma membrane permeability (139). Electroporation can be performed as a reversible or irreversible electroporation. In the case of reversible electroporation, the cells fully recover after electric pulse application, while in case of irreversible electroporation after the pulse application cells die, due to the loss of cell homeostasis. The device that generates and enables the delivery (via the application-specific electrodes) of electroporation pulses to biological tissue, is named an electroporator (149,150).

With the development of the electroporation field, new medical therapies, new clinical electroporators and innovative delivery systems, questions regarding cost-effectiveness arise. In fact, while the technology is entering into national and international guidelines (151), its value for money remains largely unexplored in mostly all settings. Indeed, only one cost-effectiveness analysis has been published to date (152), and due to the lack of information available, especially about the quality of life impact, results are incomplete (153). The UK National Institute for Health and Care Excellence (NICE) deemed electrochemotherapy is a safe treatment for primary basal cell carcinoma and primary squamous cell carcinoma, however, it also warned about the limited evidence around its efficacy (154). The recent study by Clover et al, has provided further evidence of electrochemotherapy on basal cell carcinoma in relation to surgery, in a randomized study (155).

Even though electrochemotherapy treatment of skin melanoma and basal cell carcinoma has been accepted in Slovenian clinical practice, evidence on its clinical effectiveness and cost-effectiveness in this indication is still sparse. Therefore, there is a need to discuss whether the use of ECT in this patient population is cost-effective for the Slovenian healthcare system, and whether the overall cost of the procedure is acceptable. Based on the results of a real-world small clinical study, the aim of this case study was to perform an early economic evaluation of electrochemotherapy, with the Cliniporator™ (IGEA S.p.A., Capri, Italy) as a treatment modality for stage IIIc and IV skin

melanoma in Slovenia. Electrochemotherapy was compared to the standard of care (SoC) consisting of symptomatic therapy and palliative treatment.

Materials and methods

Patients

Twenty-three patients were enrolled and treated with electrochemotherapy at the Institute of Oncology Ljubljana in Slovenia between June 2014 and March 2019. Patients were treated according to Slovenian Recommendations for the treatment of patients with cutaneous melanoma and standard operating procedures for electrochemotherapy (156). Inclusion criteria were:

- cutaneous melanoma metastases, which are symptomatic due to bleeding, ulceration, oozing, odour, or pain;
- progression of cutaneous metastases, where development of symptoms as listed above, is expected; primary skin cancers, including recurrent tumours, where other treatment modalities (surgery, radiotherapy, and systemic therapies) have failed or are not possible;
- patients who are receiving systemic therapy, but where cutaneous metastases are progressing or not responding despite satisfactory response to systemic therapy in internal organs;
- patient preference for electrochemotherapy, after other treatment possibilities have been thoroughly explained to the patient.

Exclusion criteria for electrochemotherapy was one of the following:pregnancy, lactation, allergy or hypersensitivity to bleomycin or lifetime dose of bleomycin exceed 400,000 International Units (IU). At the time of writing eight of the patients were still ongoing patients while 15 died or refused further treatment. The average patient age at the first electrochemotherapy procedure was 78.1, with a standard deviation of 12.3 years, range from 48 to 96 years. Patients were diagnosed with stage IIIc or IV malignant melanoma and were not amenable to other treatments. Patients' characteristics are presented in Supplementary material S3.1. In

Slovenian guidelines for the treatment of skin melanoma, electrochemotherapy is indicated as a treatment modality for recurrence at the extremities where simple excision is not possible (> 3–5 metastases) or for recurrent relapses (sooner than 3–6 months) (157). Patients were treated with electrochemotherapy in accordance with standard operating procedure with the Cliniporator™ device (156,158). Follow up examinations were conducted 14 days after the intervention and after 1, 2, 4, 8 and 12 months. Altogether 38 electrochemotherapy procedures were performed, meaning on average each patient was treated 1.6 times, range 1 to 5 procedures per patient. At each electrochemotherapy procedure, all skin lesions present at the time of procedure were treated. Electrochemotherapy was repeated when new lesions were presented or when only a partial response was obtained.

Cost-effectiveness analysis

A cost-effectiveness analysis was conducted to estimate the lifetime costs and consequences of using electrochemotherapy in the target patient population from the perspective of the Slovenian health care system. The model used a time horizon of 10 years, which was considered sufficient to capture the life time costs and consequences for these patients. All future costs and consequences were discounted at a 3.5% discount rate.

Patients included in the study are mostly elderly with stage IIIc and IV skin melanoma and are generally not amendable to any other treatment option. Therefore, in the present study, electrochemotherapy was compared to the (SoC) consisting of symptomatic therapy and palliative treatment. The cost-effectiveness of electrochemotherapy compared to standard of care was expressed in terms of incremental net health benefit (iNHB) which is calculated based on:

$$iNHB_{ECT\ vs\ SOC} = \Delta QALY - \frac{\Delta Cost}{k}$$

Where $\Delta QALY$ and $\Delta Cost$ are the differences in the expected QALYs and costs between electrochemotherapy and standard of care, and k is a constant

value representing the payers cost-effectiveness threshold (159). Electrochemotherapy is then considered cost-effective for any value of the iNHB greater than 0. The value of the cost-effectiveness threshold for Slovenia was assumed to be equal to the 2018 Slovenian real Gross Domestic Product (GDP) per capita, and set equal to €20,000.00 per QALY (160). Other thresholds have been suggested in the literature for Slovenia which are lower compared to the one used here (161). The impact of using different thresholds is discussed in the discussion.

The CEA analysis was conducted separately for all patients and for the subset of patients with bleeding lesion, because of the higher procedure costs and higher improvement in the quality of life was expected.

A discrete-time Markov model was used to model patients' lifetime costs and consequences of either electrochemotherapy or standard of care. A previous Markov model used for skin melanoma (153) was modified to better fit the disease progression of patients with stage IIIc and IV skin melanoma after electrochemotherapy. The model used in the study is presented in Figure 4.1. In the initial state, patients have predominately cutaneous symptoms and are treated with electrochemotherapy, they can either not respond (i.e., they remain in the same state) or move to a response state if they experience complete or partial response (response to electrochemotherapy). Patients then remain in the response state unless they develop new metastases, then they are moving to a relapse state, where they can again receive repeated electrochemotherapy. Patients in the initial state were also allowed to move to the relapse state in case they worsened their health status by developing new methastases. The cycle length used in the model is two months.

Bimonthly transitions probabilities between states were directly derived from fully observed patient-level data collected during the study. First, the total person-months of exposure in each state and the number of transitions to any other state were used to calculate a 4 by 4 transition-rate matrix Q in a Bayesian framework using the data from the study and uninformative prior distributions. The model was run using the OpenBugs Markov Chain Monte Carlo (MCMC) software (162) (Supplementary material S4.1). In the model,

an initial run of 10,000 iterations was considered as 'burn in' (these values were discarded). Subsequently, two independent chains, starting from randomly assigned values were run, and convergence was monitored by looking at the ratio of the within-chain to between-chain variance to be about one, and by using Heidelberger-Welch (163) and Gelman-Rubin (164) diagnostics. Second, the transition-probability matrix *P(t)* was estimated by taking the matrix exponential $P(t) = Exp(Qt)$ using the *expm* package in Rstudio (165).

Figure 4-1 Four state Markov model of skin melanoma, adopted for electrochemotherapy treatment of stage IIIc and IV melanoma.

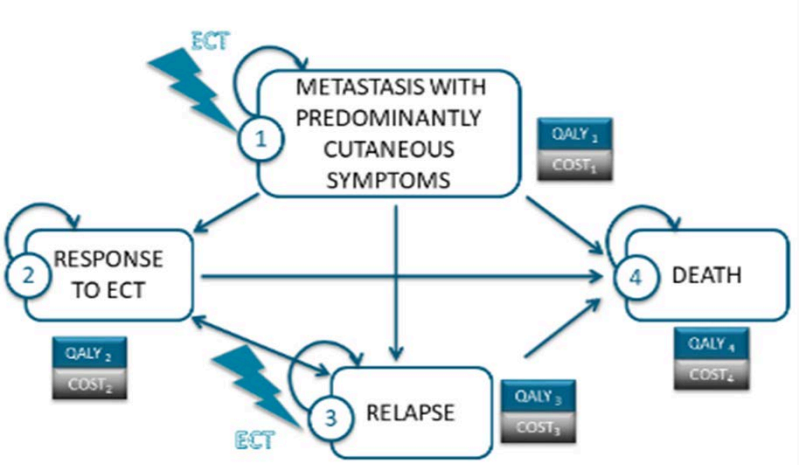


Table 4-1 Two-month transition probabilities for patients receiving electrochemotherapy

	Mean	95% credible interval
Electrochemotherapy Group		
From State 1 to:		
State 1	0.15	0.06 - 0.28
State 2	0.61	0.48 - 0.72

State 3	0.11	0.07 - 0.16
State 4	0.13	0.06 - 0.26
From State 2 to:		
State 1	0	-
State 2	0.73	0.64 - 0.81
State 3	0.20	0.13 - 0.27
State 4	0.08	0.04 - 0.13
From State 3 to:		
State 1	0	-
State 2	0.24	0.13 - 0.36
State 3	0.54	0.39 - 0.68
State 4	0.22	0.11 - 0.36
Standard of care group*		
From State 1 to:		
State 4	0.22	0.11 – 0.36

** In the absence of electrochemotherapy, since standard of care only provides palliative care, it was assumed that patients would remain in state 1 until they die.*

In the absence of electrochemotherapy, since SoC provides only palliative care, patients were assumed to remain in state 1 until they die, with probability, which was assumed to be equal to the one of relapsed patients in the electrochemotherapy group (state 3 to state 4). This assumption was discussed and validated with expert clinicians at the Institute of Oncology, Ljubljana. Lastly, transition probabilities were assumed to be the same for both the whole sample and the subgroup of patients with bleeding lesions.

The estimated transition probabilities and their credible intervals for electrochemotherapy are provided in Table 4-1.

The average cost and standard deviation of a single electrochemotherapy procedure were estimated using patient-specific data on resource consumption during electrochemotherapy procedure and subsequent follow-ups and attaching the corresponding unit costs. All the costs are presented in EURO and valid for Slovenia. Costs related to the hospital and procedure were obtained from the Institute of Oncology, Ljubljana, whereas IGEA S.p.A. (the manufacturer of the Cliniporator™ device) provided prices related to the medical device.

Table 4-2 Costs included in the evaluation of the cost of electrochemotherapy in Slovenia

Cost item	Cost per unit [€]
Overnight stay in the hospital (one night after electrochemotherapy + one day)	240.00
Price of intervention and cost of staff (average duration: 45 minutes to1 hour)	
Staff one hour of procedure	128.25
Supplies for personnel and venue	66.13
Consumables during operation	99.25
Depreciation of apparatus in an operating room	22.57
Chemotherapeutic drug	
Bleomycin vial	30.00
Cisplatin bottle	23.00
Anesthesia	
General anesthesia	225.88
Local anesthesia	4.40
Sedation	14.60
Spinal block	15.90
Electrodes EPS series	1,200.00

The overall costs of electrochemotherapy is presented in Table 4-2. The average cost of single electrochemotherapy procedure was estimated to €2,757.00 with the standard deviation €707.30. In addition, the Institute of Oncology also provided an exact expense for eight electrochemotherapy procedures, and these data was used to validate the costs figures reported in Table 4-2. Overall, the difference in the mean costs using both data sources was in the range of 10%.

General anaesthesia was used in 25 out of 38 treatments. On average 1.5 to 2 bleomycin vials and one cisplatin bottle per patient was used, and the cost of bleomycin was thus set to €52.50 per procedure. The cost of cisplatin is less but was only used in two electrochemotherapy procedures. Electrodes represent almost half of the price for electrochemotherapy procedure and new versions of electrodes (EPSA series) are even more expensive, with a price of €1,600.00 per single electrode (VAT excluded). All electrochemotherapy electrodes manufactured by IGEA S.p.A are for single

use, meaning one electrode can be used for one patient, but for multiple metastases, however, in some cases more than one electrode geometry is used due to the difference in metastases. In the scope of our study 1.19 electrodes per patient were used, in range from 1 to 2 per procedure. Also, other patient-specific costs were analysed and evaluated, such as analgesics or antibiotics, but because these costs did not exceed €10.00 per treatment, they were neglected in the analysis. Follow-up specialist visits were conducted after 1, 2, 4, 8 and 12 months following the intervention with an estimated cost of €22.50 per examination.

Cliniporator™ device, manufactured by IGEA S.p.A. is essential part to perform electrochemotherapy. The price of the device Cliniporator™ EPS02 is €100,000.00 (VAT excluded). The maintenance cost is €3,000.00 and the maintenance is due every 24 or 36 months according to the country specific (e.g. Germany: 24 months, while Italy, U.K. and Slovenia: 36 months). The Cliniporator™ device is considered as a highly stable device, therefore IGEA as a manufacturer requires the maintenance only every 36 months. The device lifetime is specified as 500 treatment sessions or 10 years (according to user manual).

Following an extensive literature review, we could not identify baseline cost data for skin melanoma in the Slovenian setting. However, an extensive cost of illness study is available for Croatia (166). Because Croatia and Slovenia are neighboring countries and not long ago, were even both part of the same country, data collected for Croatia were considered in this study. For the calculations, the average cost per single patient for stage IV melanoma was set to €4,333.00 per year. This value is also in accordance with the European average (167) and close to value available for Italy (168).

Table 4-3 Costs and QALYs values and credible intervals used in the model (1-year values)

	Mean	**95% Credible interval**
Cost state 1		
All patients	€4,333.00	€4,139.00 – €4,533.00

Patients with bleeding lesions	€7,784.00	€7,586 .00–€ 7,978.00
Cost state 2	**€4,333.00**	**€4,139.00 – €4,533.00**
Cost state 3	**€4,333.00**	**€4,139.00 – €4,533.00**
Cost of ECT	**€2,757.00**	**€2,095.00 – €3,690.00**
QALY state 1		
All patients	0.66	0.54 - 0.74
Patients with bleeding lesions	0.40	0.23 - 0.58
QALY state 2	**0.72**	**0.66 - 0.80**
QALY state 3		
All patients	0.66	0.54 - 0.74
Patients with bleeding lesions	0.40	0.23 - 0.58

With the help of Slovenian palliative experts, we evaluated the worst-case cost of care for a patient with bleeding lesions to be €3,450.00 per patient per year (includes only the care of bleeding wound). This cost can be completely eliminated after the electrochemotherapy procedure, however, only 5 % of melanoma patients develop bleeding lesions (169). Patients were given an EQ-5D-3L questionnaire at each examination, and the following results, which present health utilities, were obtained: before the procedure the average quality of life was equal to 0.66, the patient that responded to electrochemotherapy had an increase in quality of life to 0.72 and in case of relapse of metastases, the quality of life decreased to 0.66. A significant increase in quality of life was expected only in patients with bleeding nodes, which after electrochemotherapy gain the most, however, only 5 % of all melanoma patients have bleeding lesions, meaning 1.15 patients in our study. The quality of life for patient with bleeding lesions was in initial state (state 1 in Figure 4.1) thus reduced to minimal obtained quality of life, which was equal to 0.4.

The baseline mean estimates and 95% credible intervals for costs and QALY data used in the model are reported in Table 4-3.

The face validity of the model has been verified by discussing both the model structure and the data sources with the expert clinicians that collected the primary data in Slovenia. The internal validity of the model has been checked by performing extensive analysis of the model behaviour in extreme scenarios.

Scenario and sensitivity analysis

Probabilistic sensitivity analysis (PSA) was performed by assigning probability distributions to all parameters used in the cost-effectiveness model. For transition probabilities, samples were taken directly from the joint posterior distribution of the transition probability matrix, calculated with the MCMC simulation in OpenBugs. For costs and QALY data, samples were derived from Gamma and Beta distributions respectively, which were previously characterised using mean and standard error estimates from the study data and the literature. The results of the probabilistic sensitivity analysis are then reported in a cost-effectiveness plane and used to calculate cost-effectiveness acceptability curves for electrochemotherapy compared to standard of care.

In addition, the probability of electrochemotherapy being cost-effective was also estimated as a function of the cost of the electrodes used for each electrochemotherapy procedure in both patient subgroups, because the cost of one electrode geometry represent almost half of the estimated electrochemotherapy procedure cost. This analysis is a modified version of the headroom analysis and it presents the probability of the intervention of being cost-effective at a certain cost-effectiveness threshold and different prices of the device. As mentioned, in this case study a probabilistic headroom analysis was conducted where rather than establishing a maximum price achievable, different prices are assessed in their impact on the likelihood for electroporation to be cost-effective.

Lastly, since electrochemotherapy is usually considered to be a outpatient procedure, not requiring hospitalization, the results of a scenario analysis are reported, where electrochemotherapy procedures were assumed to be provided in an outpatient care setting, without any hospitalization costs.

Results

In the whole sample, electrochemotherapy is expected to improve QALYs by 0.29 QALYs (95 % CrI 0.10, 0.50) (Table 4-4) over patients' lifetime, at an increased cost of €6,568.00 (95 % CrI €4,593.00, €8,928.00). The expected iNHB of electrochemotherapy compared to standard of care is equal -0.04 QALYs (95 % CrI -0.19, 0.11), meaning that at the used cost-effectiveness threshold of €20,000.00 per QALY gained, electrochemotherapy is slightly less cost-effective compared to standard of care, although the uncertainty over this estimate is quite large.

For the subgroup of patients with bleeding lesions, electrochemotherapy is expected to yield a higher quality of life by 0.34 QALYs (95 % CrI 0.18, 0.56) (Table 4-4) at a higher cost of €4,863.00 (95 % CrI 2,479.00, 7,177.00). Compared to the whole sample, providing electrochemotherapy only to patients with bleeding lesions is expected to be more cost-effective with an expected iNHB of 0.11 (95 % CrI -0.06, 0.27), Although uncertainty remains high.

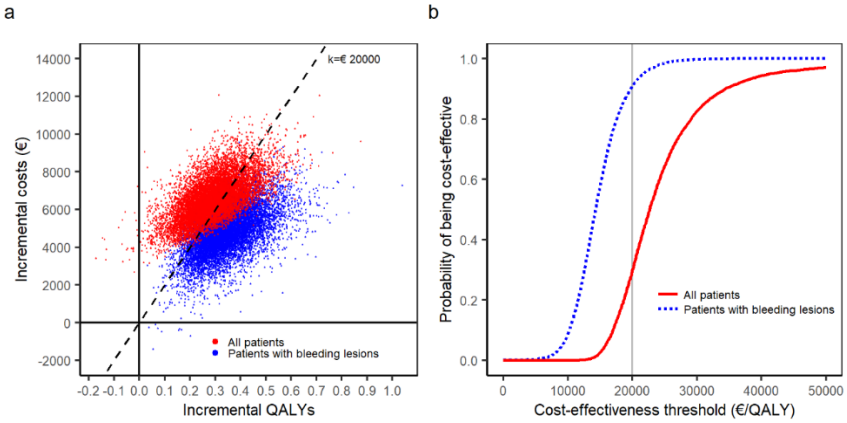
Table 4-4 Results table, presenting QALY, costs and NHB for each patients group

	ECT		**Standard of care**		**Incremental Results**	
	All patients	**Patients with bleeding lesions**	**All patients**	**Patients with bleeding lesions**	**All patients**	**Patients with bleeding lesions**
QALYs	0.74	0.62	0.45	0.27	0.29	0.35
Costs [€]	9,539	10,199	2,971	5,336	6,567	4,863
NHB	0.26	0.11	0.30	0	-0.04	0.11

QALYs, Quality Adjusted Life Years; NHB, Net Health Benefit

The results of the probabilistic sensitivity analysis show a considerable uncertainty on the incremental costs and QALYs of electrochemotherapy for both patient groups with the simulated costs and QALYs pairs being spread widely in the cost-effectiveness plane.

Figure 4-2 Cost-effectiveness plane and cost effectiveness acceptability curves of electrochemotherapy versus standard of care.



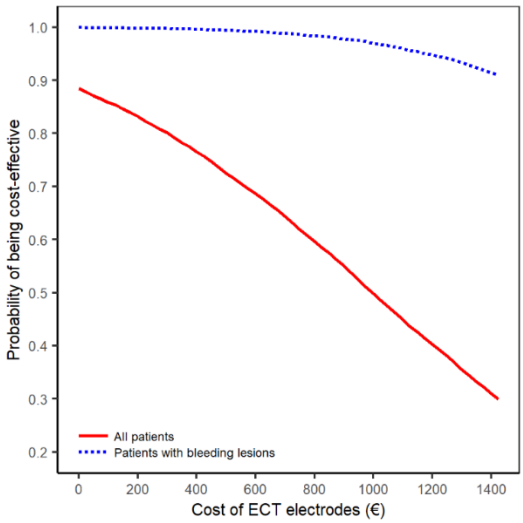
(a) Cost-effectiveness plane; (b) The probability of electrochemotherapy being cost-effective for all patients and patients with bleeding lesions.

However, it must be noted that most of the simulations remain in the first quadrant of the cost-effectiveness plane, meaning that electrochemotherapy is highly likely to be more effective and more costly compared to standard of care (Figure 4.2, panel a). The probability of electrochemotherapy being cost-effective is estimated to be 30% for the whole sample, and 91% in patients with bleeding lesions (Figure 4.2, panel b)

A reduction in the cost of the electrodes used in the electrochemotherapy procedure is not going to greatly affect the probability of cost-effectiveness in patients with bleeding lesions, since in this patient population, electrochemotherapy is already highly likely to be the best treatment option even at the base case cost of the electrodes. However, the cost of the electrodes has a considerable impact when considering the whole patient sample (Figure 4.3). For example, a reduction by half in the average cost of

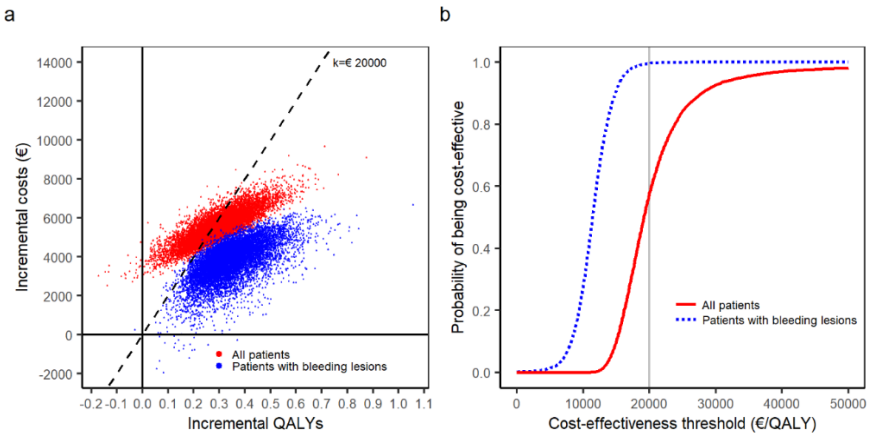
the electrodes used during the electrochemotherapy procedure would increase the probability of cost-effectiveness from 0.30 to approximately 0.64.

Figure 4-3 The probability of electrochemotherapy being cost-effective as a function of electrode cost for both patients groups



Finally, in the scenario without hospitalization costs, the expected incremental costs of electrochemotherapy were estimated to be 16 % lower compared to the base case analysis, which in turn resulted in a higher iNHB of 0.014 (95 % CrI - 0.12, 0.15) and a higher probability of being cost-effective equal to 0.58. (Figure 4.4), for the whole patient sample.

Figure 4-4 Cost-effectiveness plane and electroporation in case of zero hospitalization costs



(a) Cost-effectiveness plane, in case of elimination of hospitalization cost; (b) The probability of electrochemotherapy being cost-effective for all patients and patients with bleeding lesions, without hospitalization costs.

Discussion

This case study presents a cost-effectiveness analysis of electrochemotherapy based on early observational study in the Slovenian setting. To date no other cost-effective studies have been found in this patient population. Previous studies (170–176) have reported increases in quality of life after electrochemotherapy, however, quantitative information is often lacking or not sufficiently detailed to allow estimation of the cost-effectiveness of electrochemotherapy in patients with stage IIIc and IV skin melanoma. For example, even when quality of life estimates were reported, it was not possible to differentiate between cancer types and stages (177,178). The primary data collected in this study aims to fulfil this gap; however, the study has several limitations. First, by only collecting data on patients receiving electrochemotherapy, improvements in QALYs were measured using a before and after evaluations, which is prone to biases. Second, the estimation of the relative effectiveness of electrochemotherapy compared to standard of care is based on the assumption that patients not receiving treatment would remain with the same baseline utility values for the rest of their life, which may not be the case.

When considering all patients with stage IIIc and IV skin melanoma, electrochemotherapy is expected to be less cost-effective compared to palliative care and symptomatic treatment (iNHB -0.037 QALYs, with probability of being the most cost-effective strategy equal to 0.32). Conversely, electrochemotherapy is expected to be more cost-effective in patients with bleeding lesions as both the relative improvement in QALYs after successful electrochemotherapy and the expected savings in management costs are expected to be higher, although with a considerable uncertainty in the model estimates. It should be also noted that the NHB estimates are calculated using a cost-effectiveness threshold of €20,000.00 per QALY, which approximates the 2018 Slovenian GDP per capita. While

the use of thresholds based on GDP has been recommended by the WHO (160), other estimates have also been proposed in the literature. For example Woods et al. provide country specific values using empirical estimates of the threshold for the UK; estimates of the relationship between country GDP per capita and the value of a statistical life; and a series of explicit assumptions (161). For Slovenia the authors estimate a threshold in a range between $11,374.00 and $15,690.00 purchasing power parity (PPP) which correspond to a range in Euro between €6,710.00 and €9,257.00. It is expected that electrochemotherapy would not be considered cost-effective in any case at these lower thresholds.

Nevertheless, even at this higher threshold, the cost of electrodes remains a critical issue that may hinder a broader adoption of electrochemotherapy in clinical practice. Indeed, electrodes represent almost half of the procedure costs, and this study showed the extent to which their price is likely to affect electrochemotherapy's value for money. The type of headroom analysis using the probabilistic output of a PSA may provide useful information during the negotiation of the price of the electrodes for ECT at the different decisional levels (e.g., at the hospital, regional or national level). This analysis clearly sets some criteria to define a value-based price for the device based on the existing uncertainty and evidence supporting its cost-effectiveness claims. This study, and particularly the headroom analysis on the cost of the electrodes also shows that cost-effectiveness of electrochemotherapy is highly dependent on which patients subgroups are considered in the analysis, suggesting that the optimal price of the device is likely to be indication specific. Nonetheless, the analysis is based on small sample of patents, especially for the subgroup of patients with bleeding lesions. Therefore, results for this patient's population should be carefully interpreted. More research is needed to estimate the cost-effectiveness of electrochemotherapy in other less severe patient populations. Systematic collection of EQ-5D-3L questionnaires or any other quantitative reporting of quality of life during the electrochemotherapy treatment are essential for further economic evaluations of electrochemotherapy. It seems that even if it is obligatory to

collect quality of life data (e.g., due to the commitment in the study application), it is not done on a regular basis, as it is not considered important.

4.2. Early cost-effectiveness model on a novel total artificial heart

Motivation

This case study reports on an early CEA for a novel total artificial heart (TAH), a life-saving, mechanical circulatory support (MCS) device for patients with end-stage heart failure waiting for a heart transplant.

TAH are the most radical of MCS solutions and consist of the complete replacement of the diseased heart with a mechanical one.

The evidence available at market launch, for MCS devices in general, and TAH specifically, is usually very sparse and is mostly based on the limited studies used to achieve regulatory approval. Well planned randomized clinical trials are almost inexistent due to ethical issues or practical challenges with conducting RCTs, for example because of the limited patient population. Consequently, evidence for both clinical and economic evaluations usually accumulate post-launch through the collection of data from national or international multicentre registries, or the experiences of individual implanting centres.

Therefore, performing CEAs to inform healthcare payers adoption and reimbursement decisions on these types of devices is challenging. After simulating effectiveness data from a small single-arm, pivotal study of an hypothetical novel artificial heart, PBA was then applied to conduct sensitivity analysis over the cost-effectiveness of the devices compared to available alternatives.

In addition, the lack of pre-market evidence for TAH makes them particularly suitable for CED schemes. The existence of potential misalignment in the perception of the value of collecting further evidence

between manufacturers and healthcare payers has also been explored using VOI and different utility functions.

Disease problem and the technology under assessment

Heart transplantation is the optimal treatment for patients with advanced heart failure (179,180). Unfortunately, donors' heart availability is not sufficient to satisfy the demand for transplantations thus making it necessary to use ventricular assist devices to bridge patients until a new heart becomes available. In most cases, left ventricular assist devices (LVAD) have demonstrated good survival outcomes. However, approximately 40% of patients receiving LVAD suffer from right ventricular failure (179,181,182) leading to worse quality of life and increased mortality. There exist a variety of devices for temporary right ventricular support, however, worse long-term outcomes of patients surviving to right ventricular failure suggest the need for a long-term biventricular support. The use of biventricular assist devices, which provide support to both the left and right ventricles of the heart, has shown good outcomes, but for a limited patient population there exists a benefit from totally removing the sick heart and implanting a TAH (179). Until recently, there existed only one TAH licensed as a bridge to transplant device, the SynCardia TAH (SynCardia Systems, LLC Tucson, AZ, USA), which has evolved from the Jarvik 7, the first TAH model to be implanted in a human patient in 1981 (183). Despite being a life-saving device for patients awaiting in the transplant list, survival and complications rates and quality of life outcomes remained relatively poor (184,185) so that in the quest to improve patients outcomes, other TAH devices are under development or have been recently licenced for bridge-to-transplant therapy. Evidence on ventricular assist devices is generally poor at market launch with very few randomized clinical studies comparing device types and therapeutic strategies. In fact, most of the evidence comes from international registries, such as the INTERMACS, EUROMACS registries (182,186) or individual cohort studies. For the TAH, likely because until recently there was only one model available, providing a life-saving treatment to a very sick population,

no previous cost-effectiveness studies exist. Nevertheless, with the advent of novel models from different companies, more evidence will be needed to inform decisions about which device should be reimbursed, purchased, and implanted. Consequently, device manufacturers will also have an interest in understanding which type of evidence provides the best value for money when building their value proposition.

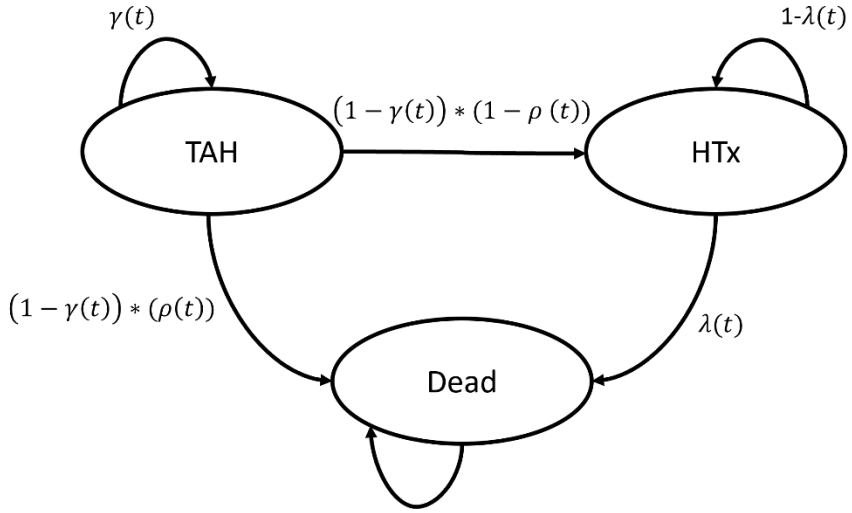
In the present analysis, we used Italy as the reference jurisdiction for the case study. From the payer side, we adopted the perspective of the national healthcare system considering only direct healthcare-related costs and consequences. From the manufacturer's side, we adopted the perspective of an individual company planning to submit for reimbursement to the national authority. The utility functions and decision rules for both manufacturers and payers are specified below in section 2.5.

Materials and methods

Cost-effectiveness model

A discrete-time semi-Markov model was adopted to estimate the cost-effectiveness of two different models of a TAH, the Syncardia TAH and a hypothetical novel TAH. The model was assumed to have three states, alive with TAH support (TAH state), heart transplantation (HTx) or dead. Patients in the TAH state can either survive until transplant or die with probability that is a function of the time from implant. Patients who receive a donor's heart move to the HTx state and remain in such state until death. The probability of death $\lambda(t)$ from the HTx state was also modelled as a function of time from transplantation. A graphical representation of the model is reported in Figure 4.5. Since the probability of transitions across states depends on the time of entrance in each state, following Sharples et al (187), the model was decomposed into two simple sub-processes, one representing time pre-transplantation and the second representing time post-transplantation. This approach allows to overcome the memoryless characteristic of Markov models and to correctly estimate transitions of patients according to their time-dependent probabilities.

Figure 4-5 Graphical representation of the cost effectiveness model for the total artificial heart



TAH, Alive with total artificial heart implanted; HTx, alive after heart transplantation

After implant of a TAH the survival of patients was modelled considering death and transplantation as competing events. For the Syncardia TAH, survival data were taken from a retrospective observational study reporting data on 450 patients implanted with a TAH in the INTERMACS registry (184). Kaplan Maier curves were digitized, and individual patient level data were reproduced using the algorithm proposed by Guyot et al. (188). The study reported survival curves with censoring at transplantation, however, since no other censoring occurred due to lost to follow up or other reasons, with the only exception of 6 patients still alive at the end of the observation period, it was possible to replicate full information on the survival of patients for both events of death and transplantation. The time to any event, was modelled as a Generalized Gamma distribution based on visual inspection and the AIC/BIC criterion. Then, conditional on experiencing an event, patients were distributed among death or transplantation states based on the estimated number of events in each cycle. The probability of transition to either transplantation or death was calculated as

$$(1-\gamma(t)) * \rho(t)$$

and

$$\left(1-\gamma(t)\right)*\left(1-\rho(t)\right)$$

where $\gamma(t)$ is the survival function for any event and $\rho(t)$ is the proportion of deaths conditional on leaving the TAH state.

The cumulative hazard of death rapidly decreases in the first three months and then remain almost constant afterwards (184). Therefore, the proportions of deaths and transplantations were estimated individually for the first 3 months and then assumed constant thereafter. Estimated and reported cumulative incidence functions from competing risk analysis were then compared to check the validity of the approach resulting in nearly identical curves. Four major complications were considered and defined according to the INTERMACS definitions of adverse events for mechanical circulatory support (189): ischemic strokes, major bleedings, major infections and major device malfunctions. Strokes were further classified into disabling (moderate or severe disability) and non-disabling strokes. Patients surviving to a disabling stroke were assumed to have lower quality of life after transplantation compared to their non-stroke peers.

Expert clinicians agree in that the novel TAHs under development will improve the quality of life of patients. Such improvement is expected to be realized through a lower rate of non-fatal adverse events, higher mobility and discharge rates while on support, as well as improved hemocompatibility and flow regulation (181). In the absence of evidence on the quality of life of patients on support with Syncardia (182,190,191) we presumed that data on quality of life was available from the pivotal study on the novel TAH and assumed that the differences in quality of life between TAH could be elicited through expert opinion.

Disease specific mortality after HTx remain high in the first three months and then become almost zero (192–195). Similarly to survival curves with the TAH, parametric extrapolation was done from digitized data of the KM curves from David et al. (192) who provide mortality data with the longest

follow up of 12 years. Italian life tables were then used to incorporate all-cause mortality in the cohort assuming an average age of 50 years (184,193). The cost of the device for the Syncardia TAH was retrieved using the Italian device national classification and the public expenditure data from the Ministry of Health. Patients were assumed to either remain hospitalized until death or HTx occurred or to be discharged after an average stay in the cardiac ward of 50 days (184). Costs of complications were valued using Italian DRG tariffs.

Cost data of patients receiving a TAH were not available in Italy or elsewhere. Therefore, we used data from Sharples et al. who did collect micro-costing evidence from 70 patients receiving a left ventricular assist device in the UK (187). With the only exception of the cost of the devices, we assumed that the cost of the index procedure including implant operating theatre costs, initial ICU stay, maintenance drugs and maintenance tests, were similar among left ventricular assiste devices and TAHs and had similar standard deviations. Similarly, monthly costs after HTx were assumed equal to patients undergoing left ventricular assist devices. All cost figures from Sharples et al. have been inflated to 2019 and converted into EURO using the average 2019 conversion rate between Pound sterling and Euro from of 0.878 (196). The model estimated the lifetime costs and consequences of patients until death. The assumptions on costs were validated with an expert cardiac surgeon in Italy.

The face validity of the model has been verified by discussing both the model structure and the data sources with cardiac surgeons with experience in MCS and VAD implantation. The internal validity of the model has been checked by performing extensive analysis of the model behaviour in extreme scenarios, and by building the same model using two different softwares (Rstudio and Microsoft Excel) and ensuring that results were consistent.

Simulation of data for the novel TAH

A hypothetical comparison was created by micro-simulating data based on the expected improvements of a novel TAH under design compared to the

Syncardia TAH. We presumed that 24 months data from a single arm pivotal study with 35 patients were available and that the study collected information on main clinical endpoint and quality of life but not resource use and costs. The simulated evidence for the study is consistent with the data that is expected for a new medical device in this field. For example, the ongoing clinical study to support the Humanitarian Device Exemption (HDE) application of the Syncardia TAH as destination therapy for patients who are not eligible for transplant (ClinicalTrials.gov identifier: NCT02232659) has an open-label, single arm design aiming to measure safety and clinical benefit in an expected number of 38 enrolled patients. Similarly, the pivotal study to support CE mark application for the CARMAT TAH (ClinicalTrials.gov Identifier: NCT02962973) envisioned the enrolment of 20 patients in a single-arm study with a follow up of up to 2 years, measuring safety, and effectiveness endpoints, including quality of life measures.

We assumed that patients would have similar times on support before experiencing either death or HTx, but that the risk ratio of death between the novel and Syncadia TAH at each time point was lower by 0.25. Therefore, time to any event was simulated from a Generalized Gamma distribution and then competing events were classified as death or HTx based on the calculated proportions. Patients still on support at 24 months were considered as censored. Similarly, a 0.15 reduction in the incidence of any of the 4 considered major adverse events was simulated. The number of adverse events in the pivotal study was calculated by multiplying the number of enrolled patients for the adjusted incidence and then by rounding the results to the nearest integer. These numbers were then used as inputs in the model to estimate the posterior probability distribution of adverse events. Procedure costs were considered equal to the Syncardia TAH except for the cost of the device. During support, monthly differences in costs were originated from the model based on differences in the number of discharged patients and the incidence rates of adverse events. Patients surviving to HTx

were assumed to have same monthly costs with the two devices. Due to the simulated innovativeness of the device and the likely development costs, we presumed that the cost of the novel TAH would be set at €200,000.

The model estimated the costs and consequences for the whole patients' lifetime with future costs and consequences equally discounted at a discount rate of 0.035. For the full set of parameters, we performed a probabilistic sensitivity analysis (PSA) by running a Bayesian model in Jags from RStudio to obtain samples from the joint posterior distribution of model parameters for both the novel TAH and Syncardia TAH. For the PSA we used a simulation size of 50,000 estimated across three chains with a burn-in of 5000 each. The full table of model parameters with 95% credible intervals and the distributions used in the PSA is reported in Table 4-5.

Table 4-5 Parameters of the cost-effectiveness model

Values	Mean (95% Credible Interval)	Source	Parametric distribution
Probability of remaining in the TAH state – Survival curve parameters			
Location parameter – Syncardia TAH	5.1 (4.6; 5.7)	(184)	Generalized Gamma
Location parameter – novel TAH	4.14 (2.8; 6.02)	Simulated	Generalized Gamma
Shape parameter (assumed equal for both TAH devices)	0.9 (0.8; 0.9)	(184)	Generalized Gamma
Probability of death while on support conditional on leaving the TAH state			
1 month			
Syncardia TAH	0.7 (0.3; 0.8)	(184)	Beta
Novel TAH	0.61 (0.34; 0.84)	Simulated	Beta
2 months			
Syncardia TAH	0.5 (0.4; 0.6)	(184)	Beta
Novel TAH	0.12(0.003; 0.41)	Simulated	Beta
3 months			
Syncardia TAH	0.4 (0.2; 0.5)	(184)	Beta
Novel TAH	0.14(0.004; 0.43)	Simulated	Beta
> 3 months			
Syncardia TAH	0.2 (0.2; 0.3)	(184)	Beta
Novel TAH	0.13(0.02; 0.33)	Simulated	Beta
Complications while on support			
Strokes			
Syncardia TAH	0.23 (0.18;0.27)	(184)	Beta
Novel TAH	0.21 (0.1; 0.36)	Simulated	Beta
Proportion of disabling strokes	0.12 moderate-severe disability; 0.39 Severe	(184)	Beta
Major bleeding			
Syncardia TAH	0.21 (0.10; 0.36)	(184)	Beta
Novel TAH	0.19 (0.08; 0.33)	Simulated	Beta
Major Infection			

Values	Mean (95% Credible Interval)	Source	Parametric distribution
Syncardia TAH	0.7 (0.65; 0.74)	(184)	Beta
Novel TAH	0.59 (0.44; 0.74)	Simulated	Beta
Major device malfunction			
Syncardia TAH	0.106 (0.08; 0.13)	(184)	Beta
Novel TAH	0.108 (0.03; 0.22)	Simulated	Beta
Survival after HTx (all devices)			
Location parameter	-5.24 (-5.52; -4.96)	(192)	Generalized Gamma
Shape parameter	0.543 (0.051; 5.80)	(192)	Generalized Gamma
Quality of Life			
Quality of life while on support			
Syncardia TAH	0.64 (0.58; 0.69)	Assumed as informed by Expert opinion	Beta
Novel TAH	0.70(0.66; 0.73)	Assumed	Beta
Quality of life after HTx – without disabling stroke	0.76 (0.64; 0.86)	(187)	Beta
Long term utility decrement with disabling stroke	-0.18 (-0.23; -0.13)	(197)	Beta
Costs			
Cost heart transplant procedure and first month in the hospital			
TAH cost			
Syncardia TAH	€ 100,000	Expenses data from the Italian MoH	-
Novel TAH	€ 200,000	Assumed	-
TAH implant procedure + first month hospitalization	156,625 (106,382; 216,435)	(187)	Gamma
Monthly hospitalization cost after the first	Decreasing from 19,224 (3,730; 47,241) for the 2nd month to 1,744 (31; 6,705) in the 8th month	(187)	Gamma
Proportion of TAH patients discharged			
Syncardia TAH	0.11 (0.09; 0.14)	(184)	Beta
Novel TAH	0.73 (0.58; 0.86)	Assumed	Beta
Average LOS before discharge (for those discharged)	50 (37; 64)	(184)	Gamma
Costs of Stroke	5,587 (1,039; 17,794)	Mean of DRGs 559; 14 and 15	Log-normal †
Cost of Major bleeding requiring re-operation	4,321 (802, 13,762)	Italian DRG 174	Log-normal †
Cost of major infection	7,148 (1,320, 22,926)	Italian DRG 576	Log-normal †

Values	Mean (95% Credible Interval)	Source	Parametric distribution
Cost of major device malfunction	81,551(15,259, 260,400)	Italian DRG 103	Log-normal †

† for all costs of complications standard deviation on the log scale was arbitrarily set at 0.5

Perspective and decision rules

The payer was presumed to be risk neutral, taking decisions based on expected incremental net monetary benefit (INMB). We assumed that the payer could take "accept" or "reject" decisions and mandating further research by making coverage conditional to further evidence generation such as in "only in research" or "approval with research" schemes. These are the decision rules and policy options that would be available in Italy according to the proposed new national plan for the HTA of medical devices (198). Nonetheless given the low incidence of patient eligible to TAH and the severity of the condition only research with approval (e.g., only with research) were considered applicable, meaning that the payers decision could prevent access of the device to the market, but not limit its market shares. . Following the national appraisal, it was assumed that regional health authorities would comply to the national decision, but practitioners would still be free to choose which device to utilize, that is, implementation would not be guaranteed after a positive reimbursement decision at the national level.

The manufacturer was presumed to make decisions on whether to submit or not for reimbursement based on the expectation of a positive net present value (NPV) of the investment. Initial investment to enter a country was assumed to be €500,000. This cost may represent for example the irrecoverable costs required to preparing and going through submission, such as the costs of hiring national consultants to build the submission dossier or the costs of negotiating with the national authorities. If the company took a "no go" decision, NPV was set to zero (i.e., zero revenues and costs). In case of a "go" decision, revenues were modelled as a function of a positive reimbursement decision from the payer, and the parameters that manufacturers believe to affect market shares in case of adoption. In case of

a negative appraisal, market access of the technology would be prevented, and therefore market shares were set to zero and the manufacturer incurred a loss of the initial investment. In case of a positive appraisal, market shares were assumed to be based on the risk-ratio of successful heart transplantations and modelled using a truncated function assuming 0 values for relative risks lower than 1 and following a cumulative normal distribution with mean 1.2 and standard deviation 0.07 otherwise. This function is similar to the one proposed by Willan et al. (Willan, 2008) and was assumed to reflect the company's belief around the link between the expected treatment effect and market shares. Operative income was calculated as revenues, minus direct and indirect operating costs. Direct operating costs were set at 20% of the price of the device whereas indirect operating costs were assumed to be €250,000 per year.

For both manufacturers and payers, we assumed that approximately 10 patients per year would receive the TAH and that the novel device would stay on the market for T=10 years before becoming obsolete. We made the simplifying assumption that the price of the device, the market shares and the number of eligible patients would remain constant in each year considered for the calculation of the NPV and NMB, however, these assumptions could be easily relaxed in the model. More in general we assumed that during this time horizon no other changes such as price modifications, other competing technologies entering the market, or new evidence becoming available would occur and affect the estimation of the NPV, the NMB and the value of information for both payers and manufacturers.

The population INMB from the healthcare payer perspective was calculated as:

$$INMB_{HC} = (k\Delta u_{HC} - \Delta c_{HC})\left(\sum_{t=1}^{T} \frac{I}{(1+dr)^t}\right)$$

Where Δu_{HC} and Δc_{HC} are differences in costs and consequences between the novel TAH and the Syncardia TAH that are directly relevant to the

healthcare payer; *k* is the applied cost-effectiveness threshold; I is the annual incidence of patients eligible to an artificial heart; and *dr* is the annual discount rate which was assumed to be the same for manufacturers and payers. The NPV for the company deciding to enter a market was calculated as:

$$NPV_M = \begin{cases} -c_M^{inv} & for\ INMB_{HC} < 0 \\ \sum_{t=1}^{T} \frac{(p_M - c_M^{dir}) I \delta_M - c_M^{ind}}{(1+dr)^t} - c_M^{inv} & for\ INMB_{HC} \geq 0 \end{cases}$$

Where c_M^{dir}, c_M^{ind} and c_M^{inv} are the direct, indirect operative costs and the initial investment cost incurred by the manufacturer; p_M is the price of the device; and δ_M is the company's annual market share.

For both manufacturers and payers, the same discount rate was assumed. This assumption was then explored by conducting deterministic sensitivity analysis on the discount rates for manufacturers. The inclusion in the sensitivity analysis of a different discount rate for manufacturers was justified by the fact that discount rates for healthcare payers and manufacturers represent different types of opportunity costs; the opportunity cost related to the future marginal productivity of healthcare spending for the former, and the opportunity costs related to the returns that could be earned from other investments for the latter.

Two additional constraints were set for the manufacturer to reflect his beliefs about additional market barriers beyond the payers' approval: a minimum clinical difference (MCD) and a maximum budget impact (MBI) constraint. The MCD for the risk-ratio of successful transplantation was set to 1.2 representing the threshold below which the manufacturer does not believe that clinical practice (and market shares) would change even if reimbursement was granted. The MBI of implanting the novel TAH was set at €80,000. This constraint may reflect for example the fact that hospitals may be reluctant to purchase the novel TAH if the expected budget impact was too high and no immediate adjustments of the Diagnosis Related Group

(DRG) tariff to accommodate the extra cost of the device was done. For both constraints, it was assumed that the manufacturer would take a "no go" decision if the risk of not complying with any of these constraints was higher than 0.5.

Value of information analysis

The EVPI for manufacturers and payers was then calculated as reported in section 3.6, i.e., by taking the difference of mean of maximum values in each sample of the MCMC samples, representing payoffs with perfect information, and the maximum of the mean of MCMC samples, representing payoffs with current information.

Following Koffijberg et al. (199) the expected value of information in presence of a constraint was calculated through different steps.

First, the best option with the current knowledge that provided the highest expected payoffs (i.e., either a positive operating income or zero) to the company was identified. The optimal option took also into account the compliance with the applicable constraints and with the acceptable risk of exceeding these constraints. Second for each sample from the Monte Carlo Markov Chain (MCMC) simulation the highest *acceptable* payoff was derived: the go decision was selected if the operative income was positive, and all applied constraints were met; alternatively, the no-go decision was selected instead. The expected value of perfect information (EVPI) with a constraint is then calculated by subtracting the highest acceptable payoff from the best option with current knowledge (199).

Value of information for manufacturers can also be affected by other uncertain parameters such as their direct and indirect operating costs or market dynamics like the time before a competitor would enter the market. Since this uncertainty cannot be resolved through further generation of evidence, they have been considered fixed in the base case analysis. One-way deterministic analysis was then conducted to estimate the impact of these parameters on value of information. Specifically, the following parameters were varied by plus, minus 25%: the yearly indirect operating costs, the initial

cost of the submission (e.g., the initial investment required to enter a specific market); the annual market size, the time horizon considered for calculating the NPV and the interest rate used by the manufacturer to discount future cashflows.

Probability Bound analysis

PBA was estimated from just the perspective of the healthcare payer. In order to apply PBA, it was assumed that for some of the parameters of the novel TAH that were simulated in the base case analysis, data was not sufficient to fully characterize uncertainty through a probability distribution. The following steps were followed (§3.5):

S1: A subset θ_p of model parameters θ_i were chosen. These parameters were: (1) probability of dying in the first month, (2) probability of dying in the second month, (3) probability of dying in the third month, and (4) probability of major device malfunction.

S2: For each parameter in θ_p, It was presumed to have different sets of minimal information on the parameter statistics including: the minimum, maximum, mean, median, and standard deviation, comprising $\mathcal{D}_i$.

S3: For each θ_i, the interval [0,1] was partitioned into 15 sub-intervals (n_i =15), and subsequently, a uniform weight $w_{i_k} = \frac{1}{15}$ was then assigned to each sub-interval. The number of sub-intervals was chosen after testing several n_is (5, 10, 15, and 30) and determining by visual inspections that 15 provided a reasonably accurate approximation to the p-box of Y.

S4: For each sub-interval of θ_i, the corresponding upper and lower endpoints in the parameter space were calculated, using the quasi-inverses. The quasi-inverses were obtained from applying equations A.18 and A.17 in Supplementary materials S3.1 to intersect equations A.15 and A.16 and A.8 and A.6, respectively.

S5: Given the sub-intervals, all possible combinations were enumerated by intersecting a particular sub-interval for each parameter, giving a total of 15^4 combinations.

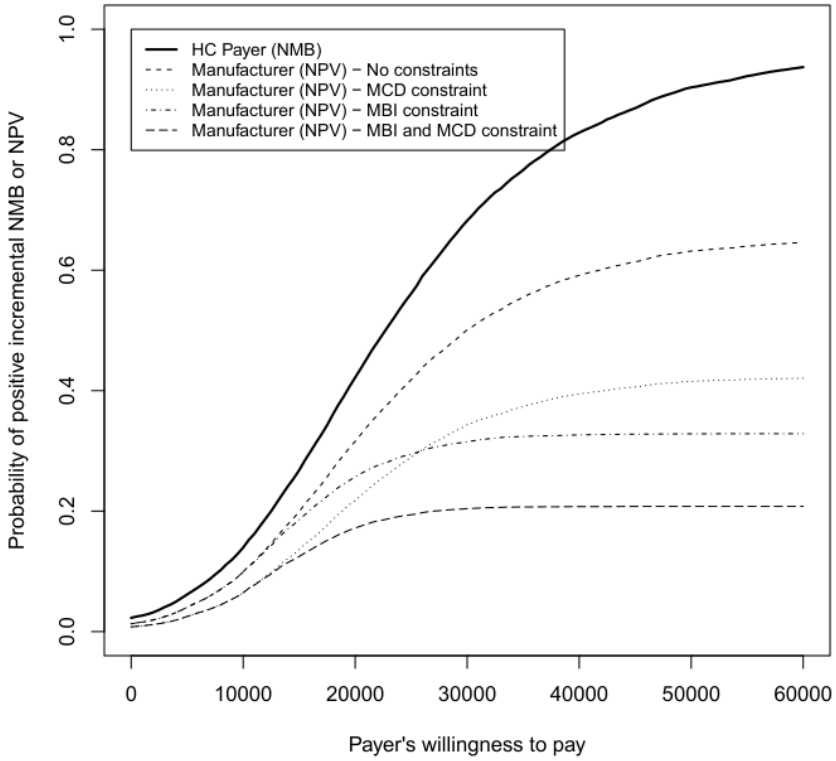
S6: To compute the minimum and maximum of the INMB, a deterministic search algorithm was employed, based on systematic divisions of the domain (each combination of intervals) into smaller hyperrectangles (200) and the non linear optimization library nplot for the R program (201).
S7: To derive the p-box of the INMB, two empirical CDFs were calculated by cumulating the weights for the values of the minimum and the maximum that are less than some arbitrary values in the support of INMB, in the interval [-€200,000 to €125,000].
As comparisons, beta distributions were used alternatively to represent the status quo practice of using "off-the-shelf" distributions given minimal data and uniform distributions to represent our complete lack of data/knowledge (ignorance) for the four parameters. for each of the four parameters, uniform distributions were constrained in the range between 0 and 1.

Results

In the base case, the novel TAH was expected to be more costly (with 0.975 probability) and more effective (0.997 probability) than the Syncardia TAH. At a willingness to pay (WTP) threshold of €30,000 per QALY gained the expected INMB of the novel TAH was €31,277 with a probability of being the cost-effective alternative for the healthcare payers equal to 0.67 (Figure 4.6). At the same threshold, the probability that the manufacturer would benefit from entering the market was lower, at 0.53. This is because the likelihood of positive operative income is affected not only by the probability of HTA approval, but also by market shares which in turn were modelled as a function of the device efficacy alone. Therefore, there may be cases in which even though the device would be considered cost-effective by the payer, the estimated clinical difference would not generate a satisfactory level of revenues. The probabilities of a profitable market entry for the manufacturer get even lower if further constraints such as MCD or MBI are introduced. In fact, constraints further limit the cases when manufacturers would choose a go decision. For example, at a CE threshold of €30,000 only 22% of cases would have positive operating income, MCD greater than 1.2

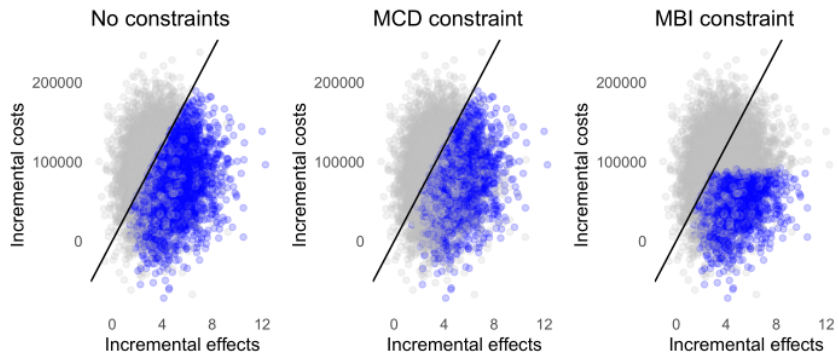
and a 24 month budget impact of less than €80,000 (Figure 4.6 and Figure 4.7).

Figure 4-6 Probability of a positive incremental net monetary benefit (for payers) and net present value of the decision to enter a market (manufacturers), as a function of payer's willingness to pay.



NMB, Net monetary benefit; NPV, net present value; MBI, Maximum Budget Impact; MCD, Minimum Clinical Difference

Figure 4-7 Samples of the probabilistic sensitivity analysis with positive net present value for the manufacturer and compliance with the imposed constraints (blue dots)



MCD, Minimum clinical difference; MBI, Maximum impact budget

Table 4-6 Estimation of the Net Monetary benefit (Payer) and Net present value (manufacturer).

HC Payer		
NMB Syncardia TAH (95% CrI)	**NMB novel TAH (95% CrI)**	**Incremental Net Monetary Benefit**
€ -31,309 (-148,807; 82,310)	€736.8 (-120,000; 126,000)	€31,277 (-79,324; 158,269)
Manufacturer		
NPV with no market entrance	**NPV of entering a market (million €)**	**Incremental Net Present Value (million €)**
0	€2.8 (-2.5; 11)	€2.8 (-2.5; 11)

NMB, Net Monetary Benefit; TAH, total artificial heart, Cri, Credible intervals

Market shares had mean value of 0.345 with truncation at 0 (44% of samples) and maximum value at 1. The expected NPV at a payer WTP threshold of €30,000 was approximately 2.8 million (95% CrI million € -2.5; 11). Results of the model for both manufacturers and payers are reported in Table 4-6

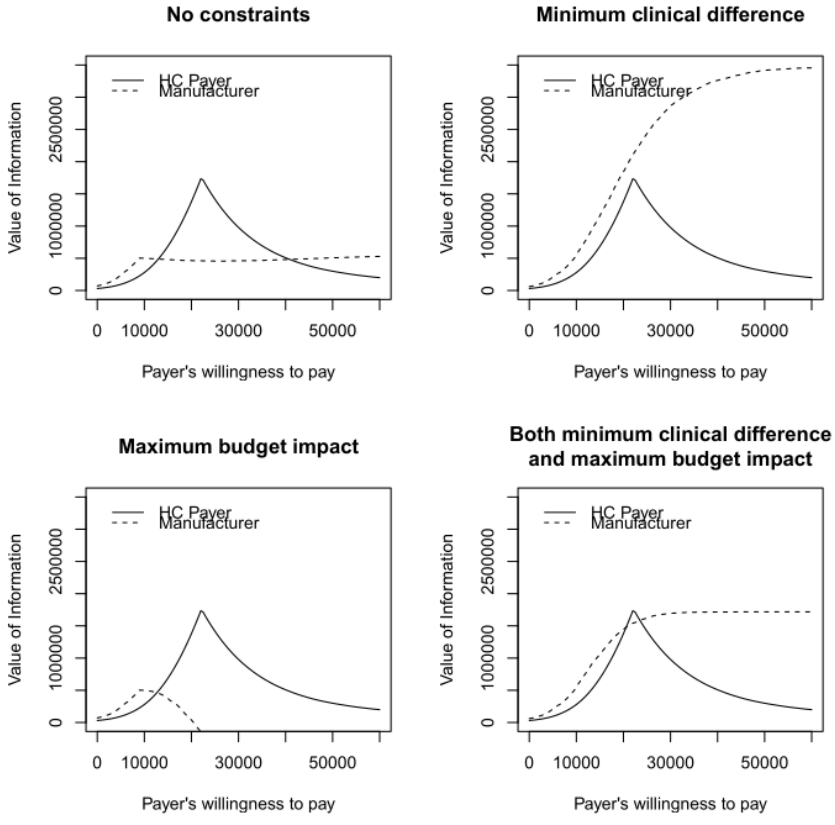
Value of information analysis

The expected value of information varied between manufacturers and payers and depended on whether constraints were added to the manufacturer's decision rules. In the absence of constraints, VOI for manufacturers increases until a payer's willingness to pay of €8,500 and then remain constant (Figure 4.8). At this payer's WTP threshold the uncertainty over the profitability of entering the market is maximum and the expectation of the NPV is close to zero. For WTP values higher than €8,500 the expectation of the NPV turns positive and the manufacturer would change its decision and opt for entering the market. Nonetheless, after deciding for entering the market, uncertainty would remain on the level of achievable market shares, which is unrelated to the payers' willingness to pay, so that the VOI remains constant.

At a payer's WTP threshold of €30,000, the population EVPI for payers is approximately €1.4 million whereas for manufacturers the value is lower at €451,335. However, if constraints are added the EVPI of the manufacturer

vary considerably. When a MCD constraint is added, the risk of not meeting the desired threshold of 1.2 for the novel TAH is about 0.57 and therefore it exceeds the maximum acceptable risk for manufacturers. Consequently, given the current knowledge, the manufacturer would not submit for approval no matter the expected NPV in the absence of the constraint. In this case, the gains from the value of information becomes higher as having perfect information would allow the manufacturer not to take such conservative decision and apply for reimbursement should the device prove to comply with the constraint and generate a positive income.

Figure 4-8 Expected Value of Perfect Information for healthcare payers and manufacturers and different constraints applied by manufacturers.



Value of information is estimated using population Net Monetary Benefit for HC payers and manufacturer's net present value, using a common time horizon of 10 years and 10 incident patients per year.

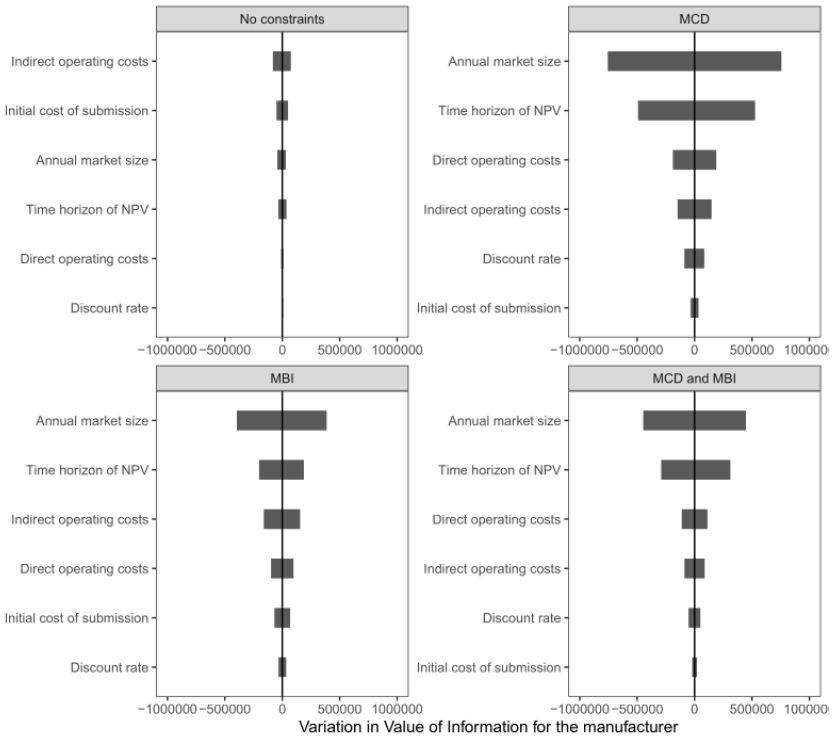
Since the probability of a positive NPV also depends on payer's behaviour, the benefit from perfect information is higher with higher values of the payer's WTP threshold and reaches its maximum when the probability of approval reaches 1 (and the VOI for payers is zero) (Figure 4.8). With the MBI constraint, the risk of the device having a 24 month budget impact higher than €80,000 is 0.47 and therefore is below the maximum acceptable risk for the manufacturer. Hence, with current knowledge, the company would choose to enter the market only based on the expectations of a positive NPV. Nonetheless, in cases when the manufacturer would choose to enter the market (i.e., for payers' WTP higher than €8,500) the application of the constraint will lower the VOI compared to the scenario with no constraints. This is because the expected value of the NPV in presence of perfect information would be lower due to the inclusion of the constraint in the manufacturer's decision rules. Specifically, any time the device would not meet the constraints in the samples from the MCMC the manufacturer would still opt for not entering the market even with positive expected NPV. If the proportion of samples exceeding the constraint is high, the expectation of the NPV with full information would be strongly reduced. Note that due to this effect, over a certain value of the WTP threshold, the expectation of the NPV with full information gets even smaller than the expectation of a go-decision with current knowledge, resulting in a negative EVPI. The full results of the EVPI for manufacturers and payers and different constraints are reported in the supplementary materials (Supplementary material S4.2)

With both MCD and MBI constraint, the EVPI of the manufacturer increases similarly to the case with the MCD constraint because similarly to that case the manufacturer would not opt for entering the market given that the risk of the device not complying with this more restrictive constraint is higher than 0.5. The value of the EVPI however, remains lower due to the effect of the MBI constraint reducing the expected value of the optimal choice.

Univariate sensitivity analysis on those parameters for the manufacturer that are not related to the performance of the device were different depending on

whether constraints were applied or not. In the absence of a constraint, variations in the value of information were limited, whereas with any of additional constraints applied the EVPPI of the manufacturer varied considerably with the annual market size and the time horizon considered for the NPV being the most influential parameters (Figure 4.9).

Figure 4-9 Univariate sensitivity analysis of parameters which may affect the value of further research for the manufacturer



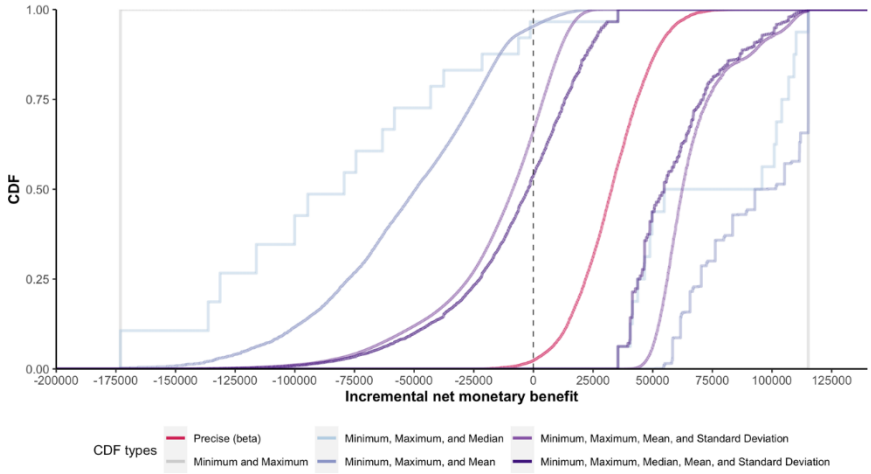
Parameters were varied by ±25%. MCD, Minimum clinical difference; MBI, maximum budget impact.

Probability Bound analysis

The first comparison of the PBA shows the difference between the results of a parameter uncertainty quantification using precise CDFs vs. PBA. For a given minimal data, a PBA results in a p-box enclosing the unknown CDF of the INMB (Figure 4.10). The p-box gives additional information: the amount of uncertainty in the INMB due to our imperfect or complete lack of knowledge about some parameters (indicated by the area enclosed by the

p-box) and the plausible values of the INMB (indicated by INMB values with non-zero CDF). The latter suggests the minimum and maximum achievable values of the INMB. We also observe that the area bounded by the corresponding UBF and LBF shrinks as we have more data. This is expected; as more data on the CDF becomes available, we are able to pinpoint the location and shape of the CDF more accurately.

Figure 4-10 Uncertainty around incremental net monetary benefits of TAH model using p-boxes under different minimal data vs beta distributions (red line)

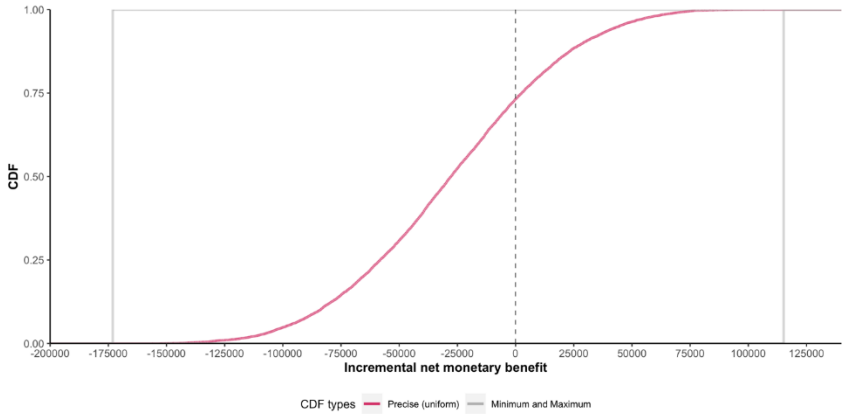


CDF: cumulative distribution function

The second comparison demonstrates the consequence of assuming more information than we have when uniform distributions are employed to acknowledge "ignorance" about the value of a certain parameter. Uniform distributions assume that each value in [0, 1] is equally likely, nevertheless, this is qualitatively and quantitatively different to knowing that the value is somewhere in the interval. A PBA does not provide any information on the uncertainty in the INMB given that the uncertainty in the parameters are represented by intervals with no measures of uncertainty assigned to the intervals. In other words, a PBA transfers our ignorance in the model parameters to ignorance in the model outcomes. In contrast to PBA, a PSA using uniform distributions does not preserve our ignorance, i.e., uniform

distributions in the model parameters do not generally translate to uniform distribution in the model outcomes unless the model is linear. Furthermore, using uniform distributions generates a "precise" CDF (Figure 4.11), hence creating a pseudo-certainty in the model outcome.

Figure 4-11 Uncertainty around incremental net monetary benefit of TAH model using p-boxes given minimum and maximum values vs. uniform distributions.



CDF: cumulative distribution function.

Discussion

In this case study, two novel methodological approaches to quantitively characterize and address parameters uncertainty in CEA models have been applied. A discussion on the relevance, limitations and further required work for each approach is provided here.

Value of information analysis

An efficient evidence generation for new medical devices would be beneficial for all, including manufacturers, healthcare payers and, most of all, the patients. Particularly, increasing the relevance of evidence for HTA purposes would improve the capacity of payers to take informed decisions at earlier stages and achieve their objective of maximizing population health. More relevant evidence would also be beneficial for manufacturers willing to maximize their probability of accessing markets and having positive incomes. Ultimately, patients will benefit from improved decision making and reduced

uncertainty over the true performance of novel technologies. Nonetheless, this analysis, and the case study on a hypothetical TAH, showed that the value of further research may differ between payers and manufacturers. Differences originate from the fact that the two parties pursue different objectives, but also may depend on other assumptions and prior beliefs specific to either one or the other. For example, when constraints were added to the decision rules of manufacturers, the value of information changed considerably (but the same would have happened if constraints were imposed to the payer). The analysis also showed that value of information for manufacturers is dependent on other factors which are not directly related to the uncertainty over the device, such as the market size of the target country or the expected time window during which manufacturers expect to have revenues before the device becomes obsolete and is overtaken by incremental innovations.

Our results confirm the importance that any potential misalignment in the (perceived) value of information is identified and addressed in order to achieve optimal evidence generation processes along the product life cycle. One ideal moment when such misalignments may be reduced is during early dialogues when the device is close to obtaining market approval and the evidence for HTA purposes is still to be generated. Although it is unlikely that dialogues on the study design during an ED are based on an explicit and transparent sharing of information and decision rules from both sides, a more consistent use of VOI may reduce asymmetric information and signal the potential reactions of each part to decision uncertainty, thus contributing to align on evidence requirements. In a typical ED for HTA, HTA bodies provide non-binding recommendations to manufacturers about their proposed pivotal study for HTA submission. The study will be fully funded by industry, so that HTA bodies just need to consider to what extent the results of such study will enable them to make informed decisions by the time the manufacturer submits for reimbursement. HTA bodies could use VOI to evaluate if, according to their utility function, the proposed design and sample-size exhaust the value of information for the technology under

assessment, or whether residual VOI would remain after the study reports its results. If manufacturers, based on their VOI calculations, propose a study design that also exhausts VOI from the HTA bodies and payers' perspective, the incentives for manufacturers and payers are likely to be similar and the appropriate evidence will be generated. If the proposed study leaves residual VOI, then HTA bodies could signal what would be the minimum level of evidence that should be generated pre-submission for the technology to be adopted. Since the recommendations of HTA bodies on a technology can prevent market access and therefore are usually considered in manufacturers' utility function this information can be used to update manufacturers' analysis on the optimal study design, thus obtaining more refined estimates of the expected payoffs for their marketing strategies and a closer alignment with the evidence requirements of HTA bodies. In theory, more complex interactions between HTA bodies and manufacturers might also allow further discussion on the intended evidence generation plan to seek agreement on the optimal sequential design of studies in a life-cycle perspective including optimal pre-launch and post-launch evidence generation. Nonetheless, such types of interactions are not actually envisioned within a typical ED Another time point in the product lifecycle when manufacturers and payers may discuss about evidence needs, is later, at the time of HTA appraisals, if adoption decisions by national or regional payers envisage the possibility of arranging performance-based risk-sharing schemes. One of the key elements of these schemes is that the price, or reimbursement of a technology is linked to its performance which is assessed through a purposeful, prospective data collection. In Chapter 5, the extent by which CED programmes are used in Europe is explored as well as the different arrangements about how they are implemented, and how responsibilities are shared between the parties involved. For example, in France the high authority for health (Haute Autorité de la Santé - HAS) may temporarily register a device in the positive list for reimbursement and make renewal conditional on the generation of further evidence to reduce some of the remaining uncertainty on the performance of the device. The

responsibility of the generation of evidence relies entirely on the manufacturer, which must report the new data at the time of the reassessment, usually after 3 or 5 years. However, it is not uncommon that re-appraisals from the HAS report that the required data is of insufficient quality or even totally missing. This may reflect the lack of incentives for manufacturers to conduct the new research once their product is approved even for a limited period of time. For example, the company may be developing and plan to launch an incremental innovation within a short time, or may know that other competitors are about to launch similar devices so that investing in further research may not be considered worthwhile.

Even if companies' private information and strategies may not be accessible to HTA bodies and payers, considerations of the expected market size for the technology under assessment and historical information on the rate of innovation in the specific field of the technology could provide insights to payers on the likelihood that manufacturers will conduct the desired research. Nonetheless, the difference in goals and incentives together with lack of transparency may still hamper any alignment process so that when agreeing CED schemes, HTA bodies should always ensure that appropriate mechanisms are in place to monitor and eventually sanction uncomplying manufacturers.

This case study has estimated the value of information for manufacturers in relation to the specific decision problem of entering a target market. In fact, the analysis included parameters that are generally considered as poorly generalizable (as for example costs and patients utilities). In addition, and perhaps even more important, the estimation of VOI for the manufacturers incorporated other aspects that are likely different across countries. These aspects include for example the decision rules adopted by payers to determine reimbursement (e.g., based on clinical criteria only rather than NMB), and characteristics of the market, such as the market size, or the existence of other competitors. Consequently, the VOI for the company estimated in one market would not be directly applicable to the company's decisions to enter other markets. However, some of the benefits of reducing

uncertainty for one specific decision problem, will also have positive spill over effect on other (future or parallel) decisions of the company, although as mentioned the extent of these spill over effects will be dependent on the type of uncertainty to be reduced. For example, in the case study proposed, future decisions of the company will benefit from a reduced uncertainty on the rate of adverse events, but less so on the costs of such adverse events, since these are likely to be more context specific. Obviously, the more this reduction in uncertainty is generalizable the higher the spill over effects for the company and the higher the global value of information. When the contributions of specific study designs are considered then also the type of study conducted may impact the generalizability of results and therefore how "reusable" that evidence would be for other company's decisions. For example, a Randomized controlled trial would be expected to be more generalizable compared to an observational study based on a registry. In any case, while spill overs effect at the global level may be difficult to estimate, calculating the VOI for single-market decisions may be still useful when the evidence required (e.g., by payers) is highly context specific, as is the case of many CED schemes focusing on the real-world performance of the device in the country's clinical practice. In addition, by calculating single-market VOI, manufacturers can already say whether the costs of further research would be justifiable even considering the expected returns of this single decision problem. So, in a sense, the single-market EVPI for manufacturers is the first check for a company on whether further research would be worthwhile. In addition, this analysis could also orient manufacturers in the choice of where to conduct a study. These decisions may also have negative spill over effects. For example, a negative appraisal might have consequences on the decisions taken by payers in other markets. Or the decision of not submitting for approval in a country may impact the company's probability of accessing late adopters' markets. The incorporation of these aspects may require more complex models to reflect the global value of research in other markets and will be the focus of further research.

To the best of the author's knowledge, no other studies applied VOI to manufacturers decision processes using a different maximization objective and incorporating the existing barriers to access and market dynamics. As discussed, this analysis may support manufactures to optimize their evidence generation plan and contribute to reduce misalignments between payers and manufacturers. This analysis has also limitations. First, the proposed case study uses a hypothetical TAH as the novel device to calculate the VOI and make several assumptions on both payers and manufacturers behaviours and beliefs. In addition, other simplifying assumptions when building the model were required by the paucity of evidence on the existing TAH. For example, it was impossible to characterize the correlations between costs and effects in the model.

While these assumptions are likely to affect the results of the economic model and VOI analysis, they do not invalidate the conclusions of the study on how VOI could be used to explore misalignements in the perceived value of research between payers and manufacturers. Nonetheless, the development of real-world applications is necessary to understand the real entity of the misalignments on VOI between payers and manufacturers, and the extent to which the proposed framework is applicable to other technologies, e.g., pharmaceuticals, and health conditions.

Second, even in the case of real applications, several parameters to calculate VOI may not be easily estimated empirically. In the case of manufacturers, the value of these parameters may be informed by the practices that are routinely performed for business intelligence. For example, studies on market dynamics (e.g., competitors' product and pricing strategies, novel devices in pipeline) could inform the time horizon to consider when calculating the NPV. Also, the link between the device effectiveness and the market shares, including any minimum clinical difference that is required to trigger adoption could be derived from heuristics based on previous experience of the company with similar devices, or by directly eliciting providers preferences e.g., using stated preferences techniques such as discrete choice experiments. Similarly, the investment costs as well as the

direct and indirect operating costs could be informed by operations management or other analyses to assess investment decisions. In the case of healthcare payers, the estimation of VOI requires an estimate of the present and future population that will benefit from the reduction in uncertainty over the decision being addressed and the time horizon before exogenous changes (such as new treatments, new evidence becoming available, changes in prices etc.) would modify the results of the underlying cost-effectiveness model and therefore the VOI. These parameters can be estimated sourcing available real-world data such as for example epidemiological data, and using approaches similar to the ones used by manufacturers to do horizon scanning of the upcoming technologies.

Second this study only estimated the expected value of perfect information for manufacturers and payers, i.e., the benefits of reducing all existing uncertainty on model parameters. Further studies may extend the analysis to calculate the payoffs of reducing uncertainty over specific parameters (expected value of partial perfect information EVPPI) or with specific study designs (expected value of sampling information, EVSI) (67). Further research may also explore the role of additional constraints on the EVPPI and EVSI and develop methods to incorporate spill over effects on the company's decision making.

Probability Bound Analysis

In the base case for this case study, several assumptions were required on the expected values of parameters used in the models and their parametric distributions to perform probabilistic sensitivity analysis and estimate VOI. Lack of information on model parameters is a common issue when doing early assessment of novel technologies, and it has been argued that performing probabilistic analysis (including VOI) with limited data on model parameters can create pseudo-certainty (202) and incorrect decisions (203). The second part of this case study introduced the PBA method as an alternative approach for modelling parameter uncertainty without necessitating the use of unverifiable assumptions regarding the functional

form of parameters probability distributions. In the PBA approach, p-boxes were propagated into a black-box model where the uncertainty of the model parameters is characterized by a combination of p-boxes and precise distribution functions.

In presence of total or partial ignorance about the distribution of a parameter, the PBA method has advantages over the standard approach in the following sense. First, in PBA, parameter uncertainties are characterized by p-boxes that provide the maximum area of uncertainty (tightest bounds) containing the unknown CDF, given some data characterizing the CDF (e.g., summary statistics). The uncertainty propagation of p-boxes into a black-box model, using optimization, generates bounds that are guaranteed to enclose *all possible CDFs* of the outcomes of interest provided that the p-boxes of the parameters enclose their respective (unknown) CDFs. Second, if the lower and upper bounds of a CDF coincide for every element in the parameter support, then a p-box degenerates to a CDF: a situation where Monte Carlo simulation is the standard approach. Therefore, a PBA is a generalization of the two standard approaches for representing parameter uncertainty. Third, the bounds on the plausible values of a model outcome are a critical decision-relevant information provided by PBA. This information is particularly useful when the outcome represents a decision-criterion, such as the INMB in our case study. The resulting p-box that spans over both positive and negative INMB values suggests that there is substantial uncertainty in the INMB being positive as well as negative. The status-quo approach may yield less uncertainty in the INMB and potentially lead to an over-confidence in the INMB being positive.

The approach can be nicely applied to decision making on regulatory and reimbursement issues. In particular, decisions on medical devices, orphan drugs, or advanced therapy medicinal products are subject to great uncertainty due to a lack of controlled clinical trials, e.g., in the case of medical devices, too small clinical studies, e.g., in the case of orphan drugs, (204) or the use of surrogate and intermediate endpoints (39,40). The use of advanced methods, such as PSA, allow health care decision-makers to

account for such uncertainties and consider the degree of confidence in their decision-making. However, health care decision-makers often do not take into account the effect of the underlying assumptions of such advanced methods. This ignorance is problematic from both a regulatory and reimbursement perspective, as it may result in technologies with an unfavourable benefit-risk profile entering the market (regulatory decision) or in products that are not cost-effective being reimbursed (reimbursement decision). The PBA we advocate does require less distributional assumptions of the input parameters. Therefore, the decision is often not as unambiguous as when using a PSA.. Despite its utility and attractiveness, PBA is computationally intensive for the following reasons. First, implementing PBA requires an optimization step over the parameter space. The higher the desired level of accuracy is, the higher number of sub-intervals n_i (i.e., the finer the discretization of the parameter space) is needed. Second, an increase in the number of p-box parameters will lead to a higher-dimensional optimization problem. Third, the computational cost is further exacerbated if the model is "expensive" to evaluate for a given set of parameter values. Fourth, if, in addition to p-box parameters, some parameters are characterized by their precise CDFs, the optimization step is embedded in a Monte Carlo sampling loop; thereby increasing the number of optimizations by a factor of N (the total number of Monte Carlo samples). To decrease the computational cost, practitioners may opt to use more efficient optimization methods (205) and fast-to-evaluate approximations of the original model or meta-models (206). Nevertheless, we expect a higher computational burden since a PBA imposes fewer restrictions (i.e., we do not assume a functional form), leading to a larger region of uncertainty over which a model needs to be evaluated. With the rapid advancement of computing power and capacity, however, it is only a matter of time before the higher computational burden is no longer a constraint. When the first health economic models and PSA emerged in the 1980s, the models were computationally intensive for the computing power available at the time but today they no longer pose a major computational challenge.

Our study has limitations in the following context. First, independence among the model parameters was assumed. However, how to model dependencies among the parameters in the context of uncertainty propagation using PBA and black-box models is an open problem and warrants further study. Secondly, a clear rule prescribing when one should adopt PBA approach to characterize parameter uncertainty is not defined. Such decisions are left to the analysts and should be taken also considering the relevance of a parameter on the model outcomes. For example, a parameter may be highly uncertain due to the lack of empirical data and/or previous knowledge and, at the same time, non-influential, i.e., the decision is not sensitive to variations in the parameter values. Thirdly, only one interpretation of uncertainty in outcomes was provided, i.e., how to make decisions using the results of a PBA. Since a PBA generates bounds on the unknown CDF, the expected value is interval-valued instead of a single value. Therefore, a re-formulation of approaches that rely on one expected value, including value of information analysis, is needed. Indeed, as mentioned in section 3.4, the prevalent decision rule in cost-effectiveness modelling using the von Neumann–Morgenstern (VNM) utility theorem is no longer directly applicable with the PBA approach.

The Hurwicz decision criteria is compatible with the PBA approach, as a unique value for the CDF could be obtained by averaging the lower and upper bounds of the CDF using the "pessimistic" coefficient as weight (§3.4). Nevertheless, a more comprehensive exploration of the feasibility and acceptability of this approach to PBA is warranted and should be the focus of future studies.

Chapter 5
Characteristics and challenges of Coverage with Evidence development schemes for medical devices in Europe

This chapter of the thesis moves away from the methodological aspects of the methods by which uncertainty can be characterised and addresses the issue of uncertainty from a more policy-oriented perspective. Indeed, as already mentioned in the introductory chapters, regardless of the quality and quantity of economic and clinical evidence on a technology, doubts will always remain about the real impact of technologies in the specific real-world context where they are intended. This also means that public decision-makers will have to take decisions facing the risk that this decision will turn out to be wrong over time. This chapter discusses relevant aspects of conditional reimbursement schemes that are used in European countries to limit the consequences of making a wrong choice due to existing uncertainty.

Materials and methods

The research for this chapter was conducted in three consecutive steps: (1) development of a structured interview guide (2) interviews with decision-makers from a sample of European countries, (3) synthesis and qualitative content analysis of the interview data, the data made available by the decision-makers during or following the interview, and data on scheme characteristics previously obtained through the systematic literature review presented in section 2.6. These steps are described in more detail below.

Development of the interview guide

A structured interview guide was developed (Supplementary Material S5.1) that consisted of three sections. Section A included general questions on whether CED programmes underpinning the individual schemes existed in the decision-maker's country and for which type of technology they were used. Section B included questions on 13 challenges for CED schemes for devices (Table 5-1).

Table 5-1 Challenges with CED schemes for medical devices

	Challenge
1	Deciding which medical devices are candidates for CED schemes
2	Obtaining stakeholder agreement on the scheme
3	Securing funding for the scheme
4	Determining the appropriate study design for data collection
5	Determining the relevant outcome measure(s) on which data are collected
6	Dealing with data collection and monitoring
7	Dealing with data analysis
8	Ex-ante definition of decision rule, based on possible outcomes of the scheme
9	Reaching an agreement on price, reimbursement or use of the device at the end of the scheme
10	Withdrawing a device from the market when evidence indicates the device is not (cost-) effective
11	Obtaining agreements about the duration of the scheme and the stopping rule
12	Adapting the scheme to account for product modifications or a learning curve
13	Dealing with the market entry of similar devices

CED, coverage with evidence development.

This list was derived from the review described in §2.6 that identified 20 challenges for CED schemes for devices (1). In order to reduce the participants' burden, we reduced the original list of 20 challenges to 13, by grouping different aspects of the same general challenge. The final list of challenges was discussed and agreed among all researchers participating in this study to ensure that all relevant aspects originally identified were covered in the interview guide (Supplementary Material S5.2).Decision makers were

asked to assess how they perceived the 13 challenges to apply to CED schemes for devices on a six-point Likert scale (ranging from 0 "not a challenge" to 5 "a major challenge"). Where CED schemes for devices existed, respondents were required to state how the challenges were met in their country, and the interview proceeded to Section C. Otherwise, the interview ended here. Section C included questions on the detailed characteristics of individual CED schemes for devices that had been either initiated or still ongoing in the past five years. These questions concerned a description of the device under evaluation, its clinical application, the objective of the scheme, key sources of uncertainty, funding of the scheme, its design, the decision rule, and outcome (if re-assessment was done), and any public source of information on the scheme.

Interviews with decision-makers

A first draft of the interview guide was circulated for comments among the COMED project partners. Subsequently the final draft of the interview guide was pilot tested during interviews with one Italian policy maker and two academic experts with extensive experience of CED in Canada and the USA, two countries with a substantial number of schemes.

The interviews were conducted face-to-face or by telephone between June and December 2019. Decision-makers from decision bodies at the central (or in two cases regional) level were identified from the professional networks of the members of the COMED project team or the websites of relevant decision bodies in the following European countries: Austria, Belgium, Bulgaria, Croatia, Czech Republic, Denmark, England, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, the Netherlands, Norway, Poland, Portugal, Romania, Scotland, Slovakia, Slovenia, Spain, Sweden, and Switzerland. Other countries from the EU/EEA were excluded because it was not possible to identify a relevant decision-making body for the technology assessment of medical devices. Decision-makers were invited to participate in the study by sending them an email with information on the COMED project and the objective of our study. When

no decision-maker could be identified from the networks or websites, the same information and invitation letters were sent to the relevant decision bodies. In three cases where no relevant decision-maker could be identified (i.e., Bulgaria, Czech Republic, and Sweden), academic researchers with relevant expertise were invited to participate. None of these countries had however any CED programme for devices in place. More than one decision maker was interviewed from a given country in cases where schemes were operated in more than one jurisdiction (i.e., Italy), where more than one decision body was involved in operating schemes (i.e., France), or where more than one decision-maker, from different parts of the relevant organization, agreed to participate (i.e., England). Croatia, Iceland, Romania, and Slovenia were excluded from the sample after repeated attempts to schedule an interview by December 2019 were unsuccessful. Information on the individual CED schemes provided by decision-makers during or following the interview was supplemented with information on individual schemes previously obtained in the systematic literature review presented in section 2.6, compiled in tabular form, and sent to the participants for a validity check.

Data analysis

The transcripts were subjected to qualitative content analysis using deductive coding to meet the objective of this research. The results of each interview were reported in a table by the candidate and assessed in parallel with another COMED researcher who independently extracted the relevant information. Agreement on the data to be reported was then reached through discussion and further analysis of the original transcripts. The data obtained from Sections A and C of the interview guide, together with the data obtained prior to and following the interviews were used to identify and classify the characteristics of the existing CED programmes for devices according to the four phases of CED schemes: 1) assessing the desirability of the scheme; 2) designing the scheme; 3) implementing the scheme, and 4) evaluating it (82). These phases are described in more detail in Table 5-2.

Table 5-2 Phases of CED schemes

Assessing the desirability of a scheme This initial phase relates to the way candidate technologies for CED schemes are identified and selected. It also concerns the criteria used to assess whether a scheme is a good policy option, compared with other available options such as, for example, fully adopting the technology despite the residual uncertainties; refusing to adopt the technology until better evidence becomes available; or negotiating/mandating a lower price for the technology. **Designing the scheme** This phase is about deciding on the specific features of the scheme design. These include, for example, the categories of patients who will have access to the technology during the scheme (e.g., Only in Research or Only With Research schemes), and the characteristics of the data collection plan, such as the study design (e.g., registry-based studies *versus* randomized controlled studies), the duration of the data collection, and the types of outcomes to be measured. **Implementing the scheme** Reflecting the previous design phase, this phase is about the different ways schemes are operated and how roles and responsibilities are distributed among the stakeholders involved (e.g., the national/regional HTA agencies, the manufacturers, or the providers collecting the data). Relevant aspects are, for example, who will initially design the study protocol, who will coordinate and/or perform the data collection, monitoring and analysis, and who will fund the provision of care and the extra costs of collecting the new evidence. **Evaluating the scheme** This phase relates to the types of decisions/policy updates that are made at the end of the scheme once the data collection is concluded and the new evidence has been assessed along with other evidence that has become available. It also concerns the way data collection is monitored during the scheme and the definition of any stopping rule or intermediate assessment of the evidence being collected.

The information collected was then synthesised in a narrative review. The data obtained from Section B of the interview guide were used to obtain insight into the participants' perceptions of the 13 challenges and into the factors that influenced their score for a particular challenge. The quantitative data obtained from Section B were used to calculate the mean (SD) and median (IQR) Likert scores for the 13 challenges (excluding the challenges that were marked as 'not applicable' by the participants).

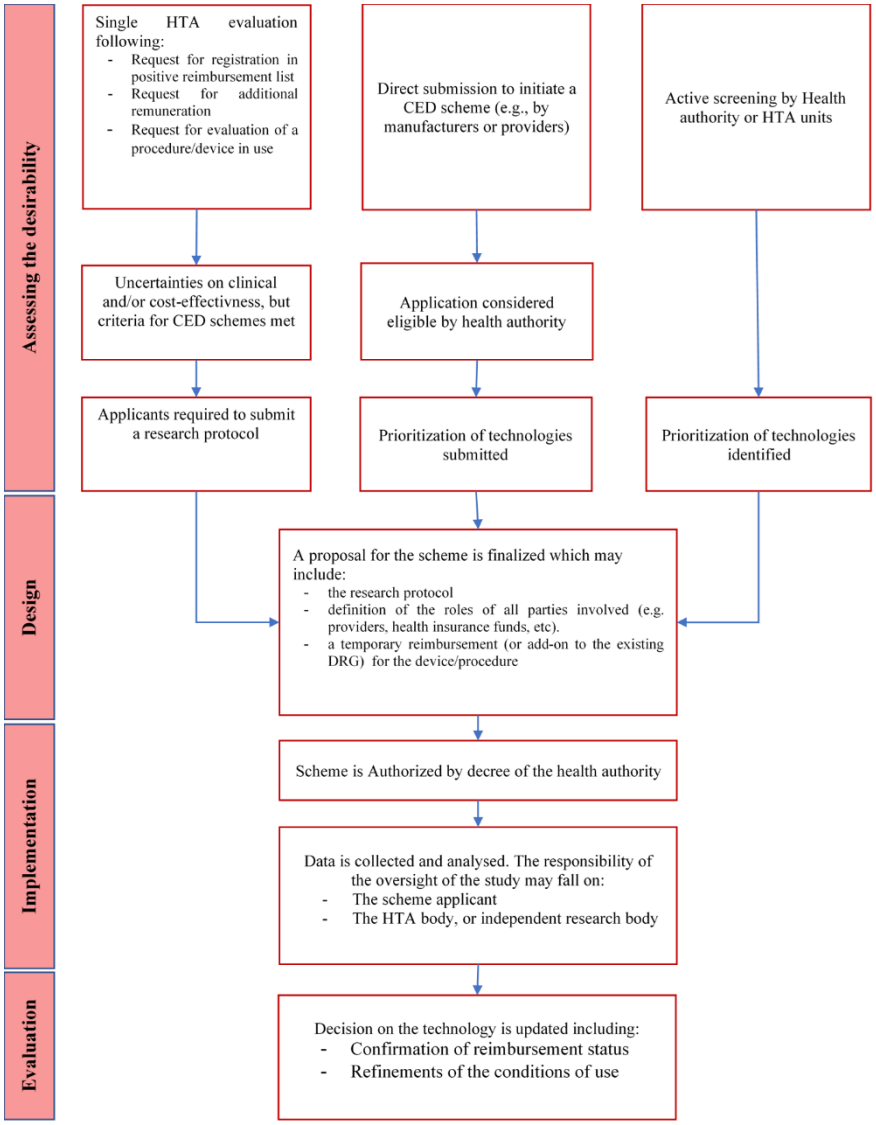
Then, these statistics were estimated separately for participants from countries with and from countries without a CED programme for medical devices. Because of the small sample sizes, the differences in scores were not assessed by performing statistical tests, but all factors which were perceived as having a positive or negative influence on each challenge were synthetized in tabular form.

Results

25 participants from 23 jurisdictions attended the interview. Respondents were from national or regional health authorities (n=15); national health insurance bodies (n=2); hospitals (n=3); and universities (n=3) (Supplementary material S5.3). Eighteen participants had high-level managerial roles related to the HTA of medical devices or services, or were responsible for the CED programme in their jurisdiction; 4 participants were technical advisers directly involved in the assessment of medical devices, and 3 were academics with an expertise in conditional reimbursement schemes. In 7 out of the 23 jurisdictions (30.4%), CED programmes existed that included (or were specific to) schemes for medical devices (i.e., Belgium, England, France, Germany, the Netherlands, Spain, and Switzerland). In France, two different programmes were identified: Post Registration Studies (PRS) for devices submitting for registration into the positive list of reimbursable products and services (LPPR list); and Forfait innovation (FI) for highly innovative technologies early in their development phase. Of the remaining jurisdictions, 5 (21.7%) operated CED programmes for drugs only (i.e., Bulgaria, Hungary, Portugal, Scotland, and Slovakia), and 11 (47.8%) did not operate any CED programmes (i.e., Austria, Czech Republic, Denmark, Finland, Greece, Ireland, Italy-Emilia Romagna Region, Italy-national level, Norway, Poland and Sweden), although some of these may have other types of PBRSAs such as performance linked reimbursement schemes (e.g., payment by results schemes). In addition, single 'one-off' experiences with schemes for specific devices were reported by participants from Emilia

Romagna Region in Italy and Ireland, in the absence of formal programmes for CED schemes for devices.

Figure 5-1 Overview of the main characteristics of CED programmes in Europe



Overall, we identified 78 CED schemes for devices which were ongoing in the last 5 years in Europe. A full overview of the characteristics of these schemes is included in on the Output page of the COMED website. Figure 5-1 and Table 5-4present an overview of how the existing national CED programmes underpinning the individual schemes address the different phases of CED schemes.

The main findings from the interviews are detailed in the following sections.

Assessing the desirability of a CED scheme

There were three main ways in which devices are selected for a scheme (Table 5-3). Firstly, a device can be selected as the direct result of a formal health technology assessment (HTA), if the decision body making the assessment identifies remaining uncertainties on the device (cost-) effectiveness and therefore propose initiation of a scheme. Such HTAs can be conducted for example, in the context of i) a request from a manufacturer to include the device on a positive reimbursement list (e.g., Belgium, France – PRS, the Netherlands and Switzerland); ii) a request from a provider for an extra remuneration of the procedure involving the device, for example on top of an existing diagnosis-related group -DRG tariff (e.g., in Germany); or iii) a request for an evaluation of a procedure or device already in use in clinical practice (e.g., in Belgium, Germany, Spain and Switzerland). Secondly, a device could be selected following an active screening of potential candidates for CED schemes conducted by the decision body or by a committee specifically appointed for this task (e.g., in England or Spain). Finally, a device could be selected following a direct application to initiate a CED scheme by manufacturers or other stakeholders, such as care providers and health insurers with an interest in the device (e.g., in Belgium, France – Forfait innovation studies, the Netherlands and Switzerland).

In all jurisdictions criteria are used to select and/or prioritize devices for inclusion in a scheme, and decisions are made either through a deliberative process or using an explicit scoring system or checklist. However, a formal assessment of the pros and cons of initiating a scheme, as opposed to other policy decisions, such as providing unconditional coverage, or refusing to adopt the device until better evidence becomes available, was never clearly defined.

Table 5-3 Overview of the characteristics of CED programmes for medical devices in Europe

	England	France	Germany	Netherlands	Spain	Switzerland	Belgium
Name of the CED Policy	**Commissioning through Evaluation**	**Forfeit Innovation (FI) / Post-Registration Studies (PRS) (études post-inscription sur les technologies de santé)**	**Evaluation of medical examination and treatment methods (§ 137e SGB V)**	**Conditional admission (Voorwaardelijke toelating)[a]**	**Postlaunch evidence-generation studies (Estudios de Monitorización")**	**Services in evaluation (Leistungen in Evaluation)**	**Limited clinical application (Beperkte klinische toepassing)**
Desirability of schemes							
Technology selection	Proposals for new schemes are co-ordinated by NHS England's CRGs during a 'Topic Selection' phase and assessed by the Clinical Panel that determines which schemes go forward for implementation.	**FI** package: proposals are submitted by manufacturers alone or in partnership with physician's associations **PRS**: if, during the assessment of a request for inscription in the LPPR, the CNEDiMTS identifies remaining uncertainties on the technology's short or long-term	During the evaluation procedure of a diagnostic and therapeutic method, if the opinion of the IQWIG reports that the benefit has not been confirmed, but the method offers the potential of being a treatment alternative. Requests for the evaluation of methods may be put forward by 1) stakeholders	Technologies can be identified in 2 ways: 1) a bottom-up process where parties can submit their own application once a year; and 2) a top down process, where the ZIN recommends, in any negative view following an assessment, whether an intervention can be eligible for conditional admission.	The DGPSPh of the MoH is responsible for the selection of topics for PLEG at the proposal of the National Commission of Provision, Insurance and Financing (CPAF)	Technologies can be identified in 2 ways: 1) following a request for verification that a medical service is effective, appropriate and efficient (WZW criteria), if during the assessment the EAMGK/CFAM A issues a "Yes in evaluation" recommendation; 2) following a direct request from manufacturers or providers for	CED schemes can be initiated top-down following a technology appraisal by the CTIIMH of the RIZIV. Bottom-up requests for the initiation of CED schemes can be submitted by scientist or participating hospitals; however, these schemes can formally only be initiated by the CTIIMH of the RIZIV.

		outcomes, it can require collection of new data through a PRS.	organizations for inpatient (§ 137c, SGB V) and outpatient (§ 135 SGB V) care, 2) directly by manufacturers (§ 137e SGB V) or 3) by hospitals, submitting a first request for NUB payment to the InEK (§ 137h SGB V).			medical devices that need to be listed under the medical device aid list.	
Criteria for eligibility and prioritization	The following eligibility criteria must be met: 1) Technology falls within NHS England's direct commissioning responsibilities 2) The treatment or care pathway shows significant promise; 3) a clinical commissioning policy is published confirming that the technology is not routinely commissioned, or there are significant	**FI:** requests are accepted if the device is expected to be innovative (4 criteria: 1) the novelty of the device, 2) early dissemination phase, 3) acceptable risk for patients, and 4) promise of significant health improvements or reduction in healthcare costs; and the protocol is considered adequate to answer the identified research questions	The new method must have positive promise of benefit, as defined in the German code of procedures: 1) potential replacement of more complex methods; 2) fewer expected side effects, 3) higher expected clinical benefits.	10 primary criteria for admissibility to a scheme (yes/no answers, all to be satisfied), and 5 secondary criteria for prioritization (score from 1 to 10). Prioritization criteria include: 1) Disease burden, 2) existence of clinical alternatives, 3) the expected added value of the intervention (health benefits/ economic/organizational/social/ethical impact) 4) existence of other	A quantitative prioritization tool is used. Criteria are defined across 4 domains: 1) Population/end users (e.g., disease burden, frequency of use); 2) Technology (innovativeness, different expectations of use); 3) Safety/adverse effects (e.g., safety issues, undetected adverse effects); 4) organization/costs and other implications (e.g.,	An explicit checklist is used for technology selection and prioritization. Main criteria are: 1) existence of a relevant evidence gap regarding efficiency, safety, cost-effectiveness and conditions of use; 2) interest for the technology (e.g., disease burden, existence of treatment alternatives, significant economic impact); 3) existence of	Main criteria used are: 1) the innovativeness of the technology; 2) feasibility of answering the identified research questions within the timeframe of the study

	remaining questions of clinical or cost-effectiveness, and/or outcomes in the routine clinical setting. 4) existing uncertainties will not be answered by current or planned clinical trials. 5) Meaningful new outcome data can be gathered within the likely timescale of a scheme (1-2 years).	(article 165.1.1 of the French social security code) **PRS:** A request for a PRS is done whenever the CNEDiMTS outlines relevant remaining uncertainties (no prioritization).		(similar) studies ongoing or planned and 5) The level of evidence of the proposed study (RCTs, observational design, comparative or not-comparative studies).	learning curve, financial impact, organizational or structural impact).	ongoing studies 4) the research proposal can answer the evidence gaps 5) feasibility of a CED scheme for the technology 6) expected positive cost-benefit ratio; and 7) capacity of the new findings to affect coverage decisions.	
Research design							
Type of CED scheme[b]	Only in research	**FI:** Only in research, **PRS:** Only with research	Only with research for Inpatient care, Only in research for outpatient care	Only in research[c]	Only with research	Only with research	Only in research
Types of study design	Mainly prospective observational studies using data collected from existing clinical databases, or by setting up a new registry.	**FI:** Highest level of evidence preferred (e.g., RCTs). **PRS**: Mainly single-arm, registry based, observational studies.	Highest level of evidence preferred (e.g., RCTs).	Highest level of evidence preferred (e.g., RCTs). Furthermore, a supplementary (observational) study may be initiated after the enrolment of the preferred study is completed.	Prospective, observational studies; focus on designs which minimize data collection effort.	Preferably RCTs, other designs may be also considered (before-and-after comparisons, case series or comparisons with historical controls).	Prospective observational studies, based on registry data.

Implementation							
Funding of the research	NHS England provides funds for service provision to the participating centres and NICE to oversee the scheme. NICE directly commissions an External Assessment Centre for data collection and data analysis.	**FI**: a forfeit payment for the procedure and the associated hospital costs is established at the start of the scheme. Costs of data collection and analysis fall on the scheme applicant. **PRS:** Following the CNEDiMTS appraisal, the device is temporarily listed in LPPR and covered by the social health insurance. Costs of data collection and analysis fall on the manufacturers.	G-BA coordinates all phases of the design and implementation of the scheme. The diagnostic and therapeutic method under evaluation is covered by the health insurance. Overheads of the study can be financed by the manufacturer of the device being evaluated or are financed by statutory health insurance via G-BA	The care provided is covered by the basic insurance package. The reimbursement rate is negotiated between the health insurance companies and the participating hospitals and included in a covenant agreement signed by all parties involved in the scheme. The costs of data collection and analysis are covered by the scheme applicants. However, there is the possibility to apply for a research grant at ZonMw.	Regional HTA agencies receive funding for data collection, analysis and reporting from the central Ministry of Health. The price of the device is negotiated individually by the regions participating in the scheme. Participating hospitals do not receive extra funding for data collection	The reimbursement of the procedure is covered by the health insurance. Costs of data collection and analysis falls on the manufacturer	Following the recommendation of the CTIIMH to initiate a scheme, the minister of health takes a decision regarding the temporary reimbursement of the care service and the reimbursement methods to be applied. Participating hospitals do not receive any funding for data collection and analysis.
Definition of study protocol	The study protocol is developed by the External Assessment Centre in partnership with NICE.	**FI:** The study protocol is directly submitted by the scheme applicant and evaluated by the HAS	The key aspects of the study are defined in the directive approving the scheme. The protocol is then	Development of the study protocol is a direct responsibility of the scheme applicant. ZIN	The study protocol is defined by the regional HTA agencies participating in the data collection.	The design of the protocol is a responsibility of the scheme applicant. The proposal is then	The relevant questions to be answered in the scheme and the set-up of the registry are

		PRS: The responsibility of defining the protocol falls on the scheme applicants. Authorities only provide broad indications on the type of uncertainties that must be addressed by the scheme, and approve the final version of the study protocol	refined by the research institution that conducts the study.	assesses the study proposal in collaboration with the the Scientific Advisory Council (WAR) and ZonMw.		evaluated and approved by the FOPH	proposed by the CTIIMH and discussed with the stakeholders involved, in order to obtain an agreement. Outcomes to be considered are discussed between the expert scientific community and the CTIIMH which also approves the final version of the protocol.
Data collection, monitoring. and analysis	The data collection is overseen by the appointed steering group, and supported by the External Assessment Centre. Periodic checks and follow ups are done on the quality and validity of data submitted to ensure meaningful data is being collected. Analysis of the data is done	**FI and PRS:** The responsibility for both the data collection and analysis falls on the manufacturer only. For PRS studies, the CNeDMTs evaluate the quality of the new evidence provided at the time of the planned re-appraisal of the technology.	Data collection and analysis are done by an external research institution which has been appointed by G-BA through a public tender, if the overheads are financed via G-BA.	The scheme applicant has the main responsibility for data collection and monitoring. The ZIN monitors the course of the scheme and reports it annually to the Minister of Health. ZIN assesses the new evidence provided at the time of the planned assessment of the technology	Data collection and analysis is coordinated by the Regional HTA agencies participating in the scheme.	The applicant (provider and/or manufacturer) are the sole responsible for data collection and analysis. Yearly reports have to be reported to the FOPH, showing how the study is proceeding. These reports may inform changes to the scheme or even cause early termination, if	Data collection is a responsibility of the hospitals that have signed the agreement to participate in the scheme. Depending on the agreement, the hospitals or an external peer-review group/scientific association are responsible for the scientific report.

	by the external assessment centre and reviewed by NICE					issues arise with data collection (e.g., poor quality of the data, slow recruitment)	
Evaluation							
Existence of ex-ante decision rules for the scheme linking the results of the scheme to a decision on price, reimburse ment or use	No	No (both FI and PRS)	No	Agreements regarding the uptake of the intervention, in case of a positive coverage decision at the end of the scheme, or exit strategies in case of a negative opinion (e.g., because the intervention is not effective, or the data quality is considered insufficient to take a decision) are defined in the convenant agreement prior to the start of the scheme.	No	No	No
Types of decisions informed by the scheme	Results of the scheme are used for the development of Clinical Commissioning	**FI**: Conditional reimbursement is provided only for the duration of the scheme, then devices enter usual	Confirmation of the reimbursement status	Confirmation of the reimbursement status	Confirmation of reimbursement status and refinements of conditions of use	Confirmation of the reimbursement status, refinements of conditions of use, and changes in the maximum	Confirmation of the reimbursement status

	policy for NHS England's directly commissioned specialised services.	evaluation pathways (e.g., a new request by the manufacturer for inscription in the LPPR). **PRS**: confirmation of the device in the LPPR and refinements of the conditions of use. Financial penalties on the price of the device may be applied in case of poor data quality at reassessment.				reimbursement rate for the technology or procedure.	

[a] Starting from 2019, conditional admission schemes have started to be gradually replaced by schemes within the new Promising Care Subsidy Fund (PCSF). The main difference between the two programmes concerns the way care provision is reimbursed during a scheme, i.e., directly subsidized in the PCSF rather than covered by the statutory health insurance as in conditional admission schemes. The schemes already ongoing will be completed according to the VT programme described in the Table. [b] "Only in research" are defined as schemes in which a device or procedure is reimbursed only for patients who enrol in a clinical study, whereas "Only with research" schemes are defined as schemes in which a device or procedure is reimbursed for all indicated patients while data are collected in a subset of patients. [c] Conditionally approved care is only covered if the patient participates in the main study. However, patients who are not able to participate can claim the conditionally approved care if they participate in a supplementary ancillary study. CNEDiMTS, French Medical Device and Health Technology Evaluation Committee; CRGs, Clinical Reference Groups (England); CTIIMH, Belgium Implant and Invasive Medical Device Reimbursement Committee; FOPH, Swiss Federal Office of Public Health; G-BA, German Federal Joint Committee; InEK, German Institute for the Hospital Remuneration System; IQWiG, German Institute for the Hospital Remuneration System; LPPR , List of Products and Services qualifying for Reimbursement (France); RCTs, Randomized Controlled Trials; RIZIV, Belgian Medicines Verification Organisation ; ZIN, Netherlands National Health Care Institute ; ZonMW, Netherlands organisation for Health Research and Development, Implant and Invasive Medical Device Reimbursement Committee

Designing a CED scheme

There were differences in the design of schemes between countries. For example, Spain and Switzerland mainly operated schemes in which a device is reimbursed for all indicated patients while data are collected in a subset of patients (i.e., only *with* research - OWR), whereas England, the Netherlands and Belgium mainly operated schemes in which a device is reimbursed only for patients who enrol in a clinical study (i.e., only *in* research – OIR). In France schemes were either OIR in the FI programme or OWR in the PRS, whereas in Germany the type of schemes depended on whether the technology was intended for inpatient use (OWR) or outpatient use (OIR).
It is worth noting, however, that within countries, the designs were relatively similar between schemes and appeared not to be tailored to (the specific characteristics of) the device under evaluation or to key sources of uncertainty. Moreover, the study designs were similar between schemes in each country and mainly concerned either observational designs utilizing real-world data, or experimental designs to ensure a high level of evidence (e.g., randomized controlled trials).

Implementing a CED scheme

There were differences in the governance of CED schemes as well as in the roles and responsibilities of the stakeholders involved, regarding the development of the research protocol and the subsequent monitoring of the scheme. Overall, two main approaches were identified. In the first approach (e.g., in France, the Netherlands and Switzerland), the responsibility for the development of the study protocol, the monitoring of the scheme and the quality of the generated data relies entirely on the scheme applicants (e.g., the manufacturer or care providers). However, in defining the study protocol the applicants typically must follow the recommendations of the relevant decision body. Usually, the protocol requires formal approval before study initiation, to make sure that it is suitable for addressing the identified uncertainties regarding the device. In the second approach, the responsibility for the development of the study protocol and the quality of the generated

data are coordinated centrally (e.g., by HTA agencies), and managed either directly or through third-party research centres (e.g., in Belgium, England, Germany and Spain).

Patient representatives may be involved in the initial assessment phase on the desirability of the scheme (e.g., in Spain or France) or later during the recruitment phase of the study (e.g., in England), but their involvement in the design phase and the development of the protocol was generally limited. During the scheme, the costs of care provision (including utilization of the device) are usually funded through the public health care system. Specific funding arrangements may be defined at the onset to cover the additional costs of the device or procedure, by either establishing a forfeit or negotiating an add-on to an already existing DRG tariff. However, different arrangements exist for covering the additional costs associated with the research, including the costs of developing the study protocol, scientific monitoring, data collection and analysis. These costs may be either entirely financed with public funds (e.g., in Belgium, England and Spain) or they may be partially or entirely covered by the scheme applicant (e.g., in France, Germany, the Netherlands and Switzerland). Notably, in some cases, funding arrangements also include resources for data collection. In other cases, health care providers are required to perform this task without any additional funding, for example, as a condition of participating in the scheme and gaining market access for the device (e.g., in Spain or Belgium).

Evaluating a CED scheme

Decisions at the end of schemes mainly concerned the confirmation of the reimbursement status of the device, the refinement of clinical indications or conditions of use. For most of the identified schemes, no ex-ante decision rules that explicitly linked the scheme results to future decisions were defined. In fact, in most countries the schemes solely concerned the collection of additional evidence to reduce the identified uncertainties, while the final decision on the reimbursement, coverage or use of the device was integrated in the routine decision-making framework. A notable exception

was the Netherlands, where the level of effectiveness that must be demonstrated during the scheme in order to obtain unconditional reimbursement was predefined at the onset of the scheme, in a covenant agreement signed by all stakeholders. Moreover, the covenant also addressed how to manage the withdrawal of a device in case it proved to be insufficiently effective or the data did not allow an informed decision (e.g., due to poor data quality or inconclusive results).

Notably, all participants reported having no, or only very little experience, with schemes that led to a negative coverage decision. Indeed, of the 24 CED schemes for which information on final decisions were available, coverage was confirmed (or conditional coverage prolonged due to data quality issues) in 22 cases.

Challenges associated with CED schemes for Medical Devices

Of the 25 participants, 18 scored the 13 challenges on the six-point Likert scales. Of these, 9 were from jurisdictions with CED programmes involving devices, and 9 were from jurisdictions with CED programmes involving drugs only. The 7 participants who did not score the challenges were from countries without CED programmes. For most of the assessed challenges, scores were observed to span across the full range of the Likert scales, indicating no clear patterns in the decision-makers' perceptions. presents the mean and median scores for each challenge (Table 5-5).

Overall respondents from jurisdictions with CED programme for medical devices tended to give lower scores to most of the challenges as opposed to respondents from jurisdictions without such programmes. However, the low sample size and the variability in responses within each challenge hampered any firm conclusion.

Supplementary material 5.4 presents the main factors that, according to the participants, positively or negatively influenced the challenges. Many of the factors identified are common to all technologies and consistent with the existing literature on CED schemes. However, some elements specific to devices could be identified. Devices were generally considered to be more

Table 5-4 Assessment of challenges by participants[a]

		Participants from countries with CED programmes for medical devices - (Belgium, England[b], France[b], Germany, Netherlands, Spain, Switzerland)			Participants from countries without CED programmes for medical devices - (Bulgaria, Hungary, Ireland, Italy[b], Poland, Portugal, Scotland, Slovakia)		
	Challenge	**n**	**Mean (SD)**	**Median (IQR)**	**n**	**Mean (SD)**	**Median (IQR)**
1	Deciding which medical devices are candidates for CED schemes	9	2.5 (1.17)	2 (2.25)	9	3.78 (1.48)	4 (2.5)
2	Obtaining stakeholder agreement on the scheme	9	2.17 (1.46)	2 (2.75)	8	2.75 (1.83)	2.5 (3.5)
3	Securing funding for the scheme	9	0.89 (1.05)	1 (1.50)	8	3 (1.69)	3 (3.5)
4	Determining the appropriate study design for data collection	9	2.39 (1.45)	2 (2.75)	9	3.33 (1.32)	4 (2)
5	Determining the relevant outcome measure(s) on which data are collected	9	2.61 (1.27)	2 (2.50)	9	2.78 (1.72)	2 (3.5)
6	Dealing with data collection and monitoring	8	2.13 (1.64)	2.5 (3.5)	9	3.78 (1.2)	4 (2.5)
7	Dealing with data analysis	9	1.61 (1.22)	1.5 (2.5)	8	3 (1.51)	3.5 (2.75)
8	Ex-ante definition of decision rule, based on possible outcomes of the scheme	3	3 (1)	3 (2)	8	3.75 (1.58)	4.5 (2.75)
9	Reaching an agreement on price, reimbursement or use of the device at the end of the scheme	5	2.1 (2.13)	2 (4.25)	7	3.57 (1.27)	4 (3)
10	Withdrawing a device from the market when evidence indicates the device is not (cost-) effective	6	3 (0.89)	3 (2)	8	4.5 (1.07)	5 (0.75)
11	Obtaining agreements about the duration of the scheme and the stopping rule	9	1.94 (1.13)	2 (1.25)	8	1.75 (1.49)	1.5 (2.75)
12	Adapting the scheme to account for product modifications or a learning curve	8	1.44 (1.45)	1.5 (2.38)	8	3.25 (1.49)	3.5 (2.75)
13	Dealing with the market entry of similar devices	9	1.83 (1.73)	1 (2.75)	8	2.25 (1.67)	2 (3)

[a] Assessed on a six-point Likert scale (ranging from 0 "not a challenge" to 5 "a major challenge"). [b]Two participants scored the challenges for this country

difficult to identify and monitor than pharmaceuticals, given that their routes to market are often less clear and may not be observed by those who are responsible for selecting potential candidates for CED schemes. The intrinsic characteristics of devices were also reported to pose additional challenges in the design and implementation of schemes. For example, device-user interactions and the context-specific factors which may affect device performance in the real-world were considered as challenges for the identification of all relevant uncertainty at the time of scheme initiation, and for the definition of the study protocol.

In addition, devices may be associated with uncertainties that cannot be easily resolved within a feasible time frame for a scheme, such as uncertainties over the devices' durability or their long-term performance in patients with different clinical conditions and physiologies. This in turn may increase the tension between the need to pragmatically rely on surrogate endpoints, which are rarely validated for MD procedures, and the relevance of the data collected to inform decision-making at the end of the scheme. In addition, routinely collected data, such as administrative datasets or electronic health records were expected to be less often available, or relevant, for devices, as compared with pharmaceuticals.

Relating to the possibility of product modifications during the timeframe of the scheme, one of the main concerns related to the fact that such modifications could bias the results of the study or compromise the relevance of the new evidence collected. In this respect, being able to anticipate product modifications by means of dialogues with manufacturers and sharing of information was considered a potentially mitigating factor. However, the possibility of product modifications was not perceived by most of the respondents as a major challenge, or something which is likely to occur during the duration of a scheme.

Similarly, about half of the respondents did not consider the possibility that similar products would enter the market during the period of the scheme to be an important challenge. Possible reasons related to the fact that most of the schemes evaluate a class of devices or a procedure rather than a single

branded device, or that, even if focused on a single product, they collected mainly non-comparative data. However, other respondents emphasised the difficulty of anticipating which products would enter the market during the schemes and the possibility that relative effectiveness estimates may not be meaningful anymore by the end of the scheme, as clinical practice changes more rapidly in the context of devices compared to pharmaceuticals.

Finally, with respect to the existence of a learning curve, interviewees acknowledged it as a challenge which affects both the collection and analysis of data, as well as the design of the study, such as deciding on the number of clinical centres authorized to use the device as part of the scheme. However, direct experience with this aspect was generally limited across all respondents.

Discussion

CED schemes and their application to medical devices are important items on the policy and research agendas. The objectives of this chapter were to explore the characteristics and use of CED schemes for devices in Europe, as well as the challenges that decision-makers face when designing and operating these schemes. This chapter importantly adds to the existing knowledge base by providing a comprehensive and multi-country overview, which was directly informed by surveys with European decision/makers.

There were 72 device-related CED schemes that have been operated over the last five years in European countries. However, only 7 countries (out of the 22 included) had CED programmes in place for medical devices. To a large extent, this result may reflect the uneven application of HTA within Europe, since it may be difficult to develop a policy for CED schemes without having an established HTA capacity. For example, deciding that more data are required post-launch implies that some form of assessment of clinical or cost-effectiveness has been made. Nevertheless, HTA capacity cannot fully explain these differences, since CED schemes seem to be less frequently used for devices than for drugs (15).

The characteristics of the identified CED programmes underpinning the individual schemes for devices varied between countries, which may reflect local differences in how HTA is organised and practised. For example, schemes were either initiated by the authorities (i.e., Ministry of Health), often as a consequence of the findings of an HTA for the technology, or as a response to a request from a manufacturer. There were similar patterns in the relative responsibilities for the funding of schemes and the design of study protocols although the authorities always played some role in study design, either by outlining a general specification or recommending that an independent research centre be involved. These differences in roles were also found in the aspects of the implementation of schemes, including the collection and analysis of data, which was sometimes the responsibility of the manufacturer and sometimes an independent party.

One aspect that deserves attention is how devices are selected for a scheme. Indeed, CED is not a costless activity and its (opportunity) costs and benefits should be considered alongside other policy options, such as adopting or refusing adoption of the technology, based on currently available data, or negotiating a lower price. Aspects to be considered should include: 1) the expected value of research option(s) in terms of reduced uncertainty; 2) the direct costs of collecting evidence; 3) the opportunity costs of any delay in providing access to the technology because of the scheme; and 4) the existence of any irreversibility in the process (e.g. difficulty to subsequently withdrawal the technology, or difficulty to conduct further research after conditional approval) (207,208). However, while all the identified programmes used criteria to identify and prioritize technologies for a scheme, a formal assessment of these aspects was generally missing. Related to the previous point. In many jurisdictions, there does not seem to be an option for choosing among different types of CED schemes, such as OWR and OIR schemes. Nonetheless, also depending on characteristics that are specific to, or particularly relevant for devices (e.g., the existence of irreversible upfront investment costs), there may be cases where either one or the other type of CED scheme would be optimal (208,209). As reported

in the recent report from the ISPOR good practice Task Force, Value of Information (VOI) analysis may be used to support formal assessments on the opportunity to initiate a CED scheme and the type of scheme which maximizes optimal allocation of healthcare and research funds (50).

In addition, one general finding across all countries was that relatively little attention seemed to be paid to the evaluation of schemes, both *in itinere* during data collection and at the time of the reassessment of the technology once the scheme reported its results. This mirrors the findings of other studies of CED and market access schemes more generally (207,210,211) and is obviously an area that requires further attention by policy makers and researchers. Indeed, issues with the quality and timely reporting of data have been mentioned as a factor hampering CED schemes (Supplementary material S5.4). For example, in France, where manufacturers are solely responsible for the collection of additional data, the lack of the requested evidence from post-registration studies was often reported in the technology re-appraisals.

The policy responses at the end of a CED scheme for devices may be more complicated than, for example, deciding on whether to include a drug on a formulary or to determine prescribing guidelines, since the reimbursement of devices, and the policies to determine their use, are often linked to the use of broader surgical, or other treatment, interventions. Therefore, policies probably involve adjustments to DRG tariffs, or changes to clinical guidelines, and/or hospital practice more generally. Hence, decision rules and policies for discontinuing the use of devices require attention in this context.

Notably, all participants reported to have no or little experience with refusing to confirm reimbursement at the end of the schemes. While this may reflect the degree and type of uncertainties existing at the beginning of the schemes, it may also signal a certain difficulty in reversing the preliminary reimbursement decision once a technology has entered a scheme (1,212,213). This aspect may be even more relevant if no ex-ante criteria for evaluating

the schemes were defined, as was the case for almost all schemes for devices in Europe.

Based on the observations of variation in the characteristics of schemes, it is difficult to prescribe a single preferred approach to CED of devices in Europe. Each country has specific local differences in HTA practices, although knowledge on how CED schemes have been used elsewhere can be used to develop local guidance. However, ideally a primary driver of the initiation of CED schemes would be the outcome of HTAs for the technologies concerned, since this can help identify the uncertainties in (cost-) effectiveness that (in principle) could be resolved through CED. The participants' perceptions of the various challenges in initiating, designing, implementing, and evaluating CED schemes were varied and did not indicate that, in general, some challenges were substantially more important than others. The reasons for this are unclear, although in some cases the participant's perception of a given challenge reflected local circumstances. For example, funding was not perceived as a major challenge in settings where funding was made available, but a major challenge in settings where it was not. In addition, the scores obtained for those challenges that were 'device specific' did not differ substantially from those for the other, more generic challenges. While this aspect requires further investigation, our general impression was that some of the low scores given for 'device specific' challenges are attributable to a lack of direct experience with addressing these issues, given that the use of CED schemes for medical devices in some European countries is generally quite recent.

Overall, one aspect that can be outlined is that the practice of CED in Europe often departs from the theoretical models for CED described in sections 2.5 and 2.6 of this thesis. Also the use of VOI as a tool to inform the need and design of CED (§2.4) is rarely applied in practice.

Deviation from theory may lead to sub-optimal decisions on coverage and evidence generation, if coverage and/or research decisions are not fully considering all the opportunity costs of available policy options. Nonetheless, it may also be that the existing methodological tools are not

suitable to the specific regulatory framework in some countries (e.g., in terms of legislation, possibility to mandate research or types of conditional reimbursement schemes available), thus requiring country-specific adaptations. A set recommendations based on the work developed in this thesis is proposed in Table 5-6. We used a combination of methods to obtain insights in the use of and challenges related to CED schemes in the relatively understudied context of devices, including a large set of European countries. The insights obtained allow learning from experiences across countries and increase the chances of having successful CED schemes in the future, by highlighting how decision makers perceive and deal with specific challenges.

Table 5-5 Recommendations to policy-makers on the design and implementation of CED schemes

Recommendation	Brief explanation
1) Define the purpose of the CED scheme in terms of the uncertainty to be resolved	The purpose of CED to reduce uncertainty over effectiveness or cost-effectiveness of a new technology should always be considered regardless of how schemes are initiated (e.g., via formal HTA appraisal, or bottom-up from hospitals and/or manufacturers).
2) Apply VOI, or at least VOI principles	If not directly used, VOI principles should be applied at least informally to inform the need to initiate a scheme. Decisions should be based on the comparison between the expected benefits in terms of reduced uncertainty and the costs of conducting new studies.
3) Study design should reflect nature of the uncertainty	the study design for CED schemes. should reflect the nature of the uncertainty surrounding the effectiveness or cost-effectiveness of the device and practicalities of conducting the research. Experimental or observational designs should be assessed based on their capacity to reduce the relevant uncertainty
4) Balance scientific and practical considerations when determining the length of CED schemes	The length of the CED should be determined by a balance of scientific and practical considerations. The capacity for the scheme to achieve its aims within a reasonable time frame needs to be carefully evaluated. Also, the likelihood that the evidence generated through the study is still relevant to inform the decision problem at the end of the data collection must be carefully assessed.
5) Define decisions to be taken at the end of the CED scheme as early as possible	Thought should be given to the decisions that will be taken at the end of the scheme (e.g., relating to the reimbursement, coverage, or price of the device), as well as the mechanism for implementing these decisions in the health care system

Recommendation	Brief explanation
6) Provide solid reasons when deviating from common CED principles	Considerations made to inform the different stages of CED schemes (initiation, design, implementation and evaluation) should be carefully and transparently described, and any deviation from the best-practices identified in the literature should be explained

Nonetheless, some limitations also need highlighting. First, although many European countries were considered, it may be that this overview is incomplete as some countries were not included in the study. Moreover, in each country it is uncertain whether the views of the participants are representative of the views of decision-makers more generally. Additionally, we focussed on the detailed perceptions of decision-makers, with a focus on HTA agencies at the national or regional level and (some) national payers because recent research suggests that decision-makers may be hesitant to engage in CED schemes (214). This makes them not only a relevant source for the current study in terms of knowledge, but also in articulating (potential) challenges and difficulties with applying such schemes. Future studies could nonetheless supplement this with information on the perceptions of other stakeholders, such as clinical professionals, patient organisations, local payers/decision makers, and manufacturers. Finally, the focus of this chapter was on schemes initiated at the national or regional level. In addition, some schemes involving devices may be negotiated at the local level directly between providers and manufacturers. Many of these may be 'pay for performance' schemes, but some could be characterized as CED schemes. These schemes were outside the scope of our current study, but their characteristics and performance are nonetheless important to investigate further.

Chapter 6
Conclusions and future work

This thesis approached the problem of uncertainty from a dual perspective: a more methodological perspective and a policy perspective. Far from being conclusive and complete, it is hoped that the approaches proposed have contributed to expand the set of tools available to address decision-making uncertainty in healthcare.

These approaches are consistent with a lifecycle approach to technology evaluation in which uncertainty is addressed iteratively (and differently) as long as new evidence accumulates.

While the concept of lifecycle evaluation is not new (215), it has found renewed vigour in recent years, in response to the challenges of innovative and novel technologies, disconnected and disparate stakeholders and data requirements, public expectations and achieving patient-centric health systems (72).

Promoting an iterative and interactive approach to technology evaluation at early stages of a product development can contribute to two very relevant and linked achievements. First it can ensure that the societal relevance of new technologies (in terms of effectiveness and cost-effectiveness) is incorporated earlier in the developers' strategic choices on portfolio management and go-no go decisions. Second it can make evidence generation processes more streamlined, timely and fit for purpose, thus allowing regulators and payers to make more informed assessments and ultimately to quickly provide access to safe and (cost-)effective technologies to patients.

The recently enforced medical device regulation (MDR) (74) and the approved proposal for a regulation on cooperation in HTA among member states (216) introduce elements that go in the direction of a lifecycle approach

to technology assessment. The MDR explicitly envisages the possibility for manufacturers of class III and certain class IIb devices to consult an expert panel with the aim of reviewing the manufacturer's intended clinical development strategy and proposals for clinical investigation. Similarly, the proposal for HTA cooperation envisions the possibility for health technology developers to engage in early dialogue with European HTA bodies as one of the four pillars of the cooperation among EU Member States on this matter. While regulatory and HTA early dialogues for medical devices are still conceived as separate moments, interest from multiple stakeholders exists over the possibility to conduct parallel early dialogues such as the ones already in place for pharmaceuticals, and involving both the European Medicine Agency (EMA) and multiple HTA agencies at the same time (8).

Another pillar of the cooperation in HTA concerns the joint activity of horizon scanning that has the aim to identify earlier emerging technologies that will likely have an impact on the European healthcare systems.. From an HTA perspective, the novelty elements introduced by the two regulations will raise awareness about new medical devices much earlier and may imply performing assessments on their clinical and economic value soon after, or even before market access. While desirable from a lifecycle perspective and a more rational, regulated introduction of medical devices in clinical practice, early assessments will have inevitably to face with much higher uncertainty, which in turn has strong political and policy implications.

It is expected that transparency and formal analytical frameworks will be required to ensure consistency and fairness, and consequently that the methodological approaches used to characterise and assess uncertainty and translate it into policy decisions will become even more relevant than they already are today.

In this changing environment developers must improve their portfolio strategies and choose only the technologies that have a chance to make it to the market. The headroom approach is a first rapid way for a preliminary assessment of the potential value of a technology. Also, a VOI analysis as the

one proposed in this thesis, that uses a specific developer's utility function, while acknowledging the approval processes and constraints by regulators, payers and other decision-makers in healthcare (e.g., clinicians) will support them in deciding whether the investments in evidence that are required to make it to the market are worth the money, in terms of positive return on investments. The estimation of VOI in presence of multiple constraints is quite recent and few studies have addressed it from either a theoretical or applied stance. Yet consideration of such constraints in calculating VOI may be important especially in settings where decision-making processes depart from the single-criterion decision-rule which is assumed in CEA and the estimation of VOI, as is often the case for medical devices and the developers' perspective.

The use of PBA could be relevant at the very early stage of product development as it consistently and simultaneously addresses uncertainty around parameters for which a probability distribution cannot be reliably estimated and parameters where parametric assumptions fail to truthfully characterize ignorance over their value.

PBA may be also used in very early assessments for example during an Horizon Scanning assessment, or when ignorance exist on some aspects of a technology that are likely to affect their cost-effectiveness profile. For example, one possible application for PBA during HTA appraisals is in cases when there exists uncertainty over parameters that affect the long or very long-term performance of a device (for example a device revision rate, or the duration of the battery for active implantable devices). In this case, only observational data collected over many years of clinical practice may reduce the existing uncertainty and PBA can be used to explore the costs and consequences of the worst- and best- case scenario in which uncertainty can resolve. These types of analysis, coupled with a higher interaction with decision-makers through early dialogues, may contribute to identify better products and generate the appropriate evidence to demonstrate their value.

Finally, the use of CED schemes may also become more important since it is likely that for promising technologies identified earlier in their lifecycle,

conditional reimbursement will be granted to allow for further evidence generation. Chapter 5 has revealed a widespread use of CED schemes for devices but has also outlined how practical implementation often departs from the theoretical foundations underlying these schemes. It is hoped that the findings of this chapter and the recommendations provided will be helpful for decision-makers willing to introduce or modify CED programmes in their jurisdictions.

Multiple streams of research can be identified starting from what has been presented. For VOI analysis, future work could explore and validate more refined utility functions for technology developers and identify relevant data sources to estimate them. Also, more sophisticated methods could be employed to predict market dynamics and other constraints faced by developers after regulatory and HTA approval. The degree of acceptance and understanding of VOI techniques among technology developers is still unknown and deserves more attention.

In addition, regardless of the utility function used, the estimation of the EVPPI and EVSI in presence of additional constraints like the one considered in the reported use-case implies additional technical challenges that have not been addressed here or elsewhere.

The relevance and acceptability of PBA to inform decision-making of both developers and payers still need to be verified through more qualitative and applied research. In addition, the technical implications of using PBA on the decision rules to be applied and the consequences in terms of technical efficiency and opportunity costs also need to be better understood and will be the focus of further research. For example, the implications of using the Hurwicz criteria at different stages of the product lifecycle, and its acceptability according to different perspectives, need to be better understood before recommending this tool for decision-making. The use of VOI alongside PBA also requires further investigation. While intuitively appealing, no research has been conducted on how VOI can be estimated in cases when, at least for a subset of parameters, no parametric assumptions are used.

Finally, there is still a paucity of evidence on whether CED schemes can achieve their planned objectives of simultaneously informing adoption and research decisions in practice. The use of VOI to inform the initiation and design of CED is key, but the work presented here has made evident that VOI is never considered explicitly in any of the countries with CED programmes in place. In commenting on the lack of adoption of VOI analyses by research funding bodies, Fleurence and Selby (217) point to the computational burden of conducting analyses on a range of competing research projects. However, new methods have been developed that provide quick regression based alternatives to VOI estimation (61). In the context of CED, it may be that decision-makers do not only view schemes as being important for gathering the necessary information to make an optimal resource allocation decision, but also as part of a broader negotiation with manufacturers concerning the pricing and reimbursement of the new health technology. In that sense, one could argue that while the rationale and health economic background for having a CED may relate to determining what the optimal decision would be, more attention could be given in future work to the political, behavioural and institutional aspects of reimbursement and pricing decisions, and the incentives regarding whether and how decisions can be informed and implemented. Both elements appear to be crucial in real-life . In this respect, the suitability of existing VOI approaches to the regulatory framework of individual countries will also have to be explored further.

This work has approached decision making under uncertainty from different perspectives, focusing primarily on the early stages of a technology's life cycle and covering both methodological and policy aspects. While methodological improvements can help to better inform decisions, any proposed new approach must be validated by decision makers to verify its appropriateness to meet their information needs. Indeed, only by operating this continuous dialogue on methods and decision-makers need, research on HTA can maximize its impact and eventually contribute to healthier societies.

Bibliography

1. Reckers-Droog V, Federici C, Brouwer W, Drummond M. Challenges with coverage with evidence development schemes for medical devices: A systematic review. Health Policy Technol. 2020 Feb;S2211883720300137.

2. Pirc E, Federici C, Bošnjak M, Perić B, Reberšek E, Pecchia L, et al. Early cost-effectiveness analysis of electrochemotherapy as a prospect treatment modality for skin melanoma. Clin Ther. 2020;42(8):1535–48.

3. Federici C, Armeni P, Callea G. A Value-based Revolution in Health Care: Perspectives, Challenges, and Emerging Approaches to Defining and Measuring the Value of Health Care Technologies. Clin Ther. 2020;42(1):11–4.

4. Fornaro G, Federici C, Rognoni C, Ciani O. Broadening the Concept of Value: A Scoping Review on the Option Value of Medical Technologies. Value Health. 2021 Jul;24(7):1045–58.

5. Federici C, Reckers-Droog V, Ciani O, Dams F, Grigore B, Kaló Z, et al. Coverage with evidence development schemes for medical devices in Europe: characteristics and challenges. Eur J Health Econ. 2021 Nov;22(8):1253–73.

6. Federici C, Pecchia L. Early health technology assessment using the MAFEIP tool. A case study on a wearable device for fall prediction in elderly patients. Health Technol. 2021 Sep;11(5):995–1002.

7. Federici Carlo, Torbica Aleksandra. Expanding the role of early health economic modelling in evaluation of health technologies. Comment on

'Problems and promises of health technologies: the role of early health economic modeling'. Int J Health POLICY Manag. 2021;10(2):102–5.

8. Blankart CR, Dams F, Penton H, Kaló Z, Zemplényi A, Shatrov K, et al. Regulatory and HTA Early Dialogues in Medical Devices. Health Policy. 2021 Jul;S0168851021001937.

9. Ciani O, Federici C. Introduction to Health Economics and HTA. In: Iadanza E, editor. Clinical engineering handbook. 2nd edition. San Diego: Elsevier Inc; 2020.

10. Castaldo R, Federici C, Pecchia L. Early Stage Healthcare technology assessment. In: Iadanza E, editor. Clinical engineering handbook. 2nd edition. San Diego: Elsevier Inc; 2020.

11. Drummond M, Federici C, Reckers-Droog V, Torbica A, Blankart CR, Ciani O, et al. Coverage with evidence development for medical devices in Europe: Can practice meet theory? Health Econ. 2022 Feb 26;hec.4478.

12. Federici C, Pecchia L. Exploring the misalignment on the value of further research between payers and manufacturers. A case study on a novel total artificial heart. Health Econ. 2022 Apr 23;hec.4520.

13. Iskandar R, Federici C, Berns C, Blankart CR. An approach to quantify parameter uncertainty in early assessment of novel health technologies. Health Econ. 2022 May 17;hec.4525.

14. Siverskog J, Henriksson M. Estimating the marginal cost of a life year in Sweden's public healthcare sector. Eur J Health Econ. 2019 Jul;20(5):751–62.

15. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press; 2005. 404 p.

16. Ng YK. Welfare Economics [Internet]. London: Macmillan Education UK; 1983 [cited 2021 Aug 12]. Available from: http://link.springer.com/10.1007/978-1-349-06612-4

17. Williams A, Sugden R. The principles of pratical cost-benefit analysis. Repr. Oxford: Univ. Pr; 1988. 275 p.

18. Sloan FA, editor. Valuing health care: costs, benefits, and effectiveness of pharmaceuticals and other medical technologies. Repr. Cambridge: Cambridge Univ. Press; 1998. 273 p.

19. Neumann PJ, Ganiats TG, Russell LB, Sanders GD, Siegel JE, editors. Cost-Effectiveness in Health and Medicine [Internet]. Oxford University Press; 2016 [cited 2021 Aug 12]. Available from: https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780190492939.001.0001/acprof-9780190492939

20. Drummond M. Methods for the economic evaluation of health care programmes. Fourth edition. Oxford, United Kingdom ; New York, NY, USA: Oxford University Press; 2015. 445 p. (Oxford medical publications).

21. Cookson R, Claxton K eds. The Humble Economist: Tony Culyer on Health, Health Care and Social Decision Making [Internet]. Monographs. Office of Health Economics; 2012 [cited 2022 May 11]. Available from: https://ideas.repec.org/b/ohe/monogr/000140.html

22. Markiewicz K, van Til JA, IJzerman MJ. Medical devices early assessment methods: systematic literature review. Int J Technol Assess Health Care. 2014 Apr;30(02):137–46.

23. Vallejo-Torres L, Steuten LMG, Buxton MJ, Girling AJ, Lilford RJ, Young T. Integrating health economics modeling in the product

development cycle of medical devices: A Bayesian approach. Int J Technol Assess Health Care. 2008 Oct;24(04):459–64.

24. Girling A, Young T, Brown C, Lilford R. Early-Stage Valuation of Medical Devices: The Role of Developmental Uncertainty. Value Health. 2010 Jul;13(5):585–91.

25. McAteer H, Cosh E, Freeman G, Pandit A, Wood P, Lilford R. Cost-effectiveness analysis at the development phase of a potential health technology: examples based on tissue engineering of bladder and urethra. J Tissue Eng Regen Med. 2007 Sep;1(5):343–9.

26. Cosh E, Girling A, Lilford R, McAteer H, Young T. Investing in new medical technologies: A decision framework. J Commer Biotechnol. 2007 Aug;13(4):263–71.

27. IJzerman MJ, Steuten LMG. Early Assessment of Medical Technologies to Inform Product Development and Market Access: A Review of Methods and Applications. Appl Health Econ Health Policy. 2011 Sep;9(5):331–47.

28. Raiffa H. Decision analysis: Introductory Lectures on Choice Under Uncertainty. New York: McGraw Hill; 1997. 309 p. (College Costom Series).

29. Hunink MGM. Decision making in health and medicine: integrating health and values [Internet]. 2014 [cited 2021 Aug 13]. Available from: https://doi.org/10.1017/CBO9781139506779

30. Weinstein MC. Clinical decision analysis. Philadelphia: Saunders; 1980. 351 p.

31. Sox HC, Higgins MC, Owens DK. Medical decision making. 2nd ed. Chichester, West Sussex, UK: John Wiley & Sons; 2013.

32. Talic S, Shah S, Wild H, Gasevic D, Maharaj A, Ademi Z, et al. Effectiveness of public health measures in reducing the incidence of covid-19, SARS-CoV-2 transmission, and covid-19 mortality: systematic review and meta-analysis. BMJ. 2021 Nov 17;375:e068302.

33. O'Brien B. Economic evaluation of pharmaceuticals. Frankenstein's monster or vampire of trials? Med Care. 1996 Dec;34(12 Suppl):DS99-108.

34. Kassirer JP, Angell M. The Journal's Policy on Cost-Effectiveness Analyses. N Engl J Med. 1994 Sep 8;331(10):669–70.

35. Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T, et al. Modelling in economic evaluation: an unavoidable fact of life. Health Econ. 1997 Jun;6(3):217–27.

36. Guyatt G. Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine. JAMA. 1992 Nov 4;268(17):2420.

37. Briggs A, Sculpher M, Claxton K. Chapter 1. Introduction. In: Decision Modelling for Health Economic Evaluation. OUP Oxford; 2006. p. 2–3.

38. Cook TD, Campbell DT. Quasi-experimentation: design & analysis issues for field settings. Boston: Houghton Mifflin; 1979. 405 p.

39. Ciani O, Grigore B, Blommestein H, de Groot S, Möllenkamp M, Rabbe S, et al. Validity of surrogate endpoints and their impact on coverage recommendations. A retrospective analysis across international health technology assessment agencies. Med Decis Making. 2021 in press;

40. Grigore Bogdan, Ciani Oriana, Dams Florian, Federici Carlo Baldassarre, de Groot Saskia, Möllenkamp Meilin, et al. Surrogate endpoints in health technology assessment: an international review of

methodological guidelines. PHARMACOECONOMICS. 2020;38:1055–70.

41. Welton NJ, editor. Evidence synthesis for decision making in healthcare. Chichester, West Sussex: John Wiley & Sons; 2012. 282 p.

42. Fenwick E, Steuten L, Knies S, Ghabri S, Basu A, Murray JF, et al. Value of Information Analysis for Research Decisions—An Introduction: Report 1 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. Value Health. 2020 Feb;23(2):139–50.

43. Chalkidou K, Lord J, Fischer A, Littlejohns P. Evidence-Based Decision Making: When Should We Wait For More Information? Health Aff (Millwood). 2008 Nov;27(6):1642–53.

44. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD. Model Parameter Estimation and Uncertainty Analysis: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group–6. Med Decis Making. 2012 Sep;32(5):722–32.

45. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ. 1999 Jun;18(3):341–64.

46. Walker S, Sculpher M, Claxton K, Palmer S. Coverage with Evidence Development, Only in Research, Risk Sharing, or Patient Access Scheme? A Framework for Coverage Decisions. Value Health. 2012 May;15(3):570–9.

47. Dixit AK, Pindyck RS. Investment under uncertainty. Princeton, N.J: Princeton University Press; 1994. 468 p.

48. Trigeorgis L. Real options: managerial flexibility and strategy in resource allocation. Cambridge, Mass: MIT Press; 1996. 427 p.

49. Palmer S, Smith PC. Incorporating option values into the economic evaluation of health care technologies. J Health Econ. 2000 Sep;19(5):755–66.

50. Fenwick E, Steuten L, Knies S, Ghabri S, Basu A, Murray JF, et al. Value of Information Analysis for Research Decisions—An Introduction: Report 1 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. Value Health. 2020 Feb;23(2):139–50.

51. McKenna C, Claxton K. Addressing Adoption and Research Design Decisions Simultaneously: The Role of Value of Sample Information Analysis. Med Decis Making. 2011 Nov;31(6):853–65.

52. Bouttell J, Briggs A, Hawkins N. A different animal? Identifying the features of health technology assessment for developers of medical technologies. Int J Technol Assess Health Care. 2020 Aug;36(4):285–91.

53. Hartz S, John J. Contribution of economic evaluation to decision making in early phases of product development: A methodological and empirical review. Int J Technol Assess Health Care. 2008 Oct;24(04):465–72.

54. Ijzerman MJ, Koffijberg H, Fenwick E, Krahn M. Emerging Use of Early Health Technology Assessment in Medical Product Development: A Scoping Review of the Literature. PharmacoEconomics. 2017 Jul;35(7):727–40.

55. Craven MP, Allsop MJ, Morgan SP, Martin JL. Engaging with economic evaluation methods: insights from small and medium enterprises in the UK medical devices industry after training workshops. Health Res Policy Syst. 2012 Dec;10(1):29.

56. Kolominsky-Rabas PL, Djanatliev A, Wahlster P, Gantner-Bär M, Hofmann B, German R, et al. Technology foresight for medical device development through hybrid simulation: The ProHTA Project. Technol Forecast Soc Change. 2015 Aug;97:105–14.

57. Steuten LM. Multi-Dimensional Impact of the Public–Private Center for Translational Molecular Medicine (CTMM) in the Netherlands: Understanding New 21 st Century Institutional Designs to Support Innovation-in-Society. OMICS J Integr Biol. 2016 May;20(5):265–73.

58. Henriksson M, Lundgren F, Carlsson P. Informing the efficient use of health care and health care research resources - the case of screening for abdominal aortic aneurysm in Sweden. Health Econ. 2006 Dec;15(12):1311–22.

59. Grimm SE, Strong M, Brennan A, Wailoo AJ. The HTA Risk Analysis Chart: Visualising the Need for and Potential Value of Managed Entry Agreements in Health Technology Assessment. PharmacoEconomics. 2017 Dec;35(12):1287–96.

60. Rothery C, Claxton K, Palmer S, Epstein D, Tarricone R, Sculpher M. Characterising Uncertainty in the Assessment of Medical Devices and Determining Future Research Needs. Health Econ. 2017;26:109–23.

61. Heath A, Manolopoulou I, Baio G. A Review of Methods for Analysis of the Expected Value of Information. Med Decis Making. 2017 Oct;37(7):747–58.

62. Retèl velasca P, Grutters JPC, van Harten WH, Joore MA. Value of Research and Value of Development in Early Assessments of New Medical Technologies. Value Health. 2013;16(5):720–8.

63. Miquel-cases A, Retèl velasca P, van Harten WP, Steuten LSG. Decisions on Further Research for Predictive Biomarkers of High-

Dose Alkylating Chemotherapy in Triple-Negative Breast Cancer: A Value of Information Analysis. Value Health. 2016;19(4):419–30.

64. Willan AR. Optimal sample size determinations from an industry perspective based on the expected value of information. Clin Trials. 2008;5(6).

65. Breeze P, Brennan A. Valuing Trial Designs from a Pharmaceutical Perspective Using Value-Based Pricing. Health Econ. 2015;24(11):1468–82.

66. Polisena J, Castaldo R, Ciani O, Federici C, Borsci S, Ritrovato M, et al. Health technology assessment methods guidelines for medical devices: how can we address the gaps? The international federation of medical and biological engineering perspective. Int J Technol Assess Health Care. 2018;34(3):276–89.

67. Rothery C, Claxton K, Palmer S, Epstein D, Tarricone R, Sculpher M. Characterising Uncertainty in the Assessment of Medical Devices and Determining Future Research Needs: Characterising Uncertainty in the Assessment of Devices. Health Econ. 2017 Feb;26:109–23.

68. Sorenson C, Tarricone R, Siebert M, Drummond M. Applying health economics for policy decision making: do devices differ from drugs? Europace. 2011 May 1;13(suppl 2):ii54–8.

69. Sorenson C, Drummond M. Improving Medical Device Regulation: The United States and Europe in Perspective: Improving Medical Device Regulation. Milbank Q. 2014 Mar;92(1):114–50.

70. Sedrakyan A, Campbell B, Merino JG, Kuntz R, Hirst A, McCulloch P. IDEAL-D: a rational framework for evaluating and regulating the use of medical devices. BMJ. 2016 Jun 9;i2372.

71. Neugebauer EAM, Rath A, Antoine SL, Eikermann M, Seidel D, Koenen C, et al. Specific barriers to the conduct of randomised clinical trials on medical devices. Trials. 2017 Dec;18(1):427.

72. Tarricone R, Ciani O, Torbica A, Brouwer W, Chaloutsos G, Drummond MF, et al. Lifecycle evidence requirements for high-risk implantable medical devices: a European perspective. Expert Rev Med Devices. 2020 Oct 2;17(10):993–1006.

73. Drummond M, Griffin A, Tarricone R. Economic Evaluation for Devices and Drugs—Same or Different? Value Health. 2009 Jun;12(4):402–4.

74. Medical Device Regulation in: Official Journal of the European Union. Official Journal of the European Union, (EU) 2017/745 2017 p. 1–175.

75. Directive 2007/47/EC. In: Official Journal of the European Union. Directive 2001/83/EC 2001 p. 21–55.

76. Community Procedures for the Authorisation and Supervision of Medicinal Products for Human and Veterinary Use and establishing a European Medicines Agency: Regulation (EC) No 726/2004 In: Official Journal of the European Union. Official Journal of the European Union, 726/2004 2004 p. 1–33.

77. Common Framework for the Marketing of Products: Decision No 768/2008/EC In: Official Journal of the European Union. Official Journal of the European Union, 768/2008/EC 2008 p. 82–8.

78. De Maria C, Di Pietro L, Díaz Lantada A, Madete J, Makobore PN, Mridha M, et al. Safe innovation: On medical device legislation in Europe and Africa. Health Policy Technol. 2018 Jun;7(2):156–65.

79. Lee L, Kelly J, Nassif GJ, deBeche-Adams TC, Albert MR, Monson JRT. Defining the learning curve for transanal total mesorectal excision for rectal adenocarcinoma. Surg Endosc. 2020 Apr;34(4):1534–42.

80. van Workum F, Stenstra MHBC, Berkelmans GHK, Slaman AE, van Berge Henegouwen MI, Gisbertz SS, et al. Learning Curve and Associated Morbidity of Minimally Invasive Esophagectomy: A Retrospective Multicenter Study. Ann Surg. 2019 Jan;269(1):88–94.

81. Bernard A, Vaneau M, Fournel I, Galmiche H, Nony P, Dubernard JM. Methodological choices for the clinical development of medical devices. Med Devices Evid Res. 2014 Sep;325.

82. Garrison LP, Towse A, Briggs A, De Pouvourville G, Grueger J, Mohr PE, et al. Performance-based risk-sharing arrangements - Good practices for design, implementation, and evaluation: Report of the ISPOR good practices for performance-based risk-sharing arrangements task force. Value Health. 2013;16(5):703–19.

83. Towse A, Garrison LP. Can't Get No Satisfaction? Will Pay for Performance Help? PharmacoEconomics. 2010;28(2):93–102.

84. Brügger U, Ruckstuhl A, Horisberger B, Gratwohl A. Development of Coverage With Evidence Development for Medical Technologies in Switzerland From 1996 To 2012. Int J Technol Assess Health Care. 2014;30(03):253–9.

85. O'Malley SP, Selby WS, Jordan E. A successful practical application of Coverage with Evidence Development in Australia: Medical Services Advisory Committee interim funding and the PillCam® Capsule Endoscopy Register. Int J Technol Assess Health Care. 2009;25(03):290–6.

86. Adamski J, Godman B, Ofierska-Sujkowska G, Osiska B, Herholz H, Wendykowska K, et al. Risk sharing arrangements for pharmaceuticals: Potential considerations and recommendations for European payers. BMC Health Serv Res. 2010;

87. Brügger U. A review of Coverage with Evidence Development (CED) in different countries : What works and what doesn ' t. HTAi. 2014;2014.

88. Drummond M. When do performance-based risk-sharing arrangements make sense? European Journal of Health Economics. 2015;

89. Carbonneil C, Quentin F, Lee-Robin SH. A common policy framework for evidence generation on promising health technologies. Int J Technol Assess Health Care. 2009;25(S2):56–67.

90. Stafinski T, McCabe CJ, Menon D. Funding the Unfundable. PharmacoEconomics. 2010;28(2):113–42.

91. van de Wetering EJ, van Exel J, Brouwer WBF. The Challenge of Conditional Reimbursement: Stopping Reimbursement Can Be More Difficult Than Not Starting in the First Place! Value Health. 2017 Jan;20(1):118–25.

92. Expert Panel on Effective Ways of Investing in Health (EXPH). Innovative payment models for high-cost innovative medicines. 2018. 39 p.

93. Walker S, Sculpher M, Claxton K, Palmer S. Coverage with evidence development, only in research, risk sharing, or patient access scheme? a framework for coverage decisions. Value Health. 2012;15(3):570–9.

94. Dhalla IA, Garner S, Chalkidou K, Littlejohns P. Perspectives on the National Institute for Health and Clinical Excellence's

recommendations to use health technologies only in research. Int J Technol Assess Health Care. 2009;25(03):272–80.

95. Martelli N, van den Brink H. Special funding schemes for innovative medical devices in French hospitals: The pros and cons of two different approaches. Health Policy. 2014;117(1):1–5.

96. Claxton K, Palmer S, Longworth L, Bojke L, Griffin S, McKenna C, et al. Informing a decision framework for when NICE should recommend the use of health technologies only in the context of an appropriately designed programme of evidence development. Health Technol Assess. 2012;16(46).

97. Mohr PE, Tunis SR. Access with evidence development: The US experience. In: PharmacoEconomics. 2010.

98. Trueman P, Grainger DL, Downs KE. Coverage with Evidence Development: Applications and issues. Int J Technol Assess Health Care. 2010;26(01):79–85.

99. Mortimer D, Li JJ, Watts J, Harris A. Breaking up is hard to do: the economic impact of provisional funding contingent upon evidence development. Health Econ Policy Law. 2011;6(4):509–27.

100. Garrison LP, Carlson JJ, Bajaj PS, Towse A, Neumann PJ, Sullivan SD, et al. Private Sector Risk-Sharing Agreements in the United States. Am J Manag Care. 2015;21(9):632–40.

101. Kanavos P, Ferrario A, Tafuri G, Siviero P. Managing Risk and Uncertainty in Health Technology Introduction: The Role of Managed Entry Agreements. Glob Policy. 2017;8:84–92.

102. Levin L, Goeree R, Levine M, Krahn M, Easty T, Brown A, et al. Coverage with evidence development: The Ontario experience. Int J Technol Assess Health Care. 2011;27(2):159–68.

103. Tunis SR, Chalkidou K. Coverage with evidence development: A very good beginning, but much to be done. Commentary to Hutton et al. Int J Technol Assess Health Care. 2007;23(4):432–5.

104. Tunis SR, Pearson SD. Coverage options for promising technologies: Medicare's 'Coverage with evidence development'. Health Aff (Millwood). 2006;25(5):1218–30.

105. Menon D, McCabe CJ, Stafinski T, Edlin R. Principles of Design of Access with Evidence Development Approaches. PharmacoEconomics. 2010;28(2):109–11.

106. Carlson JJ, Gries KS, Yeung K, Sullivan SD, Garrison LP. Current status and trends in performance-based risk-sharing arrangements between healthcare payers and medical product manufacturers. Appl Health Econ Health Policy. 2014;12(3):231–8.

107. Goeree R, Chandra K, Tarride JE, O'Reilly D, Xie F, Bowen J, et al. Conditionally Funded Field Evaluations: PATHs Coverage with Evidence Development Program for Ontario. Value Health. 2010;13:S8–11.

108. Garattini L, Curto A, van de Vooren K. Italian risk-sharing agreements on drugs: are they worthwhile? Eur J Health Econ. 2014;16(1):1–3.

109. Bishop D, Lexchin J. Politics and its intersection with coverage with evidence development: a qualitative analysis from expert interviews. BMC Health Serv Res. 2013;13:88.

110. Foster NL, Mottola K, Hoffman JM. Coverage with evidence development: What to consider. JAMA Neurol. 2014;71(4):399–400.

111. Lindsay MJ, Siegel BA, Tunis SR, Hillner BE, Shields AF, Carey BP, et al. The National Oncologic PET Registry: Expanded Medicare

coverage for PET under coverage with evidence development. Am J Roentgenol. 2007;188(4):1109–13.

112. Relyea-Chew A. Ethical considerations in CMS's coverage with evidence development. Journal of the American College of Radiology. 2011.

113. Paulden M. Calculating and Interpreting ICERs and Net Benefit. PharmacoEconomics. 2020 Aug;38(8):785–807.

114. Karlsson G, Johannesson M. The decision rules of cost-effectiveness analysis. PharmacoEconomics. 1996 Feb;9(2):113–20.

115. Thokala P, Ochalek J, Leech AA, Tong T. Cost-Effectiveness Thresholds: the Past, the Present and the Future. PharmacoEconomics. 2018;36(5):509–22.

116. Siverskog J, Henriksson M. On the role of cost-effectiveness thresholds in healthcare priority setting. Int J Technol Assess Health Care. 2021 Jan 25;37(1):e23.

117. Nixon RM, Wonderling D, Grieve RD. Non-parametric methods for cost-effectiveness analysis: the central limit theorem and the bootstrap compared. Health Econ. 2010 Mar;19(3):316–33.

118. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. Med Decis Mak Int J Soc Med Decis Mak. 1993 Dec;13(4):322–38.

119. Briggs A, Sculpher M, Claxton K. Decision Modelling for Health Economic Evaluation. OUP Oxford; 2006. 250 p.

120. Claxton K. Exploring Uncertainty in Cost-Effectiveness Analysis: PharmacoEconomics. 2008;26(9):781–98.

121. Ades AE, Claxton K, Sculpher M. Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. Health Econ. 2006 Apr;15(4):373–81.

122. McCabe C, Paulden M, Awotwe I, Sutton A, Hall P. One-Way Sensitivity Analysis for Probabilistic Cost-Effectiveness Analysis: Conditional Expected Incremental Net Benefit. PharmacoEconomics. 2020 Feb;38(2):135–41.

123. Vreman RA, Geenen JW, Knies S, Mantel-Teeuwisse AK, Leufkens HGM, Goettsch WG. The Application and Implications of Novel Deterministic Sensitivity Analysis Methods. PharmacoEconomics. 2021 Jan;39(1):1–17.

124. Briggs AH. Handling Uncertainty in Cost-Effectiveness Models: PharmacoEconomics. 2000 May;17(5):479–500.

125. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. JAMA. 2016 Sep 13;316(10):1093.

126. Bowman KO, Shenton LR. Estimation, Method of Moments. In: Kotz S, Read CB, Balakrishnan N, Vidakovic B, Johnson NL, editors. Encyclopedia of Statistical Sciences [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2004 [cited 2021 Sep 24]. p. ess1618. Available from: https://onlinelibrary.wiley.com/doi/10.1002/0471667196.ess1618

127. Briggs MA, Sculpher M. An Introduction to Markov Modelling for Economic Evaluation. PharmacoEconomics. 2012 Oct 18;13(4):397–409.

128. Grigore B, Peters J, Hyde C, Stein K. Methods to Elicit Probability Distributions from Experts: A Systematic Review of Reported Practice in Health Technology Assessment. PharmacoEconomics. 2013 Nov;31(11):991–1003.

129. Soares MO, Sharples L, Morton A, Claxton K, Bojke L. Experiences of Structured Elicitation for Model-Based Cost-Effectiveness Analyses. Value Health. 2018 Jun;21(6):715–23.

130. Grutters JPC, van Asselt MBA, Chalkidou K, Joore MA. Healthy Decisions: Towards Uncertainty Tolerance in Healthcare Policy. PharmacoEconomics. 2015 Jan;33(1):1–4.

131. Fasterholdt I, Krahn M, Kidholm K, Yderstræde KB, Pedersen KM. Review of early assessment models of innovative medical technologies. Health Policy. 2017 Aug;121(8):870–9.

132. Claxton K, Briggs A, Buxton MJ, Culyer AJ, McCabe C, Walker S, et al. Value based pricing for NHS drugs: an opportunity not to be missed? BMJ. 2008 Feb 2;336(7638):251–4.

133. Iskandar R. Probability bound analysis: A novel approach for quantifying parameter uncertainty in decision-analytic modeling and cost-effectiveness analysis. Stat Med. 2021 Dec 20;40(29):6501–22.

134. Ferson S, Kreinovick V, Ginzburg L, Sentz F. Constructing Probability Boxes and Dempster-Shafer Structures [Internet]. 2003 Jan [cited 2021 Nov 21] p. SAND2002-4015, 809606. Report No.: SAND2002-4015, 809606. Available from: https://www.osti.gov/servlets/purl/809606/

135. Williamson RC, Downs T. Probabilistic arithmetic. I. Numerical methods for calculating convolutions and dependency bounds. Int J Approx Reason. 1990 Mar;4(2):89–158.

136. Hurwicz L. A criterion for decision making under uncertainty. Tech Rep 355 - Cowless Comm. 1952;

137. Breeze P, Brennan A. Valuing Trial Designs from a Pharmaceutical Perspective Using Value-Based Pricing: VALUING TRIAL DESIGNS USING VALUE-BASED PRICING. Health Econ. 2015 Nov;24(11):1468–82.

138. Weaver JC. Electroporation: a general phenomenon for manipulating cells and tissues. J Cell Biochem. 1993 Apr;51(4):426–35.

139. Kotnik T, Rems, Lea, Munir, Tarek, Miklavčič D. Membrane electroporation and electropermeabilization: mechanisms and models. Annu Rev Biophys. 2019;48.

140. Kotnik T, Kramar P, Pucihar G, Miklavcic D, Tarek M. Cell membrane electroporation- Part 1: The phenomenon. IEEE Electr Insul Mag. 2012 Sep;28(5):14–23.

141. Campana LG, Edhemovic I, Soden D, Perrone AM, Scarpa M, Campanacci L, et al. Electrochemotherapy – Emerging applications technical advances, new indications, combined approaches, and multi-institutional collaboration. Eur J Surg Oncol. 2019 Feb 1;45(2):92–102.

142. Reddy VY, Neuzil P, Koruth JS, Petru J, Funosako M, Cochet H, et al. Pulsed Field Ablation for Pulmonary Vein Isolation in Atrial Fibrillation. J Am Coll Cardiol. 2019 Jul 15;74(3):315–26.

143. Bart Geboers, Hester J. Scheffer, Philip M. Graybill, Alette H. Ruarus, Sanne Nieuwenhuizen, Robbert S. Puijk, et al. High-voltage electrical pulses in oncology: irreversible electroporation, electrochemotherapy, gene electrotransfer, electrofusion, and electro-immunotherapy. RADIOLOGY. 295:254–72.

144. Mahnič-Kalamiza S, Vorobiev E, Miklavčič D. Electroporation in Food Processing and Biorefinery. J Membr Biol. 2014 Dec 1;247(12):1279–304.

145. Kotnik T, Frey W, Sack M, Haberl Meglič S, Peterka M, Miklavčič D. Electroporation-based applications in biotechnology. Trends Biotechnol. 2015 Aug;33(8):480–8.

146. Golberg A, Sack M, Teissie J, Pataro G, Pliquett U, Saulis G, et al. Energy-efficient biomass processing with pulsed electric fields for bioeconomy and sustainable development. Biotechnol Biofuels. 2016 Apr 27;9(1):94.

147. Davalos RV, Mir LM, Rubinsky B. Tissue Ablation with Irreversible Electroporation. Ann Biomed Eng. 2005 Feb 1;33(2):223.

148. Lambricht L, Lopes A, Kos S, Sersa G, Préat V, Vandermeulen G. Clinical potential of electroporation for gene therapy and DNA vaccine delivery. Expert Opin Drug Deliv. 2016 Feb 1;13(2):295–310.

149. Rebersek M, Miklavcic D, Bertacchini C, Sack M. Cell membrane electroporation-Part 3: the equipment. IEEE Electr Insul Mag. 2014 May;30(3):8–18.

150. Pirc E, Rebersek M, Miklavcic D. 12 Dosimetry in Electroporation-Based Technologies and Treatments. In: Markov M, editor. Dosimetry in Bioelectromagnetics. CRC Press, Taylor & Francis Group; 2017. p. 233–68.

151. Campana LG, Miklavčič D, Bertino G, Marconato R, Valpione S, Imarisio I, et al. Electrochemotherapy of superficial tumors – Current status:: Basic principles, operating procedures, shared indications, and emerging applications. Semin Oncol. 2019 Apr 1;46(2):173–91.

152. Colombo GL, Matteo SD, Mir LM. Cost-effectiveness analysis of electrochemotherapy with the Cliniporator™ vs other methods for the control and treatment of cutaneous and subcutaneous tumors. Ther Clin Risk Manag. 2008 Apr;4(2):541–8.

153. Pirc E, Pecchia L, Reberšek M, Serša G, Snoj M, Grošelj A, et al. Study design of a medical device pre-market assessment: a case study on electrochemotherapy. Zdr Vestn Slov Med J. 2018 Mar 2;87(1–2):22–40.

154. Electrochemotherapy for primary basal cell carcinoma and primary squamous cell carcinoma | Guidance | NICE [Internet]. [cited 2019 Nov 7]. Available from: https://www.nice.org.uk/guidance/IPG478

155. Clover AJP, Salwa SP, Bourke MG, McKiernan J, Forde PF, O'Sullivan ST, et al. Electrochemotherapy for the treatment of primary basal cell carcinoma; A randomised control trial comparing electrochemotherapy and surgery with five year follow up. Eur J Surg Oncol J Eur Soc Surg Oncol Br Assoc Surg Oncol. 2020 May;46(5):847–54.

156. Gehl J, Sersa G, Matthiessen LW, Muir T, Soden D, Occhini A, et al. Updated standard operating procedures for electrochemotherapy of cutaneous tumours and skin metastases. Acta Oncol Stockh Swed. 2018 Jul;57(7):874–82.

157. Hočevar M, Strojan P, Ocvirk J, Reberšek M, Boc M, Perić B, et al. Priporočila za obravnavo bolnikov s kožnim melanomom [Internet]. Available from: https://www.onko-i.si/fileadmin/onko/datoteke/Smernice/Priporocila_za_obravnavo_bolnikov_s_koznim_melanomom_2018.pdf

158. Mir LM, Gehl J, Serša G, Collins C, Garbay JR, Billard V, et al. Standard operating procedures of the electrochemotherapy: Instructions for the use of bleomycin or cisplatin administered either systemically or locally

and electric pulses delivered by the CliniporatorTM by means of invasive or non-invasive electrodes. EJC Suppl. 2006;4(4):14–25.

159. Thokala P, Ochalek J, Leech AA, Tong T. Cost-Effectiveness Thresholds: the Past, the Present and the Future. PharmacoEconomics. 2018;36(5):509–22.

160. Choosing interventions that are cost-effective [Internet]. Geneva: World Health Organization; 2014. [cited 2020 Feb 5]. Available from: http://www.who.int/choice/en/

161. Woods B, Revill P, Sculpher M, Claxton K. Country-Level Cost-Effectiveness Thresholds: Initial Estimates and the Need for Further Research. Value Health. 2016 Dec;19(8):929–35.

162. Welton NJ, Ades AE. Estimation of Markov Chain Transition Probabilities and Rates from Fully and Partially Observed Data: Uncertainty Propagation, Evidence Synthesis, and Model Calibration: Med Decis Making [Internet]. 2016 Sep 5 [cited 2019 Dec 23]; Available from: https://journals.sagepub.com/doi/10.1177/0272989X05282637

163. Heidelberger P, Welch PD. A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations. Commun ACM. 1981 Apr;24(4):233–45.

164. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. Stat Sci. 1992 Nov;7(4):457–72.

165. Jones E, Epstein D, García-Mochón L. A Procedure for Deriving Formulas to Convert Transition Rates to Probabilities for Multistate Markov Models. Med Decis Mak Int J Soc Med Decis Mak. 2017;37(7):779–89.

166. Benčina G, Buljan M, Šitum M, Stevanović R, Benković V. Health and Economic Burden of Skin Melanoma in Croatia - Cost-of-illness Study. Acta Dermatovenerol Croat ADC. 2017;25(1):1–7.

167. Krensel M, Schäfer I, Augustin M. Cost-of-illness of melanoma in Europe – a modelling approach. J Eur Acad Dermatol Venereol. 2019;33(S2):34–45.

168. Maio M, Ascierto P, Testori A, Ridolfi R, Bajetta E, Queirolo P, et al. The cost of unresectable stage III or stage IV melanoma in Italy. J Exp Clin Cancer Res. 2012 Nov 1;31(1):91.

169. Perera E, Gnaneswaran N, Jennens R, Sinclair R. Malignant Melanoma. Healthcare. 2013 Dec 20;2(1):1–19.

170. Cabula C. Neoadjuvant electochemotherapy for breast cancer. The first case to be treated in Italy: Results after 25 months of follow-up. Eur J Surg Oncol. 2012 Oct 1;38(10):1004.

171. Solari N, Spagnolo F, Queirolo P, Ponte E, Cafiero F. Electrochemotherapy for cutaneous metastases: The IRCCS San Martino Hospital and National Cancer Research Institute of Genoa's experience. Eur J Surg Oncol. 2012 Oct 1;38(10):1004.

172. Caracò C, Mozzillo N, Marone U, Simeone E, Benedetto L, Di Monta G, et al. Long-lasting response to electrochemotherapy in melanoma patients with cutaneous metastasis. BMC Cancer. 2013 Dec 1;13(1):564.

173. Valpione S, Campana LG, Pigozzo J, Chiarion-Sileni V. Consolidation electrochemotherapy with bleomycin in metastatic melanoma during treatment with dabrafenib. Radiol Oncol. 2015 Mar;49(1):71–4.

174. Bianchi G, Campanacci L, Ronchetti M, Donati D. Electrochemotherapy in the Treatment of Bone Metastases: A Phase II Trial. World J Surg. 2016 Dec 1;40(12):3088–94.

175. Plaschke CC, Johannesen HH, Hansen RH, Hendel HW, Kiss K, Gehl J, et al. The DAHANCA 32 study: Electrochemotherapy for recurrent mucosal head and neck cancer. Head Neck. 2019;41(2):329–39.

176. Kaehler KC, Egberts F, Hauschild A. Electrochemotherapy in Symptomatic Melanoma Skin Metastases: Intraindividual Comparison with Conventional Surgery. Dermatol Surg. 2010 Jul;36(7):1200.

177. Campana LG, Mocellin S, Basso M, Puccetti O, De Salvo GL, Chiarion-Sileni V, et al. Bleomycin-Based Electrochemotherapy: Clinical Outcome from a Single Institution's Experience with 52 Patients. Ann Surg Oncol. 2009 Jan 1;16(1):191–9.

178. Campana LG, Testori A, Curatolo P, Quaglino P, Mocellin S, Framarini M, et al. Treatment efficacy with electrochemotherapy: A multi-institutional prospective observational study on 376 patients with superficial tumors. Eur J Surg Oncol J Eur Soc Surg Oncol Br Assoc Surg Oncol. 2016 Dec;42(12):1914–23.

179. Arabía FA. The Total Artificial Heart: Where Are We? Cardiol Rev. 2020 Nov;28(6):275–82.

180. Copeland JG, Smith RG, Arabia FA, Nolan PE, Sethi GK, Tsau PH, et al. Cardiac Replacement with a Total Artificial Heart as a Bridge to Transplantation. N Engl J Med. 2004 Aug 26;351(9):859–67.

181. Melton N, Soleimani B, Dowling R. Current Role of the Total Artificial Heart in the Management of Advanced Heart Failure. Curr Cardiol Rep. 2019 Nov;21(11):142.

182. Kormos RL, Cowger J, Pagani FD, Teuteberg JJ, Goldstein DJ, Jacobs JP, et al. The Society of Thoracic Surgeons Intermacs Database Annual Report: Evolving Indications, Outcomes, and Scientific Partnerships. Ann Thorac Surg. 2019 Feb;107(2):341–53.

183. DeVries WC, Anderson JL, Joyce LD, Anderson FL, Hammond EH, Jarvik RK, et al. Clinical Use of the Total Artificial Heart. N Engl J Med. 1984 Feb 2;310(5):273–8.

184. Arabía FA, Cantor RS, Koehl DA, Kasirajan V, Gregoric I, Moriguchi JD, et al. Interagency registry for mechanically assisted circulatory support report on the total artificial heart. J Heart Lung Transplant. 2018 Nov;37(11):1304–12.

185. Vierecke J, Gahl B, de By TMMH, Antretter H, Beyersdorf F, Caliskan K, et al. Results of primary biventricular support: an analysis of data from the EUROMACS registry. Eur J Cardiothorac Surg. 2019 Dec 1;56(6):1037–45.

186. de By TMMH, Mohacsi P, Gummert J, Bushnaq H, Krabatsch T, Gustafsson F, et al. The European Registry for Patients with Mechanical Circulatory Support (EUROMACS): first annual report. Eur J Cardio-Thorac Surg Off J Eur Assoc Cardio-Thorac Surg. 2015 May;47(5):770–6; discussion 776-777.

187. Sharples L, Buxton M, Caine N, Cafferty F, Demiris N, Dyer M, et al. Evaluation of the ventricular assist device programme in the UK. Health Technol Assess [Internet]. 2006 Nov [cited 2021 Jul 7];10(48). Available from: https://www.journalslibrary.nihr.ac.uk/hta/hta10480/

188. Guyot P, Ades A, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published

Kaplan-Meier survival curves. BMC Med Res Methodol. 2012 Dec;12(1):9.

189. Lenderman JC. Intermacs Appendices - School of Medicine - Interagency Registry for Mechanically Assisted Circulatory Support | UAB [Internet]. 2021. Available from: https://www.uab.edu/medicine/intermacs/intermacs-documents

190. Canada JM, Evans RK, Abbate A, Arena R, Tang DG, Kasirajan V, et al. Exercise Capacity in Patients with the Total Artificial Heart. ASAIO J. 2019 Jan;65(1):36–42.

191. Streur MM, Beckman JA, Dougherty CM, Li S, Mahr C. Quality of life and rehabilitation after total artificial heart. Ann Cardiothorac Surg. 2020 Mar;9(2):128–30.

192. David CH, Lacoste P, Nanjaiah P, Bizouarn P, Lepoivre T, Michel M, et al. A heart transplant after total artificial heart support: initial and long-term results. Eur J Cardiothorac Surg. 2020 Dec 1;58(6):1175–81.

193. Levin AP, Jaramillo N, Garan AR, Takeda K, Takayama H, Yuzefpolskaya M, et al. Outcomes of contemporary mechanical circulatory support device configurations in patients with severe biventricular failure. J Thorac Cardiovasc Surg. 2016 Feb;151(2):530-535.e2.

194. Kirsch MEW, Nguyen A, Mastroianni C, Pozzi M, Léger P, Nicolescu M, et al. SynCardia Temporary Total Artificial Heart as Bridge to Transplantation: Current Results at La Pitié Hospital. Ann Thorac Surg. 2013 May;95(5):1640–6.

195. Shah KB, Thanavaro KL, Tang DG, Quader MA, Mankad AK, Tchoukina I, et al. Impact of INTERMACS Profile on Clinical

Outcomes for Patients Supported With the Total Artificial Heart. J Card Fail. 2016 Nov;22(11):913–20.

196. Bank EC. ECB euro reference exchange rate: Pound sterling (GBP) [Internet]. European Central Bank. 2021 [cited 2021 Jul 7]. Available from: https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates/html/eurofxref-graph-gbp.en.html

197. Luengo-Fernandez R, Violato M, Candio P, Leal J. Economic burden of stroke across Europe: A population-based cost analysis. Eur Stroke J. 2020 Mar;5(1):17–25.

198. Tarricone R, Amatucci F, Armeni P, Banks H, Borsoi L, Callea G, et al. Establishing a national HTA program for medical devices in Italy: Overhauling a fragmented system to ensure value and equal access to new medical technologies. Health Policy. 2021 May;125(5):602–8.

199. Koffijberg H, Knies S, Janssen MP. The Impact of Decision Makers' Constraints on the Outcome of Value of Information Analysis. Value Health. 2018 Feb;21(2):203–9.

200. Gablonsky JM, Kelley CT. A Locally-Biased form of the DIRECT Algorithm. J Glob Optim. 2001;21(1):27–37.

201. Ypma J. Introduction to nloptr: an R interface to NLopt. 2021;

202. Grutters JPC, Govers T, Nijboer J, Tummers M, van der Wilt GJ, Rovers MM. Problems and Promises of Health Technologies: The Role of Early Health Economic Modeling. Int J Health Policy Manag. 2019 Jun 22;8(10):575–82.

203. Lofgren K. Informing Financial Investments in Health: Applications of Decision-Analytic Methods. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences. [Internet]. Harvard University,

Graduate School of Arts & Sciences; 2020. Available from: https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37365875

204. Pontes C, Fontanet JM, Vives R, Sancho A, Gómez-Valent M, Ríos J, et al. Evidence supporting regulatory-decision making on orphan medicinal products authorisation in Europe: methodological uncertainties. Orphanet J Rare Dis. 2018 Dec;13(1):206.

205. Deng W, Yang X, Zou L, Wang M, Liu Y, Li Y. An improved self-adaptive differential evolution algorithm and its application. Chemom Intell Lab Syst. 2013 Oct;128:66–76.

206. Ellis AG, Iskandar R, Schmid CH, Wong JB, Trikalinos TA. Active learning for efficiently training emulators of computationally expensive mathematical models. Stat Med. 2020 Nov 10;39(25):3521–48.

207. Garrison LP, Towse A, Briggs A, de Pouvourville G, Grueger J, Mohr PE, et al. Performance-Based Risk-Sharing Arrangements—Good Practices for Design, Implementation, and Evaluation: Report of the ISPOR Good Practices for Performance-Based Risk-Sharing Arrangements Task Force. Value Health. 2013 Jul;16(5):703–19.

208. Rothery C, Claxton K, Palmer S, Epstein D, Tarricone R, Sculpher M. Characterising Uncertainty in the Assessment of Medical Devices and Determining Future Research Needs: Characterising Uncertainty in the Assessment of Devices. Health Econ. 2017 Feb;26:109–23.

209. Grutters JPC, Abrams KR, Ruysscher D, Pijls-Johannesma M, Peters HJM, Beutner E, et al. When to Wait for More Evidence? Real Options Analysis in Proton Therapy. The Oncologist. 2011 Dec;16(12):1752–61.

210. Gerkens S, Neyt M, San Miguel L, Vinck I, Thiry N, Cleemput I. How to improve the belgian process for managed entry agreements? An

analysis of the belgian and international experience [Internet]. Belgium KCE; 2017. Available from: https://kce.fgov.be/sites/default/files/atoms/files/KCE_288_Improve_Belgian_process_managed_entry_agreements_Report.pdf

211. Kanavos P, Ferrario A, Tafuri G, Siviero P. Managing Risk and Uncertainty in Health Technology Introduction: The Role of Managed Entry Agreements. Glob Policy. 2017 Mar;8:84–92.

212. Stafinski T, McCabe CJ, Menon D. Funding the Unfundable: Mechanisms for Managing Uncertainty in Decisions on the Introduction of New and Innovative Technologies into Healthcare Systems. PharmacoEconomics. 2010 Feb;28(2):113–42.

213. van de Wetering EJ, van Exel J, Brouwer WBF. The Challenge of Conditional Reimbursement: Stopping Reimbursement Can Be More Difficult Than Not Starting in the First Place! Value Health. 2017 Jan;20(1):118–25.

214. Schaffer SK, Messner D, Mestre-Ferrandiz J, Tambor E, Towse A. Paying for Cures: Perspectives on Solutions to the "Affordability Issue". Value Health. 2018 Mar;21(3):276–9.

215. Sculpher M, Drummond M, Buxton M. The iterative use of economic evaluation as part of the process of health technology assessment. J Health Serv Res Policy. 1997 Jan;2(1):26–30.

216. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on health technology assessment and amending Directive 2011/24/EU COM/2018/051 final - 2018/018 (COD) [Internet]. [cited 2021 Dec 1]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52018PC0051

217. Fleurence RL, Selby JV. A Step Closer to Better Research Prioritization? The ISPOR Value of Information Task Force Reports. Value Health. 2020 Mar;23(3):287–8.

218. Carino T, Williams RD, Colbert AM, Bridger P. Medicare's coverage of colorectal cancer drugs: A case study in evidence development and policy. Health Aff (Millwood). 2006;25(5):1231–9.

219. De Pouvourville G. Risk-sharing agreements for innovative drugs: A new solution to old problems? Eur J Health Econ. 2006;7(3):155–7.

220. Tunis SR, Pearson SD. Coverage options for promising technologies: Medicare's 'Coverage with evidence development'. Health Aff (Millwood). 2006;25(5):1218–30.

221. Boggild M, Palace J, Barton P, Ben-Shlomo Y, Bregenzer T, Dobson C, et al. Multiple sclerosis risk sharing scheme: Two year results of clinical cohort study with historical comparator. BMJ Online. 2009;339(7734):1359–63.

222. Pickin M, Cooper CL, Chater T, O'Hagan A, Abrams KR, Cooper NJ, et al. The multiple sclerosis risk sharing scheme monitoring study - Early results and lessons for the future. BMC Neurol. 2009;9:1–8.

223. J. McCabe C, Stafinski T, Edlin R, Menon D. Access with Evidence Development Schemes. PharmacoEconomics. 2010;28(2):143–52.

224. Stafinski T, McCabe CJ, Menon D. Funding the Unfundable. PharmacoEconomics. 2010;28(2):113–42.

225. Espín J, Rovira J, Gracia L. Experiences and impact of European risk-sharing schemes focusing on oncology medicines. Res Gate - EMINET. 2011;(January):1–39.

226. Garattini L, Casadei G. Risk sharing agreements: What lessons from Italy? Int J Technol Assess Health Care. 2011;27(2):169–72.

227. Lexchin J. Coverage with Evidence Development for Pharmaceuticals: A Policy in Evolution? Int J Health Serv. 2011;41(2):337–54.

228. Neumann PJ, Chambers JD, Simon F, Meckley LM. Risk-sharing arrangements that link payment for drugs to health outcomes are proving hard to implement. Health Aff (Millwood). 2011;

229. Launois R, Navarrete LF, Ethgen O, Moine J gabriel Le. Scope and Framework of Future Market Entry Agreements. 2014;1:1–7.

230. J.R.R. L, I. K, W. L. Coverage with evidence development and managed entry in the funding of personalized medicine: Practical and ethical challenges for oncology. J Clin Oncol. 2015;

231. Mohseninejad L, Van Gils C, Uyl-De Groot CA, Buskens E, Feenstra T. Evaluation of patient registries supporting reimbursement decisions: The case of oxaliplatin for treatment of stage III colon cancer. Value Health. 2015;18(1):84–90.

232. Navarria A, Drago V, Gozzo L, Longo L, Mansueto S, Pignataro G, et al. Do the current performance-based schemes in Italy really work? 'Success fee': A novel measure for cost-containment of drug expenditure. Value Health. 2015;18(1):131–6.

233. Thompson M, Henshall C, Garrison LP, Griffin AD, Coyle D, Long S, et al. Targeting improved patient outcomes using innovative product listing agreements: A survey of canadian and international key opinion leaders. Clin Outcomes Res. 2016;8:427–33.

234. Clopes A, Gasol M, Cajal R, Segú L, Crespo R, Mora R, et al. Financial consequences of a payment-by-results scheme in Catalonia: gefitinib in

advanced EGFR-mutation positive non-small-cell lung cancer. J Med Econ. 2017;20(1):1–7.

235. Gerkens S, Neyt M, San Miguel L, Vinck I, Thiry N, Cleemput I. How to improve the Belgian process for managed entry agreements? An analysis of the Belgian and international experience. 2018;75.

236. Nazareth T, Ko JJ, Sasane R, Frois C, Carpenter S, Demean S, et al. Outcomes-Based Contracting Experience: Research Findings from U.S. and European Stakeholders. J Manag Care Spec Pharm. 2017;23(10):1018–26.

237. van de Wetering EJ, van Exel J, Brouwer WBF. The Challenge of Conditional Reimbursement: Stopping Reimbursement Can Be More Difficult Than Not Starting in the First Place! Value Health. 2017;20(1):118–25.

238. Bouvy JC, Sapede C, Garner S. Managed entry agreements for pharmaceuticals in the context of adaptive pathways in Europe. Front Pharmacol. 2018;9(MAR):1–8.

239. European Commission. Expert panel on effective ways of investing in health (EXPH). 2018. 39 p.

240. Letters HE. Health Economics Letters Incorporating Feelings Related To the Uncertainty About Future Health in. 2008;1213(December 2007):1207–13.

241. Makady A, van Veelen A, de Boer A, Hillege H, Klungel OH, Goettsch W. Implementing managed entry agreements in practice: The Dutch reality check. Health Policy. 2019;123(3):267–74.

242. Pouwels XGLV, Grutters JPC, Bindels J, Ramaekers BLT, Joore MA. Uncertainty and Coverage With Evidence Development: Does Practice Meet Theory? Value Health. 2019;1–9.

Supplementary material S2.1

Table S1 Overview of challenges with CED schemes

#	Author(s) [ref]	Year	Country	Type of health technology[a]	Challenges associated with CED schemes
1	Carino et al. (218)	2006	USA	Pharmaceuticals	*Implementation:* Uncertainty associated with the (effectiveness of) off-label use of oxaliplatin, irinotecan, cetuximab, and bevacizumab in other cancer indications than colorectal cancer (as off-label use is often not subjected to the same level of scrutiny as on-label use) and the potential high cost of these therapies. *Evaluation:* Questions are raised relating to the quality of the additional data collected and the suitability of the sample included in the study. *Other:* There is a lack of transparency regarding the choice of specific trials and of discussing this choice with stakeholders and the public. Questions are raised relating to the relationship of CED schemes with respect to existing policies, and to CMS general authority to link coverage with additional data collection (by law they have to pay for these drugs if drugs are captured in approved compendia or published in a recognized journal). There are concerns regarding access to treatment for patients who do not wish to enrol or are not eligible for clinical trials (with respect to OIR schemes).
2	De Pouvourville et al. (219)	2006	NS	Pharmaceuticals	*Desirability:* Manufacturers may risk producing data that are useful to competitors (fast-followers). Payers may face risk paying a price that is higher than initial value of the drug. *Research design:* Challenges of designing a control scheme acceptable for both parties. The design should depend on the nature of the expected outcomes, and on the expected level of results. Particularly the choice is on what is the appropriate level of proof (e.g. a RCTs or other observational studies). For certain outcomes such as costs or quality of life, generalisability issues (e.g. transferability of the study results to another country) may be an issue. *Implementation:* The monitoring of schemes implies the availability for the payer of analytical expertise and good information systems.

					Evaluation: Schemes may require collection of cost data. Cost data require careful sampling of the healthcare provider population to ensure that no uncontrolled source of variability is introduced that may confound the predictions of the pivotal study leading to the claims made by the company.
3	Tunis and Pearson (220)	2006	USA	All technologies	*Desirability:* An appropriate application of CED requires an evidentiary "middle ground". Technologies whose supporting evidence falls into this category, making them eligible for CED, can be selected reliably only when clear evidentiary boundaries, both above and below, are presented in a transparent manner. Given the expense and effort involved in conducting prospective studies through CED, it will be important to have a robust approach to ensure that technologies are selected based on the quality of existing evidence, potential for reducing the burden of suffering for Medicare beneficiaries, and potential to produce savings for the program. *Implementation:* CED schemes require substantial and sustainable funding, and to attract such funding from public or private resources (or both), it will probably be necessary to establish some organization with the scientific credibility, political independence, and technical expertise to manage these projects successfully and efficiently. An example provided on a CED scheme for FDG-PET reports that it took several years before a trial protocol for the scheme was funded and deemed by the CMS to meet the requirements of the coverage decision. The examples highlights the challenges of getting the necessary clinical research designed, funded, and implemented in a time frame consistent with the needs of clinicians, patients, and other decision makers. *Evaluation:* There are difficulties with encouraging product developers to conduct needed clinical research after obtaining CMS coverage. Since 1999, CMS chose in several cases to provide coverage for technologies supported by evidence that was marginally sufficient, while including specific statements in the decision memos that additional research would be desirable. In some cases, the decision memo indicated that the technology would be reviewed again in several years and that the agency would consider restricting coverage if no further evidence was produced. However, the suggested research was never done, and the CMS did not pursue any effort to narrow coverage. *Other*: Various stakeholders have different views about which technologies require further study, what sorts of questions require answers, and what methods are necessary to answer those questions. Policy makers, patient groups, and other stakeholders have questioned the ethics of the entire CED approach by suggesting that it is coercive to link insurance coverage to requirements for participation in patient registries and clinical trials. Questions are raised related to the appropriate role of protecting human subjects in the context of CED.
4	Hutton et al. (103)	2007	NS	All technologies	*Desirability*: The point in the life-cycle at which a technology should be assessed remains a contentious issue. From the perspective of a decision maker, assessment of technologies close to the time of their regulatory approval and/or launch allows for a timely decision to be made regarding their coverage

					and availability. Any assessment of a technology at this point in its life-cycle will inevitably involve a degree of uncertainty around the clinical and economic data. *Research design:* Agreement on data requirements and study design, time horizon of additional data collection (in view of the pace of technological change in healthcare, a CED of more than three years may be of limited relevance, especially for MDs), finance and management of studies, data collection and analysis (should be undertaken by an independent body), decision with further evidence (even when quality of additional evidence is suboptimal). *Implementation:* There is an additional burden associated with monitoring and review in the light of further evidence (and possible costs of data collection if not fully borne by manufacturer). *Evaluation:* There are difficulties associated with withdrawing technologies that prove not to be cost-effective. There are also risks involved in using technologies that are not fully evaluated or recommended by guidance. There is an additional burden associated with data collection/analysis. *Other:* CED schemes may increase exposure to litigation as there is a possible delay in market access for effective technologies.
5	Lindsay et al. (111)	2007	USA	MDs	*Implementation:* The initial implementation of the NOPR was marked by minor difficulties, including security concerns with unencrypted health information in email communications, programming errors, and some ambiguity of specific items on the pre- and post-PET forms. Either before or on arrival at a PET facility, each patient receives a standard NOPR information document describing the registry and requesting that the patient provide oral consent for the use of identified data for research purposes. If the patient withholds consent, the identified data are still collected by the PET facility, sent electronically to the NOPR, and submitted to CMS; however, the data are not used for research. In such cases, CMS nevertheless pays for the PET scan. After the PET examination is performed and the PET report is entered into the database, the referring physician must complete an indication-specific post-PET form to assess the physician's intended patient management in light of the PET findings. The post-PET form also solicits consent from the referring physician to allow the response data to be used for research purposes. Physicians who elect to withhold consent nevertheless must submit all of the required data elements as a condition of reimbursement by CMS; however, those data are excluded from the research data set used by NOPR investigators. *Evaluation:* As is the case with any large registry, the accuracy of the data submitted to the NOPR depends on the submissions of referring physicians. Busy clinicians or their designees (e.g., nurse practitioners) may not always take the few minutes necessary to complete the forms accurately. Further, the data collected by the NOPR on physicians' intended patient management may or may not reflect actual management.
6	Boggild et al. (221)	2009	UK	Pharmaceuticals	*Evaluation:* The two-year results of the scheme failed to show clear definitive evidence for decision making. The data on patients for the control population (Ontario dataset) have been collated retrospectively with the disability status scores smoothed to eliminate short term fluctuations. As a

					result there are no "regressions" in EDSS in the Ontario dataset; disability scores for individual patients can only worsen over time and the ScHARR model had to reflect this. As a consequence, the data from the risk sharing scheme had to be modified to allow comparison with the predicted disease progression in the model, confounding the results. In fact, using the raw data from the risk-sharing scheme would have overestimated the true benefits of treatment. Since data suggest that patients with multiple sclerosis now progress more slowly compared to the past, comparison with historical data should overestimate the potential benefits of treatment. In addition, the Ontario dataset differed in other covariates, such as age at presentation and sex that are associated with disease progression. Therefore, from the outset that a major limitation was the validity and generalisability of the comparison dataset. Another potential source of bias, which would overestimate the benefits of treatment, is incomplete follow up. Completeness of follow-up might well differ between the treated and reference cohorts. Since the onset of the risk sharing scheme, there have been changes in the management of patients with multiple sclerosis. The updated guidelines from the Association of British Neurologists have widened the eligibility criteria for disease modifying treatments, though funding has not been agreed with the Department of Health. This itself will not impact on the results of the scheme, although it might make it more difficult to generalise the findings to current practice. With the licensing of new classes of drugs such as natulizamab, patients whose relapses do not respond to the scheme drugs (either because they have an aggressive form of multiple sclerosis or because they have neutralising antibodies to interferon beta) might be switched to newer non-scheme drugs. To avoid bias, it is important that such patients are included in the analysis, even if only in sensitivity analyses.
7	Carbonneil et al. (89)	2009	Australia, Belgium, Canada (Ontario), UK (England/Wales), Germany, Spain, USA	All technologies	*Research design:* There are difficulties associated with agreeing on data and study requirements. There are limited funds to finance the generation of evidence that meets HTA agency and decision-maker requirements. *Evaluation:* The evidence generated may not meet the quality criteria and, therefore, cannot tinform a decision. *Other:* Lack of coordination among the partners and bodies overseeing data collection. No well-defined regulatory framework governing coordination and financing.
8	Dhalla et al. (94)	2009	UK	All technologies	*Desirability*: Although formal methods for informing eligibility exist (e.g. VOI analysis) it was noted that these methods may be difficult to apply in specific cases and it was suggested that NICE develop a formal process using experts to help decide whether research in a particular area would be practical and likely to reduce uncertainty. In addition, an OIR decision should be accompanied by a clear statement regarding what data are required to reduce uncertainty. Where the nature of the evidence required by NICE is clear, it would also be appropriate for NICE's Appraisal Committee to suggest

					the type of study that would be most appropriate. Also the importance of developing a transparent pathway to prioritise and implement OIR decisions is highlighted. *Research design:* Ensuring that research funding arrangements do not discourage the participation of healthcare providers in OIR-related research is of vital importance. The current situation in the UK, where costs associated with research studies are divided into different categories (research costs, which are paid for by the research funding agency, and support costs and treatment costs, which are generally paid for by the care provider) was widely viewed by interviewees as a barrier to research.
9	O'Malley et al. (85)	2009	Australia	MDs	*Research design:* No funding, guidelines, or infrastructure assistance for the data collection was made available from MSAC, the Australian Government Department of Health and Ageing, or GESA. Therefore the sponsor had to seek independent professional advice and assistance for the establishment of the PillCam® Data Register. The sponsor also provided the required software infrastructure and defined the procedure and forms for data collection. The questionnaire for data collection in the registry was designed by physicians and, as a consequence, may not have been ideal for economic analysis. Although the register provided data on the reduction in hospitalisation resulting from the diagnosis by capsule endoscopy and thus treatment, details that would have enhanced the economic analysis, such as length of stay, were not captured by the questionnaire. *Implementation:* There was a decline in participation in 2007, perhaps reflecting one of the main problems associated with a voluntary register continued for an extended number of years. In Australia, the majority of interim funding recommendations by MSAC are for three years. However, there does not appear to be any statistical rationale to support the use of this three-year time limit. Physicians were not paid for data collection and reporting. *Evaluation:* Data collection envisaged collection of follow up data. Only 16.6 percent of the patients had follow up. This highlights another potential problem and source of bias with the use of data collection for procedures that require follow-up. The MSAC Assessment Report was based on evidence from only one brand of capsule, and at the time of the commencement of the register, this was the only capsule available on the Australian market. However, within the first year of interim funding, a second brand of capsule entered the Australian market, followed by a third brand before the report on the PillCam® capsule Endoscopy Register was submitted to MSAC in 2007. Despite the availability of the three brands of capsules, only data from PillCam® was collected.
10	Pickin et al. (222)	2009	UK	Pharmaceuticals	*Evaluation:* Some problems with missing data arose for patients' assessment after the first year [of the scheme]. Without a control group, it is hard to say whether observed regressions of patients are due to the therapy under assessment or not. *Other:* During its implementation, the scheme was re-tendered but the original consortium undertaking the scheme (ScHARR Consortium) did not apply as they were not happy with the proposed arrangements for data access and publication rights. This led to a change in the

					organisation managing the scheme which in turn may lead to differences in the data collected. There were different interests and expectations between the Department of Health, the pharmaceutical companies, researches and patients regarding the scheme.
11	Adamski et al. (86)	2010	USA, Canada, Australia, EU	Pharmaceuticals	*Desirability*: Concerns include potentially high administration costs, lack of transparency, conflicts of interest, and whether health authorities will end up funding an appreciable proportion of a new drug's development costs. Pharmaceutical companies may be tempted to initially over price their new drug in expectation of price cuts/cost shifting downstream as the evidence base grows. Proposed schemes must be based on robust evidence for potential consideration. *Research design:* Future schemes must have realistic time scales, must not involve appreciable administrative burden if part of routine clinical care unless addressed, and must take cognisance of any likely changes in care during their lifetime, i.e. standard drugs losing their patent and/or clinical standards changing. *Implementation:* The administrative burden, lack of communication, and concerns with passing on savings have all been highlighted as key issues with current schemes for cancer drugs in the UK. There may exist a compliance issue—especially for long term chronic conditions. This issue must be fully addressed where pertinent else all key stakeholder groups will lose out. *Evaluation:* There are high ethical standards in the evaluation of proposed schemes. This includes the declaration of any contacts and conflicts of interest between experts and pharmaceutical companies that could potentially jeopardise evaluations. 'Exit' strategies must also be considered in advance should the effectiveness and/or safety of new drugs turn out to be worse in reality leading to their possible withdrawal during the lifetime of the scheme. *Other*: Concerns about [CED] schemes can be diminished if subsequent studies are undertaken by independent organisations. Transparency and ethical considerations include: 1) funding arrangements for any registries as well as administration costs must be transparent, 2) any ethical, legal, and clinical governance considerations must be fully addressed when proposing and developing future schemes (this includes issues of ownership of the data, especially if schemes are operated within health services, intellectual property rights and opportunities for appeal), and 3) future risk-sharing schemes should be open to all pertinent companies and not just selected companies.
12	Goeree et al. (107)	2010	Canada (Ontario)	MDs	*Research design:* By nature, CFFE is a lengthy complex process. It requires the creation of working groups made up of key stakeholders and opinion leaders who are involved in designing the study questions and methods from the beginning of the process. Then there is protocol development, sample size and site determination, case report form development, contracts with sites and investigators and dealing with multiple research ethics board submissions. Therefore, study initiation is often subject to contractual and legal delays. Well designed and conducted CFFEs are similar to any clinical trial and often take considerable time before results become available. With a typical

					three- to five-year political cycle, there is often tension between research and political needs as senior management of government typically try to enforce a quick turnaround of research studies and this is not always conducive to conducting the high quality evaluations that are required to inform health policy. *Implementation:* The biggest challenge for government is finding the resources to fund the infrastructure associated with an evidence-based decision-making platform and process, which can also include the option of funding for the conduct of CFEEs of technologies with uncertain risks, benefits and costs. The cost per evaluation is a reasonable cost for conducting primary data collection studies, however, the overall cost of such a program can become large if there are a number of CFFEs undertaken each year. *Evaluation:* The biggest concern from a research perspective is the compromising of scientific rigor because of time pressures and restrictions. *Other:* Differences in patient characteristics like demographics or rates of compliance with therapies, or provider characteristics such as level of expertise or training, or healthcare system characteristics like payment incentives or available infrastructure, can all affect whether, and to what extent, a technology works in a particular jurisdiction.
13	McCabe et al. (223)	2010	UK	Pharmaceuticals	*Research design:* Data collection and analysis should be conducted by an independent academic research group. The test criterion was accumulation of disability, and this is implemented using the physician assessed EDSS. The EDSS is a multidimensional measure of disability that relies upon a substantial amount of clinical judgement. In addition, it includes some patient self-reported elements. Thus it is substantially subjective, with associated concerns about both its inter- and intra-observer reliability. *Evaluation:* The Health Service Circular did not include plans for an evaluation of the scheme. An interim analysis of the MSRSS was undertaken in 2005. However, the report of this evaluation has not been published. The governance arrangements of the scheme do not appear to be designed to promote and protect the independence of the scheme. *Other:* The contract for running the MSRSS is managed by the MS Research Trust. The MS Research Trust also owns the data collected by the scheme and therefore has control over what the data can be used for. It is difficult to see how this arrangement provides the independence that the scheme requires. The manufacturers make contributions to the running costs of the scheme and have representatives on the MSRSS Steering Group. Clinicians undertake the clinical assessments that are used to assess outcomes within the scheme. Oversight of the MSRSS is provided by a Scientific Advisory Group, which includes representatives of the four pharmaceutical companies, the UK Department of Health Medicines and Industry Directorate, the MS Research Trust, two independent academics, and the Director of the Academic team operating the data collection and analysis project.

					As well as providing oversight to the operation of the scheme, this body appears to have had the right to withhold permission for academics to publish results from the scheme. *Other:* Seven years after the scheme was announced, published details are relatively sparse.
14	Menon et al. (105)	2010	Australia, Canada (British Columbia, Ontario, and Alberta), UK, USA	All technologies	*Research design:* The decision problem is rarely stated explicitly, thereby creating a risk that the design of the approach or model may not address the decision problem. It is also a failure to meet the standard of transparency that is increasingly recognized as an important characteristic of a fair public decision-making process. *Other:* The scheme's governance should ensure the independence of the scheme from any parties with a vested interest in its outcomes.
15	Mohr and Tunis (97)	2010	USA	All technologies	*Desirability*: Members of health plans express concern that AED could lower the evidentiary threshold for coverage without ever producing the robust information needed to make a coverage determination. By contrast, representatives from industry express concern that by involving health plans in the study design, evidentiary standards may be raised, and may thereby slow innovation and increase their costs of product development. In addition, payers are sceptical that AED will reduce costs in the long run. *Research design:* One critical success factor for AED was the need to clearly define the decision problem prompting consideration of AED, articulating a study objective that was directly targeted to the decision problem, and ensuring that the AED study was designed in a way that could feasibly address that objective. *Evaluation:* Evidence development part of AED rarely worked well when the policy was applied simply as a means of avoiding or delaying a difficult decision. *Other*: Patients and consumers may distrust the motives of payers in their efforts to support evidence development through coverage, and may assume that the primary objective is cost containment, and not a genuine effort to support early access to innovations and clinical research. As with any clinical trial, patients may choose to participate because they see the research study as a way to attain healthcare that they would otherwise not get. Here they only have access to treatment if they participate in research. There has been a healthy discussion of whether or not this is coercion.
16	Stafinski et al. (224)	2010	NS	All technologies	*Research design:* 'Deal breakers' (e.g. no, poor or uninformative data, resistance from providers, etc.) need to be specified from the outset. Some schemes have required unanticipated and significant commitments from providers and administrators who did not have the resources to handle the burden introduced by the scheme, placing the health system at risk of breaching contractual arrangements with manufacturers. In other cases, the 'right' outcomes were not identified until the scheme had been implemented, resulting in failure to capture the data needed to reduce the decision-making uncertainty. There is a lack of reliable funding mechanisms to support the data collection required as part of AED schemes. Some have suggested establishing public-private partnerships

					between payers and manufacturers, while others have stressed the importance of locating publicly funded research organizations who may be perceived as neutral and, therefore, better able to manage vested interests. *Evaluation:* Many of the AED schemes carried out to date have involved observational studies (e.g. registries). Some have argued that data from such studies are not sufficient for making coverage decisions, citing examples of mistakes made in the past when relying on observational data alone. Others have referred to past AED schemes as 'informative failures', lacking the ability to generate the quality of data needed to inform definitive coverage decisions. *Other:* In general, there is a lack of information on AED schemes in the public domain. Consequently, payers considering them as a potential policy option have little information upon which to base a decision. Furthermore, disclosure of the results of previous schemes related to a technology of interest may reduce duplication of efforts. The lack of information has been attributed to 'commercial confidentiality'. Mechanisms for increasing transparency around key components of the scheme (e.g. objectives, conflicts of interest, data collection management and oversight, etc.) that respect commercial interests are required.
17	Trueman et al. (98)	2010	NS	All technologies	*Desirability*: Payers may be interested in restricting coverage to a subpopulation and a CED program may be appropriate for defining the characteristics of the target subpopulation. CED could stifle innovation by creating a disincentive to develop new products for conditions for which the evidence base is not well developed. *Research design:* While evidence suggests that public sector coverage of research costs could be offset by price discounts or other pricing agreements, payers are concerned that publicly funded, evidence development could inadvertently incentivise producers to provide incomplete submission packages. Other experts believe that payer-funded research is necessary to provide control over research design and data. The preferred study design required to answer questions of evidence development has not been clearly defined, especially among the need for RCTs or observational/not experimental designs) A disincentive related to research design and funding is that restricted access to new products could inadvertently delay the collection and communication of new data because the treatment population would remain limited in size. Reaching consensus on standards for study design. *Other:* It may be unethical to withhold a potentially beneficial innovation from a subset of patients while providing it to another. There may be unfairness in the allocation of new treatments. If a potentially beneficial treatment is withheld from half of the eligible population as part of randomization, patient advocacy groups may be unwilling to accept this option, especially if the treatment has demonstrated safety and efficacy.

18	Espin et al. (225)	2011	EU	Pharmaceuticals	*Desirability*: There is no common approach across European countries regarding RSS for oncology products. Opportunity costs of implementing RSS should be taken into account and at the same time their impact on international reference pricing should be taken into account. *Research design:* European policy makers also identify some limitations and caveats, for instance: 1) additional work time, mainly for hospital pharmacists; 2) the need to have a well-designed and easy-to-use computer system; and 3) the need to ensure that the cumulative burden of schemes is manageable for the health system. *Implementation:* Implementation of RSS is considered difficult. For example, because substantial resources are required for following up on the schemes. *Evaluation:* The absence of an evaluation process by the different stakeholders involved undermines the effectiveness of such schemes. Many countries have not included any plan aimed to evaluate the schemes they have in place. *Other:* Several other general limitations and caveats were pointed out, e.g. there are potential problems with getting money back from the manufacturer and considerable resources are needed to ensure adequate implementation.
19	Garattini and Casadei (226)	2011	Italy	Pharmaceuticals	*Research design:* The scientific rationale of the non-responder criteria for each drug has not been made public, so it is impossible to discuss their reliability and applicability, also in the light of continuing scientific progress. For example it can be argued that time frames appear too short to allow a reliable assessment. *Implementation:* Payback for non-responders is centrally managed via an Italian Web database. A key question is whether all patients are registered in the national database: if a drug is administered outside the registry, the INHS reimburses the product outside any outcome-based arrangement, de facto. In addition another question is whether regions, which are financially accountable in Italy for healthcare expenditure (including hospital drug budgets), are really able to claw back refunds from manufacturers.
20	Levin et al. (102)	2011	Canada (Ontario)	All technologies	*Research design:* The total cost of each field evaluation is estimated at CAN$600,000 which includes protocol development and implementation and costs attributed to data collection, analysis and reporting. This estimate does not include costs absorbed by institutions or of the technology being tested. The process provides an opportunity to bend the diffusion—and, therefore, the cost-curves—for these technologies. Ideally, post-market studies should not restrict access to the technology while they are being conducted, but there must be an understanding that definitive funding will be predicated on the results of these studies. Results from the completed CED schemes raise questions regarding the future scope for CED. In some situations, the lack of external validity from efficacy RCTs may mean that definitive funding decisions should not necessarily be based on this evidence

					alone and that CED should be considered for technologies to test generalizability from RCT data before making definitive funding decisions. *Implementation:* Methodologies are needed to increase the scope and reduce timelines for CED schemes, such as the use of linked comprehensive and robust data sets and collaborative studies with other jurisdictions. CED schemes, before making long-term funding decisions, especially where there is uncertainty of effectiveness, safety or cost-effectiveness, should be increasingly funded by health systems. *Other:* There is a tension between conducting these evaluations expeditiously and ensuring the evaluations are methodologically rigorous and defensible. Furthermore, if the scope of CED schemes expands, greater efficiencies will need to be introduced through interjurisdictional collaboration, collaboration with industry partners, and by undertaking some studies in the pre-marketing phase. Policy makers and opinion leaders should be involved in the formative process of field evaluations to ensure relevance of the outcome to decision making and to influence adoption of the technology according to the results.
21	Lexchin (227)	2011	USA, UK, Canada, Australia	Pharmaceuticals	*Desirability*: Posing a series of questions is a necessary first step, but is not sufficient before embarking on a CED project. How should the questions be prioritized? Should the importance of the questions vary depending on the drug and the disease it is designed to treat? Different stakeholders will view the importance of questions differently, and some stakeholders such as professional associations and pharmaceutical companies may have preferred access to decision makers to make their case. Other relevant unresolved issues with CED are: to what degree should coverage decisions in CED take into account the preferences of patients, and how should those preferences best be determined? Should the focus be solely on the quality of the evidence, or should the potential budgetary impacts and clinical importance of the practice be considered when issuing a CED recommendation? How can decisions regarding CED be made in a transparent, consistent, and methodologically sound manner? Will CED create new uncertainties that reduce capital investments in healthcare technology innovation, as the pharmaceutical industry claims? *Research design:* When NICE recommends OIR, it generally specifies what sort of further research is required and the questions that should be answered, but not how to answer them—that is, it does not specify the particular research methodology. There also are no formal arrangements between NICE, industry, and the clinical research community for pursuing the research recommended by NICE in a timely manner. As a result, recommendations are hard to implement, particularly when there is no ongoing research to resolve the questions. On the multiple sclerosis scheme in the UK, the lack of a coordinated approach between NICE and other groups in this project has "made it difficult to ensure that the information collected will ever inform an update of the NICE guidance". Another question is: where will the money to undertake CED come from? Will the financing of the

					research be done publicly, privately, or through a combination of both? Who should undertake the management, data collection, and analysis? *Implementation:* Manufacturers will be especially reluctant to fund additional research if it involves head-to-head trials, since a negative result (their drug is less effective or causes more adverse effects) could lead to a significant decline in sales. In general, it is important for decision makers to move beyond regarding CED as a set if discrete ad hoc decisions and to commit to seeing this as a policy worth developing in a systematic way. To a certain extent, NICE in the United Kingdom and the CMS in the United States have already done this, but these organizations are largely restricted to making recommendations that a drug should undergo CED, with no power to act on those recommendations. Specifically, they do not have the resources or the authority to dictate the terms of the studies that drugs should be subject to, and they do not have the funds to commission CED projects. *Other:* To what extent is CED seen by policy makers, patient groups, and other stakeholders as a coercive way of linking insurance coverage to requirements for participation in patient registries and clinical trials? Should CED be used as a means of controlling costs? Even if CED is not a means of controlling costs, should cost be a practical and ethical element in prioritising among the many technologies that could be eligible for CED? How can agencies best convey to the general public and patients that CED is not about cost cutting and, in particular, what the benefits of a CED recommendation could be?
22	Mortimer et al. (99)	2011	Australia, UK, USA	All technologies	*Desirability*: FED schemes may have the unintended effects of lowering industry investment in evidence development and shifting research costs to public fund holders. In addition, it is politically more difficult for decision-makers to withdraw coverage – even if formally temporary – than to refuse coverage in the first place. *Evaluation:* Patients may be more motivated to exert political pressure to secure or maintain coverage of last-line treatment for life-threatening illnesses than for preventative or 'me-too' interventions. All else equal, the role of political pressure will also depend upon the extent to which the relevant technology has diffused across eligible patients and providers. Other barriers to delisting: there may be considerable inertia in clinical practice; particularly for interventions with a long-standing place in both formularies and clinical practice. For the most part, explicit delisting of drugs/devices from formularies has previously been carefully avoided by bodies with direct control of coverage such as the PBAC and MSAC; perhaps because of the anticipated political resistance. Instead such bodies have adopted a largely passive role and have relied upon clinicians to modify their prescribing practice to replace inferior interventions with more effective or better-tolerated alternatives as and when they become available. This passive approach has the potential to achieve a sharp curtailment of use where interventions are revealed to have a poor risk/benefit trade-off and where physicians

					face strong financial incentives and clear ethical imperatives to 'do no harm'. Where the required change in prescribing practice entails reversion to an older and possibly less-effective (but more cost-effective) intervention, the passive approach is unlikely to curtail the natural diffusion of an FED-listed technology. Control of coverage provides an obvious mechanism for curtailing use and diffusion, but additional enforcement action may sometimes be required to obtain the required change in practice. Provisional funding is often simply extended until adequate supplementary data are available to support a formal review; suggesting that an absence of further evidence may be enough to avoid delisting or restriction. Depending upon the extent of any restrictions on patient access during the FED period, a similar incentive may operate to delay evidence development if the default position is simply to extend funding until the data become available.
23	Neumann et al. (228)	2011	USA, UK	Pharmaceuticals	*Desirability:* Payers and manufacturers must be willing to confront the legal and financial complexities involved. Drug companies must have enough confidence in their claims of product effectiveness or efficiency to be willing to accept rewards or penalties based on observed performance. Danger for companies is that their products will be affected by factors outside of their control that can compromise outcomes—for example, inefficient health systems, local practice styles, or poor treatment adherence by patients. *Research design:* A major challenge has involved the specification and determination of treatment effects in nonrandomized settings. Only certain types of outcomes may prove suitable. Ideally, they should be objective, clearly defined, reproducible, and difficult to manipulate. They should be valid measures of the desired treatment effect and should not be confounded by patients' characteristics or other therapies. Agreements with shorter time horizons—no more than one to two years—have an advantage over longer-term arrangements, which are difficult to enforce and execute. An additional question is whether manufacturers can reap any "upside" if products offer unexpected benefits. Will payers, for instance, increase prices in the event of favourable outcomes? Risk-sharing agreements require high-quality information systems, data-bases, and operational and analytic expertise. Health systems often do not capture the level of clinical detail required to link payment to specific indications or patient subgroups. For many payers, linking payment to specific outcomes will require upgrades to existing infrastructure. *Implementation:* Barriers include high implementation costs, measurement challenges, and the absence of a suitable data infrastructure. Many of the steps involved in risk sharing—such as developing data collection protocols, negotiating arrangements, assessing product performance, policing contractual arrangements, and designing procedures to adjudicate disputes—can be costly and time-consuming. *Evaluation:* Gaining a clear understanding of the status and performance of the models is challenging, however, because little formal evaluation has occurred and because agreements may not be in the public domain. Successful agreements will also depend on favourable conditions for physicians and

					patients. For physicians, key factors are objective outcome measures, the availability of training to standardize protocols, and well-defined patient populations. Physicians will have more interest in participating if the process is economically advantageous to them, not simply another administrative.
24	Relyea-Chew (112)	2011	USA	All technologies	*Other:* The primary ethical questions raised by CED are: 1) whether Medicare beneficiaries who either cannot, or will not, participate in a clinical trial are deprived of a right to coverage, 2) whether Medicare beneficiaries are an appropriate population to study and upon which to base coverage decisions; and 3) whether conditioned participation in a CED study is coercive. Medicare beneficiaries are entitled to items or services that are "reasonable and necessary" and for NCDs have been issued. Not all eligible Medicare beneficiaries have access to or consent to participation in CED research as a condition of coverage. Fair distribution and access to benefits is an issue, especially as all potential CED research participants do not reside in geographic regions where new medical technology is studied. Proponents have rebuffed the concern that CED is unfair, arguing that when patients are not otherwise entitled to a particular treatment, there can be no harm to those who are unable to participate in CED research. For those Medicare beneficiaries who do participate in a CSP clinical trial, randomization and blinding pose the risk of not receiving the purported benefit but having to undergo the burden of the process, follow-up, and monitoring. Are Medicare patients appropriate subjects for CED research? Only individuals who qualify as research subjects may participate in a CED clinical trial, and the sample should include representatives of the Medicare population with the health condition described in the NCD. Enrolment in clinical trials is historically low for Medicare beneficiaries as they are predominantly women and people with many comorbidities. Is CED coercive? The main concern expressed by critics of CED is that the ethical principle of respect for persons is compromised by offering Medicare patients medical coverage in exchange for participating in research. The transaction, it can be argued, is coercive. A research subject's informed consent should be voluntary and knowing (competent); if given under coercion, or duress, it is invalid. When an elderly, ill patient lacks alternatives to treatment save enrolment in a coverage with appropriateness determination registry or CSP, arguably consent is not volitional. "Physicians," states an authority, "should be aware of how vulnerable patients may be to the coercive influence of unrealistic hope, especially those suffering from chronic, life-threatening disorders". Nevertheless, proponents of CED argue that Medicare beneficiaries could pay privately for access to innovative technology as an alternative.
25	Claxton et al. (96)	2012	UK (England/Wales)	All technologies	*Desirability*: The general issue of balancing the value of evidence about the performance of a technology and the value of access to a technology can be seen as central to a number of policy questions. Establishing the key principles of what assessments are needed for OIR or AWR recommendations, as well as how these assessments should be made, will enable them to be addressed in an explicit and transparent manner. Although [for legal reasons] AWR is not currently a

formal policy option available to NICE, it is able to issue specific research recommendations as part of any guidance and can link this to the timing of any reappraisal. Given the current directives and remit of NICE, the research recommendations issued as part of an AWR decision are not a mandatory requirement of approval. As a result, inevitably there exists some uncertainty following an AWR recommendation over whether or not the stated research recommendations will actually be conducted. It is important that policy provides appropriate incentives for manufacturers to conduct the type of research needed to support NICE guidance at launch. Although the NICE appraisal process may be well suited to identifying the need for evidence when assessing cost-effectiveness, these other critical assessments (the type of research and its priority) are not necessarily ones for which NICE and its Advisory Committees, as currently constituted, have particular expertise, not least because they reflect the decisions of those responsible for research design, prioritisation and commissioning.

Research design: The information required to assess whether sources of uncertainty will resolve over time requires information that is not commonly sought as part of NICE appraisal. It requires information about 1) likely changes in prices of the technology and its comparators, 2) the emergence of new technologies that might make existing ones obsolete or change their cost-effectiveness and 3) other relevant research reporting. In the UK, following an OIR recommendation, there are no formal arrangements to develop the research study required to reduce uncertainties. NICE does not hold a budget to commission research so unless it is publically funded by research commissioners it will be undertaken only if manufacturers conduct it, with the NHS contributing excess treatment costs. The lack of coordination also makes it difficult to ensure an update of the recommendation following production of new evidence. In the USA, although the CMS will cover the costs of a trial or registry associated with a CSP decision, there is currently no Medicare-specific funding mechanism for the additional data collection under CAD and hence there exists similar uncertainty concerning who will be responsible for paying for the additional data collection under CAD. In addition to general issues that need to be resolved to ensure the effective use of OIR/AWR policy options, there are a number of specific issues that need to be addressed. The design of the OIR/AWR study will ultimately determine its success. Perhaps the most important consideration emerging from the literature is the issue of which type of study is most appropriate for an OIR/AWR scheme. OIR/AWR research (unlike licensing research) is not confined to RCTs and, depending on the source of uncertainties, other types of evidence may be sufficient. The choice of study is ultimately context specific and related to the source of uncertainty; however, it may also be influenced by factors such as cost and availability of suitable patients, collaborating clinical centres and potential ethical considerations. Good routine data capture mechanisms have a potentially crucial role to play in the feasibility of any scheme and the development of health informatics could greatly reduce the cost of evidence.

	Whichever study type is chosen, the study design should not take place when the decision is made over who should pay for the study as this imposes restrictions. Clarification is also needed on how the evidence collected as a result of an OIR/AWR policy will be used in an updated coverage decision and also how many data are enough to inform subsequent decisions. Allowing prices to change as part of an OIR/AWR scheme also further extends the options available to decision-makers. *Implementation:* How the NHS and manufacturers are likely to share the value of evidence might inform whether manufacturers should be expected to conduct the research specified in AWR or OIR guidance or contribute to the costs of publically funded research that may ultimately benefit their product. Two issues need to be considered: 1) the resource constraints on publically funded research may mean that other research priorities (often without commercial interest) may be more valuable to the NHS and 2) the success of AWR recommendations when manufacturers are asked to conduct the research will depend on whether NICE and/or the Department of Health are able to establish contractual arrangements as part of an AWR recommendation, that is, arrangements that can be monitored and enforced with credible penalties to ensure that agreed research is conducted and in the way intended. At present, NICE does not have a credible mechanism because removing approval of a technology simply because recommended research has not been conducted is not considered ethically appropriate or a credible threat. The assessments that need to be made can also be used to consider what would be the value of 1) being able to conduct research while a technology is approved, 2) making evidence that is needed by the NHS available at launch and 3) being able to acquire evidence more quickly. This might inform a range of policies, such as early advice, public investment in early transitional and evaluative research or better data collection or information systems that might make AWR possible. Understanding the relationship between the time taken for research to report and the value of the evidence to future populations can also help to inform 1) investments that might make research findings more quickly available, 2) the trade-off implicit in the choice of alternative research designs and 3) identification of those areas where, if research is to be undertaken, there must be confidence that it can report quickly. There are many issues that need to be resolved to enable the successful implementation of an OIR or AWR scheme both generally and also within the specific constraints of NICE. Central to this is the need to clarify the objectives of these schemes and the relevant criteria for their use. The general lack of clarity on the principles and criteria for using these schemes is reflected in many commentators' views that the development and use of schemes internationally has appeared to be rather ad hoc to date. Practical issues are associated with significant opposition from the clinical community, the significant level of funding required, the length of time required to complete data collection and limited access for patients in remote areas. The importance of interagency collaboration, achieving consensus on acceptable quality of evidence,

	external peer review, predefined clinical benefit and determining who pays for treatment was also apparent. There also remain other important challenges, including the need to ensure that research is actually conducted and is fit for purpose, as well as ensuring that the process is undertaken in a legal, ethical and acceptable manner. Another important consideration is that these schemes need to be designed in order to develop appropriate incentives to produce evidence in a timely fashion and strategies need to be put into place to ensure that the research is actually carried out. Within the UK, the importance of interagency collaboration has been highlighted as a key issue in ensuring the success of OIR/AWR schemes. Acquiring appropriate evidence following an OIR or AWR policy is of paramount importance. OIR and AWR policies allow evidence to be generated specifically to inform decisions, a role not intended for traditional regulatory trials. This raises a number of issues and potential challenges related to the design and funding of further research studies. First, there is currently very little in the way of formalised arrangements following an OIR/AWR recommendation in many countries. One exception to this are CFFEs conducted in Ontario, which have a specific funding stream (albeit modest) covering the evaluation by PATH and additional monies for the fieldwork itself. However, in many instances this budget is not sufficient to cover the full costs of a CFFE and therefore other avenues must be explored, such as cost-sharing. A key issue identified in determining the success of these schemes is the development of working partnerships between stakeholders (clinical community, decision-makers and manufacturers). Related to this is the issue of obtaining funding for OIR/AWR studies and establishing who pays for the research. Without secure funding the research may never be undertaken and thus the uncertainties leading to an OIR/AWR recommendation will remain. [In this review] a number of implications for future studies were also noted, in particular the importance of interagency collaboration, achieving consensus on acceptable quality of evidence, external peer review, predefined clinical benefit, determining who pays for treatment and establishing longer-term follow-up. Investment and reversal costs have also been identified as relevant considerations in the existing literature. In particular, NICE needs to determine whether a fully supportive decision (as opposed to OIR) would lead to significant irretrievable costs of implementation and if it would lead to termination of ongoing research or prevent future research. The need to consider what an intervention that has been subjected to a CED scheme is displacing before any assessment of potential cost savings through such schemes can be made [is also highlighted]. Because publically funded research also consumes valuable resources that could have been devoted to patient care, or other more valuable research priorities, there are a number of trade-offs that must be made. In making these trade-offs consideration also needs to be given to uncertain events in the near or distant future, which may change the value of the technology and the need for evidence. When considering OIR or AWR guidance there must be some assessment of 1) the type of research needed to address the key uncertainties, 2) whether or not this will be regarded as ethical and

	can be undertaken while the technology is approved for use, 3) whether or not it is likely to be a priority for public funding and be commissioned and 4) when it is likely to report. *Evaluation:* A further ethical issue raised specifically by AWR recommendations involves the mechanisms used to give 'teeth' to the research requirement: how will NICE ensure that the relevant research is carried out? One option has a subtle ethical dimension. NICE might threaten that, if the research is not satisfactorily completed (e.g. by the relevant manufacturing company), the intervention would cease to be made available on the NHS. Although this would provide an incentive for the manufacturer to carry out the research, it raises the following problem. At time T(1), the AWR decision is made, that is, the intervention is funded by the NHS. Suppose that, at time T(2), the time when NICE reconsiders the decision to fund – the research has not been carried out (or has failed to provide any further relevant information). At time T(2) NICE will be making a decision on exactly the same information and evidence as at T(1). In this case it would seem that NICE should make exactly the same decision, namely, to provide the intervention on the NHS. But if NICE decides to reject the intervention (on the grounds, for example, that the manufacturer had failed to carry out the relevant research) then patients could claim unfairness. The unfairness is that they were provided with the treatment on the NHS between T(1) and T(2) but not after T(2) even though the evidence is exactly the same in both situations. Among the appraisals with OIR/AWR recommendations in the final guidance, ten were later reviewed by NICE, including two that were incorporated into NICE clinical guidelines. In the majority of reviewed appraisals (n=7), new evidence informing the OIR/AWR recommendation was available for the review. In four of these reviews, the OIR or AWR restriction was removed and the technology was recommended routinely. In two cases the additional evidence was considered insufficient to warrant a change in the OIR recommendation. In the remaining appraisal the OIR was revised so that some technologies within the class were recommended routinely whereas OIR recommendations were issued for others. In all cases the changes in the guidance were owing to the new data relating to the evidence gap identified in the OIR/AWR recommendation. In three cases no new evidence was provided on the OIR/AWR indication. If research is recommended in OIR or AWR it might not be undertaken by manufacturers or commissioned by research funders. Even if undertaken or commissioned there is no guarantee that research will be able to recruit or it may not complete for other reasons. This indicates that the potential gains depend on a judgement of whether the research recommended as part of OIR or AWR will be successfully completed. *Other:* The potential ethical issues arising from the use of OIR/AWR schemes is another important theme emerging from the existing literature. For OIR, the issue of compulsory participation is often raised as a concern. Also, because of practical arrangements under OIR, treatments may not be available in all areas, causing geographical inequalities. If a RCT is commissioned following an OIR

recommendation, this raises a greater issue in terms of participation than a simple registry. It has been argued that denying access to a treatment demonstrated to be effective (however uncertain) is unethical. Patient advocacy groups may also be unwilling to accept this policy especially if the treatment is considered to be safe and efficacious. These issues have important implications for both the design and the successful conduct of research. There are general ethical issues associated with OIR/AWR schemes. First, they may be beneficial for future patients but they can impose significant opportunity costs on current patients. Some individuals in the present population may benefit from the research condition because they will also be members of the future population. This will not be true of all, so the issue of balancing the interests of some individuals in the present population against some individuals in the future population remains. The second issue is under what circumstances the present population is disadvantaged by the research condition compared with the alternative recommendation that NICE might make. There are also ethical issues that are specific to OIR or AWR schemes. OIR schemes: patients outside the trial, and participants randomly allocated to the standard arm of the RCT, would be denied what may almost certainly be the better treatment for their condition. A criterion established in research ethics for the legitimacy of carrying out a RCT is that there is substantial uncertainty as to which of the treatments being compared – that is, an innovative treatment and a standard treatment – is the more effective. This is sometimes known as the principle of equipoise. The principle is meant to capture the intuition that no one – patient or participant – should knowingly be offered less than the best treatment for their condition. OIR decisions may be made when there is such substantial uncertainty. In such cases, the principle is respected and ethical review of the relevant research poses only issues already considered as standard by researchers and RECs. However, NICE is considering recommending OIR when the intervention in question is clearly superior to alternatives, but the degree of its superiority remains uncertain, and so its cost-effectiveness is uncertain. Evidently, in this scenario researchers are not in equipoise about the relative effectiveness of the two interventions. The substantial uncertainty relates to whether the more effective, but more expensive, treatment produces sufficient extra benefit compared with alternatives for it to be recommended by NICE, but it does not relate to whether it is more effective. Is it permissible to flout the principle of equipoise concerning effectiveness and give an OIR decision in such circumstances? An intuitive response is that patients are harmed by an OIR decision that denies them the best-known treatment for their condition in the interests of research. Given NICE values, principles and practices, it is perfectly feasible to conclude that the harm perpetrated by OIR is justified by the benefits to future patients of a better evidence base for allocation decisions. Another research ethics principle relevant to the sort of OIR decision under discussion is that competent patients have the right to consent to participate in, and withdraw from, a research project. Conversely, it is impermissible to coerce competent patients to participate in research. In OIR, the

					patient can have the more effective intervention on the NHS only if he or she agrees to be a research participant. Does this coerce patients to participate in the trial? A further ethical worry is that OIR decisions result in inequity because participants in one arm of the trial receive better treatment than both those in the other arm and those not participating in the research. AWR schemes: Two established principles of medical ethics are that competent patients have a right to confidentiality (including a right to decide whether or not to disclose their personal medical information) and a right to informed consent to participate in research (including a right to decline to participate in, or to decide to withdraw from, a study with impunity). It might be argued that some AWR decisions transgress these rights because the required research may involve collecting data on long-term outcomes and adverse events on patient registries (or some similar system of epidemiological data collection) without the explicit consent of patients. [Furthermore], any assessment of the potential benefits should account for the fact that patient populations will not benefit from the results of research until they are available and change clinical practice.
26	Walker et al. (93)	2012	NS	All technologies	*Desirability*: Whether CED schemes are recommended depends on both the characteristics of the technology (whether it is expected to have a positive net benefit, whether evidence can be generated following reimbursement, and whether there would be a cost in reversing the decision at a later data) and the range of authority of the purchaser (whether they can delay a decision or review it at later date, whether they can negotiate price, and whether they can ensure that research is actually conducted). When the purchaser is unable to delay a decision or to reconsider it at a later date, its options are heavily limited: it should accept or reject, and the coverage decision should be based solely on the expected value of the treatment (whether assessed explicitly or implicitly). For technologies that are expected not to be of value given existing evidence, the purchaser is generally limited to rejecting the technology or trying to ensure a reduction in the price. More interesting are the choices available when a technology is expected to be of value given existing evidence. As the purchaser's range of authority expands so also does the coverage options available, and the choice should be based on more than the expected value. When further research is worthwhile but cannot be generated following a decision to pay for use, the value of the evidence should be weighed against the expected benefits of immediate coverage to patients. When there are large reversal costs, the purchaser should weigh the expected benefits to patients against the probability that the decision will be wrong and should be changed with the reversal costs incurred. What should be apparent is that the choice between coverage options always revolves around the trade-offs between the expected value of the treatment given existing evidence, the amount of uncertainty (and how much of it can be resolved through research), and the magnitude of costs of reversing a decision.

27	Bishop and Lexchin (109)	2013	Australia, Canada, UK, USA	All technologies	*Desirability*: Perceptions of CED differ widely between the interviewees according to their roles and involvement in research, decision-making, and policy development. At the core of the discussion are differing perceptions regarding both the definitional and methodological issues and, as well, the scope of CED. Particularly three elements are outlined: what is worthy of research, what constitutes uncertainty and what constitutes evidence? Two key tensions were identified: the first being how the various stakeholders can affect political decisions around initiating a CED project and, the second, the influence of the pharmaceutical industry in CED (e.g. pressure to initiate a scheme, conflict of interests when the industry plays a role in funding, collecting the data and evaluating the scheme). *Implementation:* A decision maker laid much of the blame for the difficulties in doing CED on physicians' problems in collecting data and asking people to sign informed consent forms. There was also a structural issue that was identified, namely the access to data and the privacy issues that this entails. There was a general agreement that, in order to successfully undertake CED, it is necessary to have better access to data and to be able to link databases. Registries were identified as one key method of collecting the needed data by all categories of interviewees. However, with registries comes the problem of finding a sustainable funding model for them, verifying the accuracy, reliability and completeness of the information and, as well, ensuring the privacy of the information that is stored. *Other:* CED lacks a governance structure and a systematic approach. Beyond ascertaining uncertainty, three aspects of governance or approach to CED act as barriers and contribute to the difficulty in establishing norms: 1) role of different stakeholders: With respect to CED, researchers took contradictory positions around where the leadership should rest (e.g. clinicians, government) and which stakeholders should be involved (e.g. industry, department of health, patient groups and clinicians); 2) translation of research into policy: The role of different stakeholders is inherently linked to the difficulty in translating research into policy or implementing policy. This process requires guidelines and a firm grounding; 3) financial considerations also are barriers. Justifying the continued flow of funds requires policy standardization and a formal agreement of the strategy to be put in place; currently it is a cyclical schema that has a lot of holes. Indeed, making conscious and collective decisions around evaluation and cost and measuring certainty are highly difficult. The politics and history undergirding the lack of agreement acts as a strong barrier to doing so and impacts the translation of CED. Another barrier is: Corporate influence and overt politics in CED. The translation of evidence into policy is riddled with political and economic considerations, both the overt political process involved in CED and the role of the pharmaceutical industry. The most explicit evidence of relations of power comes from the hierarchy of roles in the decision-making process. Political influences play a role in determining where the money for CED will come from and where the ultimate decision-making comes from. There are imbalances and risks inherent in the

					governance of CED. In some instances agreements make "it possible for parties of the scheme to rewrite the rules subsequently," which is a dangerous practice that places the resources that the society or health system puts into the scheme at increased risk. Finally, researchers tended to blame decision and policy makers for the trouble they encounter in translating the evidence gained through CED into policy. Not only because of the perceived flaws in the research. The role of the pharmaceutical industry: Decision makers, policy makers and researchers all voiced some level of criticism about the role of the pharmaceutical industry in CED.
28	Garrison et al. (82)	2013	NS	All technologies	*Desirability*: There has always been considerable uncertainty at product launch about the ultimate real-world clinical and economic performance of new products, but this appears to have increased in recent years. PBRSAs represent one mechanism for reducing this uncertainty through greater investment in evidence collection while a technology is used within a healthcare system. *Research design:* Which research design is most appropriate depends on the nature and type of the uncertainty that the PBRSA evidence collection is trying to address.: uncertainty about whether the medical product or service will be used in the right patients or uncertainty at launch about clinical or economic outcomes (effectiveness vs efficacy, final outcomes vs. surrogates, or about the size of cost offsets). GRP in PBRSAs should build on previous GRPs for specific types of studies. What is important for a PBRSA is that the study is designed to answer the specific uncertainty that most increases the likelihood of a bad decision. GRP requires outcome measures to be selected with care. They should be clear, measurable, objective, realistically achievable [in relation to the duration of the scheme], and relevant. An important part of the design of any PBRSA is to define the metrics by which the success of the scheme can be assessed. Process indicators of success of the scheme should relate to the research questions relating to the design, implementation and evaluation of the scheme and include the questions: are the intended outcome measures collected, was uncertainty in associated parameter estimation reduced for the outcomes that were the focus of the scheme, did the scheme run to budget and time, was the integrity of the design/estimation maintained, did the governance arrangements work well, did the success to underpin a decision with further evidence prove successful? *Implementation:* Additional evidence collection is costly, and there are numerous barriers to establishing viable and cost-effective PBRSAs: negotiation, monitoring, and evaluation costs can be substantial. For good research practice (GRP) in PRBRSAs, it is critical to match the appropriate study and research design to uncertainties being addressed. Aspects of good implementation follow from clarity on the desirability of using a PBRSA and on the type of evidence being collected as part of the PBRSA: is the scheme measuring appropriate outcomes, are the costs acceptable, is the time horizon realistic, are the funding arrangements clear, how is responsibility for undertaking data collection and analysis allocated, is the data collection efficient, what will be the process for reviewing and analysing

					the evidence to make a revised decision on price, revenue, or coverage, will discounts or rebates be paid during the course of the scheme (e.g. based on provisional results)? Good governance processes are also essential. The need for formal governance structures is greater when multiple stakeholders are involved in PBRSAs. PBRSAs that have multiple stakeholders and/or the involvement of taxpayer funding of research as well as payment for the intervention should be governed through a diverse steering committee that includes patients, manufacturers, disease advocacy groups, professional associations, and other major stakeholders, with a channel for receiving broad public comment. Transparency of process is also important. A more controversial issue is transparency about pricing arrangement. Effective governance requires that the governance committee has a clear charter specifying who is involved and their respective roles, signoff procedures for research design and study protocols between the governance committee and principle investigators, the governance agreement specifies the aims of the PBRSA, who has access to the data, who can publish, the process for vetting manuscripts, and the final steps for managing and disseminating the research (i.e. the stopping rules and how the results will be used), funding arrangements are clearly specified upfront, the agreements spell out a process to ensure data quality, there is a declaration of the conflicts of interest. Most PBRSAs involve a public payer. The information generated as part of PBRSAs has public good aspects, bringing ethical and professional obligations, which need to be considered from a policy perspective. *Evaluation:* Whether the PBRSA has achieved its objectives and was good value from a health system perspective is linked to the desirability of the scheme and can be addressed from multiple perspectives: manufacturer, patient, payer, provider, and society. A comprehensive evaluation will therefore need to consider multiple perspectives. The ex-post evaluation of a PBRSA should [...] be a multidimensional exercise that assesses many aspects, including not only the impact on long-term cost-effectiveness and whether appropriate evidence was generated but also process indicators, such as whether and how the evidence was used in coverage or reimbursement decisions, whether budget and time were appropriate, and whether the governance arrangements worked well. As an innovation in and of themselves, PBRSAs should also be evaluated from a long-run societal perspective in terms of the impact on dynamic efficiency (eliciting the optimal amount of innovation). Because one cannot assess the value of information generated by a single CED scheme directly post hoc, there is a need to rely on process indicators of the schemes success. The scheme for PET scanning demonstrated the difficulty of rescinding coverage once it is offered provisionally through CED.
29	Brügger (87)	2014	Australia, Belgium, Canada (Ontario),	All technologies	*Desirability*: From a policy making perspective, CED schemes tend to be used for three purposes: 1) to generate evidence for a promising technology, 2) to monitor use or control volume and 3) to help innovations enter the market. The 'classic' purpose of CED is to generate evidence that is missing but a 'no' decision does not seem to be appropriate, i.e. balance the goals of enabling timely access

	France, Germany, Netherlands, Sweden, Switzerland, USA	for patients to medical innovations and of guaranteeing value for money under these difficult circumstances. In Canada (Ontario), there are two types of CED schemes: pre and most market. Pre-market programme is called 'excellence in clinical innovation and technology evaluation' (EXCITE), which is used to evaluate new technologies. In the post-market situation, field evaluations are used to avoid diffusion of a technology where there are still considerable uncertainties regarding effectiveness or cost-effectiveness. In France, the OWR option is used to supervise real life use of drugs and MDs or procedures. They are set up if there is a concern regarding medical, economic or organisational questions. The OIS option is used to bridge the time between market authorisation (CE mark) and approval of the Haute Autorité de Santé (HAS), the French HTA body. In Germany, CED schemes were set up for reasons of 1) better decision making and 2) helping industry to innovate. In the Netherlands CED is done to generate effectiveness data. In the US, under CED the health technology is financed only for the people participating in the study, which corresponds to an 'only in research' (OIR) regime. There are no clear criteria as to when CED has to be initiated. However, according to the interviewed expert, evidence that the technology is promising has to exist if the evidence base is still too vague for a clear decision. In the US the FDA is the regulatory body not only for drugs but also for devices. For devices, the requirements for market authorization are a lot stricter than they are for getting the CE mark used in Europe. Therefore the FDA may require further studies, sometimes under so called Investigational Device Exemption (IDE). This allows the investigational device to be used in a clinical study in order to collect safety and effectiveness data while it is reimbursed for the patients participating in the study. *Research design:* The OIR option is much more restrictive than the OWR option and therefore has the characteristics of a decision that means 'no, unless' rather than 'yes, but'. If a randomised controlled trial is chosen as the appropriate research to be carried out, it further means that patients who chose to take part in the study cannot be certain they will get the new technology because of randomisation. This might cause problems if there are strong patients' preferences for the new technology or the comparator. Consequently, there may be concerns about coercing patients to enrol in research studies. In all cases where no research is set up, the OIR option becomes a straight 'no' from a patients' perspective. The OWR option is less restrictive and such a decision is clearly of the type 'yes, but' because all indicated patients have access to a technology. There are several version of this option. It can mean that 1) data from all patients who receive the new technologies is collected in a register, 2) a randomised controlled trial (RCT) is conducted but not all patients receiving the treatment must take part in the study, and 3) results of an ongoing study is awaited which is carried out in the respective country or elsewhere. [One problem with CED schemes in the UK (England/Wales)] is that only observational (mainly through registers) are collected so far. [...] all evaluated technologies under this programme have been interventional procedures, for which the

	scientific evidence is typically weak and there is no equivalent for a thorough market approval process. Unlike in the US where the FDA can demand solid scientific evidence for market approval of high risk devices, the CE mark that is required in Europe is a far lower hurdle to take. The first attempts of CED in Australia were of limited success. The experience showed that either the research did not happen and the technology was funded anyway or the research did not deliver the evidence but the funding could not be stopped. [In the Netherlands,] before a [CED associated] study is started, a signed letter of both the relevant medical association and the relevant patient association has to be received that says that they agree with the study design. [In the Netherlands,] patients participating in an RCT have to be willing to be randomized to qualify for reimbursement. According to the [interviewee] this has caused some discussions among doctors but, the [interviewee] explained that this was in line with the medical principle of 'primum non nocere' (first, do no harm) and it would be unethical to provide a medical service where the risk-benefit profile compared to current standard of care is not known outside the context of research. *Implementation:* Costs and complexities of data collection (particularly project management and co-ordination) seem to be one obvious obstacle. Another issue concerns the mandate of an HTA organisation to commission (additional) research. In the UK, NICE does not have this mandate. The first attempts of CED [in Belgium] have not been successful. The problems were that compliance with data collection by doctors was weak, the monitoring of the study was poor because of lack of staff and there was no uptake of the research in the decision making process later on. The time frame during which the research is carried out should be kept to a minimum in order to keep it relevant to the policy decision makers. In France, there are some challenges with OWR schemes studies. The main problem is that the doctors are reluctant to provide the data that are requested. There is the possibility of sanctioning, but [according to an interviewee] this does not seem easy to do. Regarding [three] OIR schemes, there were difficulties in all of them. In the case of one medical device another study was already going on and in the case of the other medical device there was no agreement between the authorities and the manufacturer regarding the type of study. The required study by the authorities was seen as too expensive by the manufacturer. Because of these initial practical problems with OIR schemes, a revised law was introduced in 2013 to clarify three points that were seen as the major obstacles in the beginning: 1) clearly stated selection criteria, 2) clear relationship between all actors and 3) limitation of the duration of the study under forfait innovation. Learnings in Germany from past CED schemes are that Doctors play a crucial role. Incentives for doctors have to be set. Doctors have to be reimbursed for their contribution to the study. Professional societies and associations should be included. [In addition,] a big problem is recruiting the patients. If possible financial incentives for patients should be set, for example reduction of obligatory deductibles for doctor visits. [And] a professional research organization has to run the study that guarantees

compliance with GCP (Good clinical practice) principles and research governance, which also includes ethical aspects. In the Netherlands, the mandate for CED is stated in the law and each health technology is listed. For every new intervention that falls under CED, the law has to be altered each year. This means that the CED decision in each case is taken by the minister of health following a request by the Dutch Healthcare Institute (ZIN). A CED programme can be set up for a maximum of 4 years. RCTs or registers are used depending on the question and the feasibility. Issue regarding MDs concerns market access of other devices during a CED scheme and study. In Sweden, there is no authority responsible for oversight of the research [as] supervision of the research may create conflicts of interest for a HTA body [as, according to the interviewee,] they need to keep the image of being a helper for a better quality healthcare system. In the US, under CED the technology is paid by Medicare/Medicaid whilst the studies are financed either by the manufacturer or by a research funding body. Registers are often funded through user fees. The interviewed expert explained that a close collaboration with industry and physician specialist societies is crucial for success of CED. [The interviewee] explained that generally a maximum of four years until study completion and final decision seems to distinguish successful from unsuccessful cases. The following principles are derived from experiences in the countries included in the literature review and interviews: 1) there must be a clear mandate for CED and a clear process, 2) research much be clearly linked to coverage, 3) the CED recommendation should be carefully made and with a clear indication of the evidence gab, 4) research must be carried out in a scientific robust way, 5) There must be feedback of the study results to the decision making, and 6) there must be dialogue between stakeholders and agreement. If any one of the principles is violated, CED has been shown not to work satisfactorily.
Evaluation: Experiences with CED for non-drug technologies in different countries have been mixed. In some countries the use of CED has been more successful than in others. Success in this context means that 1) appropriate and scientifically sound research is carried out, 2) that the technology is only funded according to the conditions during the CED phase, 3) that the results of the research are fed back to the decision making, and 4) that the funding and use of the health technology can be adjusted according to the new evidence. Most countries can be assigned to two categories: 1) there are countries where CED was used for some time but where the use has been reduced or practically stopped because the approach was not satisfactorily working such as Australia, Belgium, and Switzerland. Problems were that data collection did not happen, that research did not answer the initial questions or the research did not feed back into the decision making. All these countries are currently thinking about redesigning and restarting the process, 2) There are countries where CED was started or restarted recently with a clear focus and process, or where the first experiences were positive such as Germany, the Netherlands and the USA/CMS. The Netherlands and the US restarted CED after partially unsuccessful experiences. Two countries do not fit into this pattern just

					described. Ontario, in Canada, a particular CED approach seems to be working quite well, potentially because of having a champion who is driving the programme. Finally, in England there are several different approaches to CED. In most cases of CED in Australia the funding just continued. There were only two examples where the committee decided to withdraw funding after the result of the study carried out after a CED decision were not as expected. The first was for kyphoplasty/vertebroplasty following publication of a randomised controlled study published by an academic group. The second was for hyperbaric oxygen treatment (HBOT) for nondiabetic ulcers. That decisions caused a heated discussion with doctors, patients and politicians and a public debate in the media took place. Since 2009 there were no new cases of [CED schemes] recommended in Australia. It is likely that CED (interim funding) will be reinvented in a better form in the future in Australia. The interviewed expert names the following factors needed for success: a clear mandate, a binding agreement with the applicant, clear measures of success and failure, and an explicit communication strategy that explains the logic of CED and its possible outcomes. The problem with the Swedish system is that there is no systematic feedback of the research results to the decision making process. In Switzerland, the CED process evolved over time but lacks uniform standards. In the US, there have been some successful cases such as PET scanning for oncology where the study was terminated and final decisions were made. So far there has been only one final 'no' decision after CED which was PET scanning for prostate cancer. The decision was well accepted by the medical professionals since the relevant medical society was involved in the decision making.
30	Foster et al. (110)	2014	USA	MDs	*Research design:* Not all CMS patients automatically qualify for CED as they may not have the required coverage. There is an inevitable trade-off between acceptable patient burden and required effort for data collection and regulatory procedures. *Implementation:* Identifying and counselling potential participants and obtaining informed consent require considerable effort and patients may decline to participate or may prematurely withdraw. Frequent reconsent or challenges occurring when the FDA changed FDG-PET regulations mid study were unforeseen. Lack of CED-study experience by clinical staff, staff turnover, billing requirement complexity, and inconsistency of nonresearch and research requirements added up to significantly more time and effort than anyone could have predicted—for a study that had no funding for administration. *Evaluation:* Final CMS decisions are critical because private insurers often deny reimbursement while CED remains in effect. Unfortunately, few technologies assigned to CED have emerged, in part, because there are no clear standards for revision. In the case of FDG-PET, for years, CED has prevented meaningful patient access. Over time, patients and healthcare providers become increasingly frustrated with the additional requirements and, as interest wanes, collecting requisite data becomes more difficult.

31	Launois et al. (229)	2014	NS	All technologies	*Research design:* There will be discordance between the expected and the observed result of a scheme because the external performance reference is built on findings from controlled trials, while performance itself is confounded by uncontrolled real life factors. The payer and the manufacturer might be able to set up a performance threshold based on clinical trial results. Yet, the uncertainty of translating knowledge originating from a controlled clinical trial environment into the standard to be expected in clinical practice could subject pricing decision-making to pure chance. Moreover, P4P evaluations do not allow the payer to understand: 1) How and why a treatment could have interrupted natural disease progression, 2) Whether the observed changes in the patient's health are directly and exclusively due to the administration of the product, 3) Whether the innovative product is comparatively effective vis-a`-vis alternative treatments. *Evaluation:* P4P schemes lean towards the concept of normative research (where judgment is based on the comparison with pre-established criteria or norms). However, although norms are negotiated through a contractual agreement, this does not alter their legal and unscientific content […]. By measuring performance, a treatment could reach the pre-established target value, but the true effect of the treatment may not be the cause behind the value attained, rendering performance-based evaluations unreliable. […] When using P4P contracts, the uncertainty surrounding the demonstrated effectiveness/ efficiency of a treatment remains. The negotiated and observed performance targets may be equal, without elucidating whether this equality is directly and exclusively attributed to the treatment. Performance studies […] do not envision a counterfactual in their design, posing a problem for P4P agreements where a drug is reimbursed on a case-by-case basis […], and the payer does not know the true contribution of the drug in attaining the outcome given the inherent lack of comparative evidence in the design. When reimbursing each individual performance target attained, the true effectiveness of the product in question cannot be known due to the absence of a counterfactual. When a P4P reimbursement commitment uses an intermediary endpoint as the external performance target, as most Outcome Guarantee contracts do, an additional problem arises. Ideally, if a reimbursement decision is based on intermediary performance targets, this target should show a strong correlation with the clinical parameters that it intends to replace. Only then, can the intermediary target be considered a valid surrogate criterion. In most cases, defined intermediate endpoints do not necessarily lead to clinical endpoints.
32	Martelli and van den Brink (95)	2014	France	MDs	*Desirability*: When innovative MDs are launched on the European market there is generally little clinical evidence regarding both efficacy and safety, both because of the flaws in the European system for regulating such devices, and because they are at an early stage of development. To manage the uncertainty surrounding the reimbursement of innovation, several European countries have setup temporary funding schemes to generate evidence about the effectiveness of devices. Only the

					CNEDIMTS is empowered to determine whether a device meets the criteria of the French CED. No official definition of an innovative product or intervention is available and this is consequently left to the interpretation of the committee. The transparency of CED processes should be enhanced, in particular by clearly stating the selection criteria for devices that may benefit from such approaches. OIR/AWR recommendations, like the French CED, are perceived by manufacturers as an additional delay in making innovations generally available. *Evaluation:* The leading French MD industry association indicates that the process for innovations is much too slow. Indeed, three health technologies were selected by the CNEDIMT since December 2011 with no further decision taken. Also, CED decisions need to be made more quickly. Effective collaboration between all stakeholders is undoubtedly the cornerstone of success for the CED approach. *Other:* The French Ministry of Health has to specify requirements for studies, but the law is unclear on this point and clinical and/or cost effectiveness studies may be involved. With the French CED, manufacturers receive financial support for two years, but the law does not specify what happens at the end of this period.
33	Garattini et al. (108)	2014	Italy	Pharmaceuticals	*Desirability*: AIFA published the revenues of MEAs for the first time. The total theoretical pay-back was €46.3 million. However, the report pointed out that one third of this could not be clawed back because of disputes with pharmaceutical companies (22 %) or late requests by hospitals (11 %). The total direct cost for managing these registries should be around €1 million according to what AIFA reports on its website, although this might be an underestimate since other costs (e.g. IT maintenance) must also be considered. Last, but not least, a cost item to be included should be the hospital consultants' and pharmacists' time for completing the forms. It is very likely to be considerable, but it is hard to estimate since at present no information is retrievable on the number of forms. *Evaluation:* To our knowledge, no published report has included clinical data on drugs subjected to RS/PbR agreements and, more in general, under AIFA registries, except for a summary report very early in 2007. Going through the forms referring to the patients' clinical status (all available on the AIFA website), it is clear that they closely reflect the approval indication, hardly asking for any additional information useful for an extended clinical assessment. So the information collected is unlikely to contribute to the existing evidence on the drugs under these agreements, beyond self-certified validation of appropriate prescription by the prescriber.
34	Drummond (88)	2015	NS	All technologies	*Desirability*: Some types of uncertainty (e.g. over whether a previously unrecognized adverse effect may emerge in the long term, or whether changes in market conditions might cause the price of the technology to drop in the future) cannot be easily resolved by further study. Therefore, one should not embark on PBRSAs to postpone a reimbursement decision that should be taken today, albeit

					with the possibility of revision later. Since it is extremely difficult, although not impossible, to secure agreement to a randomized study once a technology is reimbursed, it is worth considering whether convincing evidence on effectiveness can be gathered in an observational study. It is also important to be sure that the outcome of interest is largely influenced by the technology concerned. *Research design*: Two timelines are important: the time required to initiate the scheme and the time required to conduct the necessary research. Those planning PBRSAs should be wary of schemes that require a follow-up longer than around 3 years. This is because the kinds of policy questions that such studies inform have the habit of changing, for example as other health technologies become available for the same patient group. *Evaluation:* Optimal decision at the outset of the schemes may be more or less easy to define from one country to another, depending on the clarity of the decision rule for adopting new technologies (for example, it might be clearest in a country employing an explicit cost-effectiveness threshold). Whatever the decision rule, it is preferable that specific targets are specified, linking specific outcomes from the research with particular pricing and reimbursement decisions. *Other:* Several attempted schemes have failed because of lack of agreement between the relevant parties. Issues such as who will collect the data and who will analyse it need to be addressed. Attention needs to be paid to maintaining patient confidentiality and the independence of the analysis. In addition, there is the question of who pays for the study. The simpler the scheme and the greater the reliance on routinely available data, the cheaper it is likely to be. There are differences of opinion about the level of complexity of PBRSAs. While simpler schemes are likely to be cheaper to implement, more complex schemes make it harder to calculate the level of any implied price discount. This has some attraction for manufacturers who may be concerned about international price referencing.
35	Garrison et al. (100)	2015	USA	All technologies	*Desirability*: Risk-sharing agreements offer a number of potential advantages, including: 1) reducing the risk to payers of making a suboptimal purchase, 2) providing earlier access to medications for patients, 3) creating international pricing efficiency, especially in a world with external reference pricing and parallel trade; and 4) generating evidence on what works in the real world. The perceived value of a RSA versus a traditional discount mechanism depends on the product, disease area, and availability of the necessary data infrastructure. RSAs are typically not used in a US setting when a product is first to market or the market leader, but are more attractive to manufacturers when there is competition. The [interviewed] manufacturer seek to secure beneficial formulary placement or gain market share. They can also be a mechanism to increase patient compliance. [The interviewed] payers leverage RSAs as a way to reduce uncertainty about a product's clinical value, performance, or budget impact, as they allow payers and patients to gain experience with the medication, and [...] are

	interested in RSAs as they allow payers to ensure that the price of a drug is more closely aligned to its value. They indicated interest in RSAs for products that are more costly and for disease areas for which cost consequences are substantial. A key success factor [of RSAs] is the need for manufacturers to understand the amount of risk they are undertaking when entering into an RSA. Manufacturers feel they should be able to reasonably predict plausible outcomes of the agreement and assess the level of risk; for example, what level of compliance is required and to what extent the clinical trial population differs from the real-world population. Manufacturers are reluctant to take on risk when they cannot predict how their product will be used in the population. [Manufacturers and payers] shy away from agreements in disease areas where there are many different treatment paradigms or the relevant outcome is an intermediary outcome, because it can be challenging to attribute the outcome to the product in question. *Research design:* It is imperative that [payers and pharmaceutical companies] trust the data: clear agreements on data validation and analysis are important to create trust in the data. *Implementation:* Due to the difficulty in implementing and executing outcomes-based RSAs, interviewees indicated an interest in more financial-based RSAs (e.g. utilization or financial capitation) but less interest in clinical and health outcomes-based RSAs. Outcomes-based agreements [...] were perceived by interviewees to be difficult to execute and as having high transaction costs. Interviewees [...] cited challenges in implementing and executing outcomes-based RSAs [...], particularly given the fragmented payer system with patient movement across plans, as well as the current lack of data infrastructure that limits feasibility and, to some extent, interest in measuring long-term outcomes. Barriers to implementing RSAs that were identified during the interviews are: 1) Significant additional effort required to establish/execute RSAs (e.g. compared to traditional rebates/discounts), 2) Challenges in identifying/defining meaningful outcomes, 3) Challenges in measuring relevant real-world outcomes, 4) Data infrastructure inadequate for measuring/monitoring relevant outcomes, 5) Difficulty in reaching contractual agreement (e.g. on the selection of outcomes, patients, data collection methods), 6) Implications for federal (Medicaid) best price, 7) Payer concerns about adverse patient selection, 8) Fragmented multi-payer insurance market with and significant patient switching among plans, 9) Challenges in assessing risk upfront due to uncertainties in real-world performance, 10) Lack of control over how product will be used, 11) Significant resources and/or costs associated with ongoing adjudication. Additional challenges that were mentioned include a) the need for adequately trained staff, b) the risk to pharmaceutical companies associated with being responsible for outcomes when they cannot control the way a drug is prescribed or used, c) the management of consequences in terms of changing preferred drugs and denying coverage for drugs, and d) the identification of outcomes that are meaningful but measurable within a reasonable time frame. These interview themes were consistent with the top barriers identified in the survey.

					Top two barriers were the significant additional effort associated with the use of RSAs and the inadequate data infrastructure. Further, only large payers are likely to have the capability for multiple, simultaneous agreements, given the burden of negotiation, data collection, monitoring, and adjudication of RSAs. Another key consideration is duration, as payers and manufacturers agree that short-term deals are not desirable given the considerable investment in evidence development, [...] long-term deals are also not desirable given the costs and risks involved. Both payers and pharmaceutical companies acknowledge that medium term deals (18-36 months) are the right balance trading off the sizable up-front investment and a preference to execute agreements reasonably fast. Agreements that involve the manufacturer paying for non-pharmaceutical costs are difficult to execute. There is a risk of CED being perceived as a tool for monitoring or controlling doctors particularly with registers on interventional procedures with or without devices.
36	Lewis et al. (230)	2015	NS	Pharmaceuticals	*Desirability*: The criteria for instituting CED, such as high clinical need, and other criteria, such as timely return of research results, may also not be easily defined and may be contested. *Research design:* To make the use of observational and single-arm studies rather than RCTs studies worthwhile and valid, their design, conduct, and analysis needs to be rigorous and transparent, particularly with regard to minimizing confounding effects and bias needs. Because these studies also rely on real-world data sources, such as patient registries or administrative databases, and well-matched historical or contemporary comparative cohorts, they also must be supported by funding agencies, sponsoring companies, clinicians, administrators, and insurers. These studies should also complement, rather than replace, RCTs. *Other:* There are likely to be tensions between different and potentially competing interests. Companies may wish to obtain the highest prices possible, and payers may wish to pay as little as possible; manufacturers of generic medicines worry that these programs exist to promote early access to expensive patented medicines, and even manufacturers of patented medicines may worry that competitors will make use of the evidence they generate as part of a CED program. *Evaluation:* Pharmaceutical companies may disagree with payers about how to interpret emergent evidence—particularly whether it justifies maintenance, escalation, or reduction in drug prices. Failure to follow through on CED programs might also undermine the capacity for budgetary control if disinvestment or price reductions do not follow when medicines are found to be poor value for money. The reality is that once patients and clinicians have access to, and are familiar with, a new technology, disinvestment is highly unlikely to occur or is likely to be rigorously resisted by pharmaceutical companies. *Other:* A further challenge for CED schemes relates to ethical oversight of these programs: for example, about what kinds of CED programs count as research (e.g., whether participation in registries counts as research or simply as quality improvement) and what, if any, consent and ethical

					oversight are needed. A related question is whether it is ethical to insist that patients participate in research as a condition of coverage. Although clinical trials often provide the only means of accessing new therapies, CED programs are generally predicated on the basis that the new therapy is likely to be beneficial—meaning that there is unlikely to be genuine equipoise. Negotiations about CED programs— particularly discussions of drug prices—may be commercially confidential. Given the centrality of price negotiations to CED programs, commercial confidentiality makes it impossible for funding organizations to be transparent in their decision-making processes. This, in turn, may undermine their inclusiveness, their perceived trustworthiness, and their legitimacy in the eyes of external stakeholders to thwart efforts on the part of funding agencies to satisfy the principles of accountability for reasonableness.
37	Mohseninejad et al. (231)	2015	Netherlands	Pharmaceuticals	*Evaluation:* Like any observational study, registry data are inevitably biased. Although solutions for this exist, for the real-world data used in the present study, patient heterogeneity turned out to be too large to allow for appropriate correction of confounding in the registry data. This resulted in problems in estimating incremental cost-effectiveness using the registry data only
38	Navarria et al. (232)	2015	Italy	Pharmaceuticals	*Implementation:* A very low percentage of reimbursement procedures under PBRSAs schemes has been activated. Reasons accounting for such a high percentage of unrequested reimbursements may be found, at least in part, not only in the high percentage of patients still under treatment and in interruptions of treatment for reasons other than the ones provided in the negotiation agreement but also in patients' files that have not been closed because of the healthcare centre inefficiencies, thus preventing the activation of the reimbursement procedure. Considering the high number of registries and the heterogeneity of the mechanisms through which regions/companies/ pharmacies/hospitals are actually refunded, the complexity of this system is remarkable. Moreover, healthcare professionals are not prompted to update the registries, close the patients' files, and submit refund requests on a regular basis, possibly because the actual money to be refunded does not come to the prescribing centre itself, but rather to the hospital general budget, thus producing a responsibility gap between the stakeholder that will receive the refund (in this case, the hospital) and the actual person in charge of the reimbursement procedures (the prescribing centre).
39	Thompson et al. (233)	2016	Canada	Pharmaceuticals	*Desirability*: Unwillingness to take on risk and lack of trust among stakeholders. IPLAs can sometimes increase budgetary uncertainty. That is, the number of patients who will require coverage (e.g. based on response criteria) may remain unclear, and/or determination of a product's financial impact may be prolonged outside the timeframe of the payer's regular budget cycle. *Research design:* A clear understanding of all clinical and economic uncertainty should be obtained before IPLA development. Almost unanimously, the participants stated that outcome-based IPLAs should include agreed-upon outcome measures that are quantitative, open to little interpretation, and suitable for collection within a realistic timeframe (generally no more than three years). Difficulty

					achieving agreement on suitable clinical outcomes to measure. Siloed nature of policies and capabilities across provinces/drug funding practices; budgets within and across sectors and institutions within the health service; manufacturer budgets. The participants indicated that all stakeholders who ultimately hold budgetary and/or administrative responsibilities related to IPLAs should be involved in their design, implementation, and management. Many participants stressed the importance of involving unbiased third parties in design, data collection, analysis, and interpretation to ensure study robustness/validity in agreements that are seeking to reduce uncertainty about clinical outcomes. The survey participants offered a variety of opinions when asked about the funding of IPLAs. Some suggested that industry should be required to pay for additional real-world evaluations, as certain provinces (e.g., smaller ones) may not be able to fund them. Others indicated that payers – the ultimate "buyers" – should finance the cost to develop and implement the agreements, as the outcomes of the IPLA may offer benefit or reduce costs for the entire healthcare system. Some participants believed that funding responsibilities should be shared. *Implementation:* High administrative burden, resource demands, and costs to execute and manage the agreements. Difficulty adapting to or changing existing healthcare delivery practices (e.g., patient monitoring and follow-up). Limited data collection systems within and across provinces; for example, across-province variability in coding and collecting outcomes and lack of cross-country connectivity to share and compare data. *Other:* The participants noted that consultation and collaboration need to occur among all relevant stakeholders beginning in the design stage and continuing throughout the life of the IPLA.
40	Clopes et al. (234)	2017	Spain	Pharmaceuticals	*Desirability*: The professionals involved were able to visualize the model properly because the PbR stipulated that access to gefitinib must be offered in a manner considered correct by all. Both the institution and the professionals acknowledged that the benefits of the experience included not only the potential gains in cost effectiveness but also the additional factors derived from a more appropriate use of the product. From the perspective of AstraZeneca, the PbR allowed the product to be introduced in a specific market under appropriate and agreed conditions of use. Under other conditions, this might have been delayed or hindered *Research design:* As the PbR was linked to only one service provider, there was a risk that the negotiations and conditions would be fragmented, which could have meant a major effort for the pharmaceutical company to gain market access. However, the involvement of the health authority demonstrates their openness to adopt a generalized perspective, resulting in a unique negotiating process for Catalonia, which could lead to the inclusion of other hospitals in such schemes. *Other:* One element included in other settings is the involvement of patients in these types of agreement. This aspect was taken into consideration by the Catalan Institute of Oncology but was

					ruled out, since the use of the drug and the outcome evaluation model specified in the PbR were essentially the same as in standard clinical practice. *Implementation:* From an operational and organizational perspective, the availability of adequate information technology to measure outcomes and monitor accountability were acknowledged as essential or critical to the development of dynamic payment models [in the interviews to health authorities, clinicians and manufacturers]. The availability of diagnostic resources capable of complying with the legal framework and the agreements made was also considered essential. Governance: In organizational terms, four different aspects had the greatest importance, the first of which was leadership. This was understood as a managerial role embodied in a defined strategy for the centre, in which PbRs are a consistent feature, forming the framework of the relationship with the pharmaceutical industry. Second was the involvement of professionals in defining the PbR in order to guarantee adherence; third, funding; and fourth, a suitably sized, appropriately prepared pharmacy department. *Evaluation:* The model was, in general, appropriate to the needs of the parties involved. The involvement of the healthcare authority and the use of clinical guidelines drawn up by healthcare professionals should be highlighted.
41	Gerkens et al. (235)	2017	EU	Pharmaceuticals	*Desirability*: For the manufacturers, there is a problem of "free riding": despite the confidential nature of MEAs: some competitors can benefit from the data or information gathered by the manufacturer engaged in the MEA (e.g. if their data gathering opens the door for reimbursement. In addition, the introduction of uncertainty for manufacturers regarding a payoff for the additional research produced, and the potential impact that the new evidence could have on future prices or revenues could, in turn, dis-incentivise additional data collection after a MEA is in place. For regulators and payers there is a risk for potential disinvestment in certain disease areas, for example, those where a very limited target population is likely to be identified in the context of a MEA, which could translate into low volume use for a product. The potential benefit linked to capturing more information that could improve the cost-effectiveness and use pattern of a product, should always be confronted with their complexity and the high costs often linked to following such a route (administrative and others). The necessary resources need to be set in place to ensure an appropriate follow-up and periodic evaluation to avoid insufficient or late reporting of results. There are ethical challenges on MEAs that are often neglected for example, about the types of CED programs which count as research (e.g., whether participation in registries counts as research) and what, if any, consent and approval by ethics committee are needed. *Research design:* Challenges linked to their and transferability of results from one country to another, in particular in the case of CED-MEAs. Different standards of practice, resources used, settings, or costs make it difficult to draw clear conclusions for one country based on the additional evidence or

	information captured in another one via a MEA. The confidential nature of the data captured ads to the difficulty. If these schemes become very common, there is a risk that manufacturers may systematically ask for higher departing prices in expectation of a MEA. Also, CED lacks a governance structure, or a systematic approach, as a consequence, schemes are thought to be easily manipulated post implementation. One of the main reasons for the lack of governance comes from an absence of clear criteria to first decide whether a CED is required and then, once this is ongoing, arrive at a positive or negative decision for reimbursement based on the new evidence provided. Furthermore, there is also a lack of standardised criteria on when and how to link decisions to specific additional requirements (e.g. restricted to specified providers or the need to develop a registry). In addition, there is a lack of clarity on the role of different stakeholders. This include conflict of interests when industry is responsible for the funding and/or design of a registry or observational study as well as for the analyses of the results and their publication. There are challenges linked to data collection outside of an "ideal" RCT context: to make observational studies worthwhile and valid, their design, conduct, and analysis need to be rigorous and transparent, particularly with regard to minimizing confounding effects and bias. Because these studies rely on real-world data sources, such as patient registries or administrative databases, and well-matched historical or contemporary comparative cohorts, they also must be supported by funding agencies, sponsoring companies, clinicians, administrators, and insurers. These studies should complement, rather than replace, the evidence captured by RCTs during the drug development process. Specific challenges are linked to outcome measurements. Outcomes need to be objective, clearly defined and measurable within the time frame of MEAs, which is usually limited. Some of these MEAs make use of intermediate clinical outcomes and do not foresee a long-enough follow-up to truly assess the relevant final outcomes. Under such circumstances, the intermediate outcomes should be validated in order to ensure the financial and time consumption efforts incurred in are not wasted. There are also important hurdles directly linked to the development of effective registries, with general agreement that, in order to successfully undertake CEDs, it is necessary to have better access to data and to be able to link databases. Standardising data elements across registries would also be important. Registries are identified as one of the most commonly used tools to collect data under MEAs. However, finding a sustainable funding model for registries, verifying their accuracy, reliability and completeness of their information, as well as, ensuring the confidentiality of the information they include is respected, are some of the issues linked to the development of registries that should not be overlooked. Investment in high quality information systems is therefore thought to be needed in order to improve the current situation. *Implementation:* For regulators and public payers, 1) there are also high transactions and administrative costs (particularly for CED schemes), 2) temporary reimbursement via a MEA could discourage

					manufacturers from capturing additional data. Also, 3) a further challenge is the decision on how much evidence may be enough and when a CED should be stopped. The period between evidence generation and a final decision varies considerably from one MEA to another and is sometimes thought to be too long, while in other instances it may not be enough (depending on the indication or outcomes needing to be captured). Often there are doubts throughout the MEA period on what "corrections" may be required, including whether a cancelation of the MEA would be appropriate, 4) for performance-linked agreements, difficulties in obtaining refunds from manufacturers have been reported in the literature. These appear to be primarily linked to late requests or disagreements with pharmaceutical companies. *Evaluation:* For payers: 1) There are difficulties to de-list a drug from reimbursement, once a MEA is established. This happens because patients and clinicians may be unwilling to see the product de-listed. Also political pressure play a role, for example depending on the disease area considered, e.g. chronic conditions, or products that already achieved wide use and acceptance. Also if investment in capital equipment or in training had to be incurred, providers may also become more reluctant to stop using a treatment intervention and, consequently, de-listing will become more difficult, 2) Given the confidentiality of these agreements it is hard to even identify any attempt to withdraw or limit reimbursement. In addition 4) decisions to enter into these schemes s are often heavily influenced by direct and indirect pressure from the pharmaceutical industry, patient advocacy groups, physicians, and patients and their families. Establishing one of these schemes may also be a convenient way for politicians to postpone a difficult confrontation with patient advocacy groups over funding decisions Once more, the confidentiality surrounding these MEAs makes the identification of political weights in decision-making a challenging task. Without access to the confidential information, it is not possible to determine the extent to which decisions were based on evidence or on other grounds. There is a need for increasing transparency in general and encouraging publication of study results/registries to facilitate a more efficient decision making.
42	Kanavos et al. (101)	2017	NS	All technologies	*Desirability*: There is a group of concerns relating to the introduction of additional uncertainty for manufacturers in terms of expected returns, which may have the opposite effect of dis-incentivising additional data collection, the advantage competitors may take of data collected by the manufacturer, and related to this is the problem of free-riding. *Evaluation:* There is still relatively little evidence to support the claimed benefits of MEAs and the extent to which some of the challenges involved in MEA implementation […] impact on the final outcome. Also, despite MEAs collect very useful data, which would enable the drug to be re-assessed and its price re-negotiated, few countries appear to leverage this opportunity. With regard to schemes aiming to manage utilisation to optimise performance, it is not clear if they really succeed in limiting reimbursement to specific patient sub-groups.

					Other: There are a number of open questions, which need to be addressed such as who should finance data collection, who should be responsible for it, and how to streamline implementation of MEAs so as to reduce the management burden for healthcare staff. The complexities of MEAs may have implications for different stakeholders. For payers, additional efforts are required to make a new drug available to patients such as negotiation time, monitoring of patient response, data collection, or development of registries, among others. At the same time, there is limited capacity to implement and assess evidence – especially if clinical evidence needs to be assessed – if implementation takes place at regional or hospital level. For patients, there are, generally, limited opportunities to engage with the development of MEAs and not all patient groups are aware of what MEAs do. Manufacturers are in the majority of cases required to make concessions such as refund for non-respondent patients or collect additional data to support reimbursement negotiations. Also, there is variability in the perception of MEAs across countries and what actually is a MEA may differ across settings; this may create confusion to different stakeholders. And there is a frequent lack of transparency on the agreements implemented, their objectives, and evaluation of their impact, which is preventing cross country learning and is severely limiting the ability of patients to engage with MEA processes. There are challenges linked to the regulation of these agreements and the transferability of results from one country to another.
43	Nazareth et al. (236)	2017	USA, UK, Italy, Spain, France, Germany	Pharmaceuticals	*Desirability*: Payers generally perceived that the willingness of manufacturers to pursue an OBC was a positive sign for the value proposition of the product, while some manufacturers perceived OBCs as particularly valuable negotiation tools to demonstrate the value of their products. The value for payers of an OBC was perceived to be driven largely by economic considerations, such as reducing costs, risk, or uncertainty. For manufacturers, the main value drivers were identified as potential access or reimbursement gains. *Research design:* OBC implementation has not been without challenges. Most notably, the absence of an effective guiding framework to generate, implement, coordinate, and monitor OBCs was noted by many interviewed stakeholders as an important constraint to OBC activity. By framework, participants generally indicated an approach to OBCs that establishes key parameters or building blocks for payers and manufacturers to negotiate OBCs. OBC implementation has historically been hindered by a series of factors, including the availability, consistency, and reliability of patient data; selection of appropriate conditioning outcomes; and alignment of data needs and expectations between partners. Early and sustained involvement of health economics and outcomes research in the OBC design and implementation stages was considered important by several manufacturers and payers. *Implementation:* Feasibility concerns included limited or inappropriate data infrastructure (e.g., lack of relevant/appropriate patient outcomes data); high negotiation complexity (e.g., negotiation may

					involve numerous stakeholders on the payer and the manufacturer sides, with potential multiple rounds of negotiation in order to finalize the contract); and high internal organizational costs (e.g., as multiple internal stakeholders and resources may be required). *Other:* Lack of transparency regarding OBC activity and results has created ambiguity about whether OBCs have been a passing trend or are a rising reality and has prevented knowledge of past OBCs from being used by other manufacturers and payers.
44	Rothery et al. (60)	2017	UK	MDs	*Desirability*: Establishing the clinical effectiveness and cost effectiveness of MDs relies on evidence which is often less extensive and lower in quantity than evidence for many pharmaceutical products. This is largely because the evidence requirements for MDs to achieve a CE mark is less demanding. Unlike pharmaceuticals, where evidence on efficacy and safety is legally required before marketing authorisation is obtained, devices usually only need to demonstrate performance and safety, with the CE mark acquired close to the point of market entry. Uncertainty about the efficacy of the device and the learning or training required to achieve the desired efficacy can result in adverse consequences on patient outcomes and lead to an ineffective use of healthcare resources. Rapid approval of new entrants can also result in disincentive effects for manufacturers to invest in further research which would reduce these uncertainties. The unique characteristics associated with MDs such as rapid incremental innovation, learning effects and upfront irrecoverable costs, all present a challenge for the timing of reimbursement decisions and the value of waiting until additional evidence is conducted to support the technology. This means that conditional coverage decisions and possible risk sharing schemes (between the manufacturer and health sector) become even more important. One of the complexities associated with the evaluation of MDs is the fact that any decision about the adoption of the device will also interact with the ability to gather more evidence and may affect future commercial developments of the technology. There is also a close link between the value of the device, the value of further research to reduce uncertainty and the price of the device. These links can offer incentives for manufacturers to price accordingly and decide whether there is sufficient value from further evaluative research. It also helps to establish how the value of the device and the value of future research might be shared between the manufacturer and the health sector in order to inform who might reasonably be expected to pay for (conduct) the research. Manufacturers also need an approach to make rapid decisions at the start of product development and to revisit the decision to continue development and research at different points in the development cycle. One example is the 'Headroom Approach' which has been discussed in the context of MDs. The value of further research can be informed through methods of value of information analysis, which can also be used to inform the type and design of proposed research. Some assessment of the likelihood that research is conducted, the length of time for research to report and the costs of conducting research are also required.

					Evaluation: Further research is unlikely to be able to resolve all uncertainty. Some sources of uncertainty that cannot be reduced by further research may resolve by other changes occurring over time. For example, the effective price of the technology and/or its comparators may change in the future. The price clearly plays a key role in determining the value of the technology, but it also affects the level of uncertainty by changing the likelihood of making an incorrect decision and the value of further research. The information generated by research will not be valuable indefinitely as new and more effective interventions may become available and make the information no longer relevant to future clinical practice. Therefore, new or incremental innovation will also change the value of a technology and the future value of research. Each decision option is based on the balance of evidence supporting the value of the technology, the value of additional evidence, future changes and the likelihood of research being conducted. OIR restricts access until further research establishes value; and AWR may result in subsequent withdrawal of the technology when further research is completed. Investment costs that are sunk costs cannot be recovered if a decision about the technology is changed in the future. Funded interventions, allocation of costs pro-rata has no influence. However, if a decision to end adoption changes before the end of the lifetime of the technology, these costs cannot be recovered. Therefore, any significant irrecoverable costs should be considered when assessing the value of a technology.
45	van de Wetering et al. (237)	2017	Netherlands	All technologies	*Evaluation:* Once a technology like a pharmaceutical is used in practice, ending reimbursement maybe less feasible than deciding not to reimburse in the first place, in particular when the technology was proven to be effective in practice, but not cost-effective. This relates to the general tendency to value equally sized gains and losses differently, with losses looming larger than gains. This phenomenon of loss aversion is a well-known aspect of prospect theory. In the context of allocation decisions, loss aversion may imply that policy makers may be willing to accept higher cost-per-quality adjusted life-year (QALY) ratios for technologies already reimbursed (under conditional reimbursement) than they would accept for technologies not yet reimbursed (in the conventional decision-making context) The results of the DCE showed that both the general public and the policy makers preferred to reimburse an existing technology, which is already reimbursed and used in practice, over reimbursing a new technology, which is presently not reimbursed or used. This finding may have important implications for the consideration to use conditional reimbursement as a policy instrument for allocation decisions.
46	Bouvy et al. (238)	2018	EU	Pharmaceuticals	*Desirability*: In the framework of adaptive pathways (conditionally reimbursed) products, [interviewed] manufacturers seem more interested to explore the use of outcomes-based arrangements than payers. However, there is low appetite among European payers and HTA bodies for using agreements that involve collecting outcomes data due to their complexity.

					Research design: Coordinating multi-country data collection efforts, rather than seeking country-specific models could be facilitated through early dialogue and might facilitate outcomes-based arrangements. If a set of countries could agree on the same data to be collected within their healthcare systems once an adaptive pathways product enters the market this could substantially improve the timeliness, cost, and efficiency of data collection efforts as accrual will be faster […]. A flexible pricing model with variable discounts might be acceptable for both payers and manufacturers. Under such a scenario, payer and manufacturer would agree on a list price and conditions under which a discount would be modulated as pre-set outcomes would be met […]. Simple models rather than complicated ones: simple data collection efforts involving the collection of few but essential parameters that are normally tracked by the healthcare system might be easier to implement, less costly, and more feasible for cross-country coordination. *Implementation:* HTA bodies and payers expressed reluctance regarding outcomes-based agreements because of the added complexity and lack of administrative infrastructure readily available in most countries that would facilitate their implementation. The complexity of outcomes-based MEAs could be mitigated by existing data infrastructure to avoid lengthy study set-ups, and the use of simple outcome measures that are easier to track in clinical practice. Properly incentivizing stakeholders (especially prescribers) to participate in data collection could improve data quality and follow-up. *Other:* Apart from technical, healthcare system, and political factors, a lack of trust between payers and manufacturers might be one of the key hurdles to more extensive use of outcomes-based arrangements. Payers and HTA bodies reported that they consistently see very high prices for new products without much differentiation according to added value, whereas manufacturers reported that payers seem more concerned with budget impact and are unwilling to consider more complicated arrangements. In the absence of a comprehensive national data collection system, the use of a mutually trusted third party for outcome measurement could facilitate trust.
47	EXPH (239)	2018	EU	Pharmaceuticals	*Desirability*: The several forms and variants of these (MEA) agreements deal with different aspects, such as e.g. hidden price discounts (of value to companies as such discounts bypass international referencing practices used in many health systems), uncertainty about the performance of the product in real-world context, asymmetric information about product (potential) performance between companies and healthcare payers. The use of MEAs provides a way to have early introduction of new products "managing" the information flow. The basic issue addressed is typically related to evidence required to take final decisions, later on when more information has become available. This means that MEAs are not designed to address the issues of high prices as a result of exercise of market power by pharmaceutical companies. The managed entry agreements focus mostly on ways to deal with the uncertainty, neglecting the split of value between payer and company may lead to very unbalanced division. Some of the MEAs will address the issue of price but based on differences of

value to society and not in terms of explicit margins discussion. MEAs are administratively complex and may be difficult to negotiate and their effectiveness has yet to be evaluated. Moreover, they mainly are designed to address the issue of uncertainty about the value of the effectiveness of the medicine and not the (high) price tag or the rising pharmaceutical expenditure, although well-designed MEAs can help on price issues (especially if explicitly addressed).
Research design: There are two main issues: how to deal with uncertainty about the value of the new product and how to set its price. The great majority of discussions have the focus on the first problem. The question is whether, or not, such high prices are really the result of well-functioning system of rewards to innovation. MEAs [are] not without problems and, depending on the exact comparator situation, they may even introduce inefficiencies. One example is the moral hazard effect of the so-called risk-sharing agreements. Whenever a payment occurs only if successful treatment is achieved, decision makers in the health system may have an incentive to put too many patients into treatment as treatment failure will not have a direct financial cost to them. As the financial cost of failures passes through to prices of successful treatments by companies, health systems may end up with too many patients under treatment under a higher price, driving up healthcare expenditures (240). To companies, MEAs offer the additional benefit of setting confidential effective prices, breaking the link of external reference pricing (a policy that relies on publicly available listed prices of pharmaceutical prices in reference countries). The confidentiality of prices brings countries to a situation that is usually termed prisoner's dilemma. Individually, it is optimal to sign agreements of prices that are confidential, while globally countries could be better off by keeping a coordinated action on price determination for pharmaceuticals. MEAs may bring disadvantages, with the following ones being listed in the existing literature: MEAs 1) provide access to medicines with uncertain clinical benefit and—at a later stage—it is difficult to argue against patients why they are not reimbursed anymore (dynamic consistency problem) (see Van de Wetering et al. (237) for a discussion), 2) are associated with additional costs for implementation, especially when they are based on the clinical outcome data, 3) require clear decision rules on when to stop reimbursement or adjust use or pricing, 4) require well-functioning IT support, and 5) undermine the current system of international price comparison, since MEAs usually contain confidential agreements on discounts, while EPR is only referenced to list prices, since the discounted confidential prices are not known. As a result of the confidential agreements, the payers can claim to have completed a good deal, although there is no objective evidence on the basis of comparisons due to lack of comparative data from the other countries. There is an element of exercise of market power present in the high prices asked [by pharmaceutical companies] that is not addressed by MEAs by design. It is becoming clear that heterogeneity in MEAs, across and within countries health systems, makes it difficult to have an integrated approach at the healthcare payer level. In addition, both price setting and data collection

	(evidence) by companies may adjust to the conditions required by the MEAs. Quick examples are upward price adjustments by companies under the expectation that discounts will be part of the MEA and leaving data (evidence) collection to later stages, within the context of the MEA. That is, the starting points of the initial MEAs may not be representative of future MEAs, as economic agents adjust to their existence. On the side of pharmaceutical companies, as healthcare payers require further information and monitoring systems, costs of engaging in MEAs can escalate. Market entry agreements based on outcomes have strong demands in terms of data collection and its interpretation, making it difficult to work in every case. *Implementation:* Only some countries will have the ability to manage MEAs, and oversee the results. Replication in every country will be challenging for small countries due to costs of setting and using monitoring mechanisms. There are clear economies of scale in the management of entry agreements for new pharmaceutical products, even though it may require some uniformity or harmonization of objectives, thresholds, and methodology. There is a major governance risk that a product introduced initially based on expected valuation, later, based on its performance in a population context, is found to be a low value product, but it is not excluded from coverage due to (potential or actual) public opinion or patients' pressure. Due to a governance failure, the MEA will not work as intended in this particular type of situation. Such failure is not merely theoretical. Allocation decisions in healthcare, especially negative decisions, can be controversial, both politically and societally, which is an important factor in the current discussion. MEAs should not become a quick-fix solution to introduce expensive medicines but be integrated into a process of managed introduction of new medicines which starts from horizon scanning activities, moves to forecasting, HTA assessment, pricing and reimbursement, and continues with post-marketing studies and surveillance. Thus, MEAs are a useful tool in a more global process of setting payment models for innovative pharmaceutical products, and are complemented by other instruments available (as discussed later in this Opinion). The anticipation of entry decreases one type of problem, delayed access—a good new product reaches sooner the patients. As elements such efficacy and safety are monitored along the way, a different problem emerges—the use of products that have a performance that under normal conditions would not lead them to be approved. As withdrawal may be difficult, as would be seen as cutting access to a product by the population, unless serious issues of safety become apparent, the anticipation substitutes one type of problem by another (237). The general use of more complex payment models [such as MEAs] for new pharmaceutical products will imply changes in health system design. Some of the changes will likely create challenges in terms of political feasibility, including the delisting of products that do not materialize initial expectations based on preliminary evidence. Even if predicted in the payment model, removing products from coverage may face the opposition of patients, even in the case of smaller effects than promised.

					Evaluation: Although MEAs have been applied in many countries for several years, there is no public knowledge available whether they meet the associated expectations (a contribution to the reduction of uncertainty on actual benefit, amount of cost reductions and/or access of patients to these medicines). The ability of MEAs to bring useful information in practice seems to fall short of expectations. Aspects that seem to contribute to this finding are the short time span of the use of MEAs, the small number of patients typically involve, and selection of patients receiving the pharmaceutical product (after being approved) included in the MEA. The discussion [...] does not address the crucial issue of price determination mechanisms, the non-disclosure conditions on the exact terms and results of MEAs, part of the agreements set, lead to lack of transparency and difficulties in assessing whether or not objectives are achieved. [Regarding MEAs the] broad message [is] centred in the complexities and heterogeneity of MEAs bringing less information and higher management costs that were presumably predicted.
48	Makady et al. (241)	2019	Netherlands	Pharmaceuticals	*Desirability*: Inclusion of drugs in CF was only warranted if three criteria were met: a budget impact above 2.5 million euros/year, a proven additional therapeutic value in comparison to available comparator treatments, and a well-defined proposal for out-comes research to address uncertainties regarding appropriate use and cost-effectiveness in routine practice. *Research design:* Literature alludes to numerous reasons such as challenges with analysing and interpreting RWE generated. It may be that the lack of full incorporation of recommendations on the proposed outcomes research contributes to this. Two safeguards proposed in ZIN guidelines may have prevented such shortcomings in hindsight. Firstly, the conduct of VOI analyses at T=0 to highlight the feasibility and intrinsic value of data collection for specific parameters within the timelines projected. Secondly the mid-term reporting of outcomes research progress and interim results between T=0 and T=4 (specifically at T=1 & T=3) may have led to more timely decisions regarding continuation, adjustment or termination of the CF procedure for drugs, thereby avoiding waste of valuable time and money for all stakeholders involved. Unfortunately, both recommendations (VOI and interim reporting) were published in December 2008, more than two years after the start date of the CF scheme. Another shortcoming is the absence of an a priori strategy for the implementation of CF outputs in the actual healthcare setting. Provided the diversity of stakeholders active within the Dutch healthcare setting, the complexity of interactions between their mandates and stakeholders' differing interests, more attention should have been paid to establishing a priori strategies on how CF outputs would and should be implemented in practice by different stakeholders. *Implementation:* Subsequent [to inclusion of a drug in CF] marketing authorisation holders, in collaboration with hospitals, clinicians and clinician societies would implement the proposed outcomes research to collect RWE on appropriate use and cost-effectiveness in routine practice

throughout a period of three years, which was eventually extended to four years. Only one drug [of the 12 drugs included in the review] was completed within the envisioned four-year period. Although reasons for failure to timely processing of the [...] drugs are not directly apparent from the extracted data for this study, they may relate to a myriad of factors, including the time needed to set up registries required for data collection, to compile and evaluate data generated from outcome studies, and subsequently to assess and appraise the evidence generated. The use of tailored approaches for determining required time-frames to answer the questions raised at T=0, rather than a fixed four-year window, may provide a more intuitive design.
Evaluation: After the four-year period, ZIN would conduct a reassessment (T=4) of drugs comprising the following elements: therapeutic value, appropriate use, cost-effectiveness and budget impact. Finally, an appraisal of all available evidence at T=4 would be performed to advise on the reimbursement of drugs based on4 criteria: necessity, clinical effectiveness, cost-effectiveness and implementability within the healthcare system. The Scientific Advisory Committee (WAR; hereafter Assessment Committee) was responsible for the assessment of evidence at T=0 and T=4. Meanwhile, the appraisal of evidence at T=4 based on the four criteria was conducted by the Appraisal Committee. No systematic evaluation of CF in the Netherlands has been conducted since its inclusion (i.e. implementation) stopped in 2012. For the 13 drugs that are included in the review, reassessments are ongoing (5/13) or pending (e.g. due to extended deadlines allowing for extra data collection to supplement inadequate datasets; 8/13). For 11/12 finalised drugs, the elapsed time period between publication of the T=0 and T=4 reports extended beyond 4 years. Contrary to ZIN guidelines, the Assessment Committee (rather than the Appraisal Committee) performed the appraisal of evidence at T=4 in relation to the four package (i.e. decision) criteria for 5/12 drugs. This occurred for drugs whereby appraisal was relatively straightforward (i.e. evidence at T=4 on all four package criteria indicated a positive opinion on continued reimbursement). However, for the remaining 7/12 drugs where evidence may have led to a negative advice, the Appraisal Committee was consulted. The Assessment Committee concluded that evidence submitted at reassessment (T=4) and its analysis was of sufficient scientific quality to assess AU in Dutch clinical practice for 9/12 (75%) of finalised drugs and inadequate for 3/12 (25%) of drugs. Meanwhile, the committee concluded that evidence submitted at reassessment and its analysis was of sufficient scientific quality to assess CE in Dutch clinical practice for 7/12 (58%) of finalized drugs and inadequate for 5/12 (42%) of drugs. The committee stated that it could not reach conclusions on appropriate use for drugs for which the submitted evidence was insufficient. The Assessment Committee went on to appraise all evidence at T=4 in relation to the 4 package criteria (necessity, clinical effectiveness, cost-effectiveness and implementability in the healthcare system) for 5/12 drugs; for 4/5 drugs, continued reimbursement from the basic healthcare package was advised. For the final drug (natalizumab), the

					committee advised to postpone the decision on discontinuation of reimbursement until further evidence becomes available from a separate initiative (Round Table on Multiple Sclerosis). The Appraisal Committee appraised evidence at T=4 for 7/12 drugs. For 5/7 drugs, continued reimbursement was advised based on additional conditions, e.g. the need for exceptional financing of orphan drugs outside the basic healthcare package, tailored policies on development and pricing of orphan drugs, the establishment of necessary patient registries to monitor real-world outcomes and bundling of clinical expertise to ensure AU. Conditions for expensive drugs varied per case. For 2/7 drugs (eculizumab and ranibizumab), the Appraisal Committee advised [the Ministry of Health] to discontinue reimbursement. The Assessment Committee concluded that evidence generated through outcomes research was of insufficient quality to answer a third of research questions defined at T=0. For a third of research questions defined at T=0, insufficient evidence was generated through the implemented outcome research studies to reach grounded conclusions at T=4. Moreover, for half of the finalized drugs, reimbursement was continued based on yet further evidence generation to address remaining uncertainties.
49	Pouwels et al. (242)	2019	Netherlands	Pharmaceuticals	*Research design:* Uncertainties were neither systematically nor completely identified in the analysed CED schemes. *Evaluation:* Assessing the success of CED schemes has to be based on process indicators because quantitatively assessing the value of CED research, as a result of the reduction in decision uncertainty, is difficult. [Also] because stakeholders may have different perspectives on the success of these schemes. In the Netherlands, CED was useful in producing evidence on the use of treatments in daily clinical practice but it did not always decrease the uncertainty surrounding the cost-effectiveness of these treatments. Hence, policy makers should carefully assess whether additional evidence collection will significantly reduce decision uncertainty before engaging in CED. Stakeholders emphasized that CED research design should be adapted to the identified uncertainties and questioned whether research had reduced uncertainty.

AIFA, Agenzia Italiana del Farmaco (Italian Medicines Agency); CE, Conformité Européene; CED, coverage with evidence development; CF, conditional funding; CFFE, conditionally funded field evaluation; CMS, Centers for Medicare and Medicaid Services; CNEDIMTS, Commission nationale d'évaluation des dispositifs médicaux et des technologies de santé (National Committee of Medical Devices and Health Technologies); EDSS, Expanded Disability Status Scale; EXPH, Expert Panel on effective ways of investing in Health; FDA, Food and Drug administration; FDG, fluorodeoxyglucose; FED, Funding contingent upon Evidence Development; GRP, good research practice; GSAC, Gastroenterological Society of Australia; INHS, Italian National Health Service; IPLA, innovative product listing agreements; MD, medical device; MEA, managed entry agreement; MSAC, Medical Services Advisory Committee; MS, multiple sclerosis; MSRSS, Multiple Sclerosis Risk-Sharing Scheme; NCD, national coverage

decisions; NHE, net health effect; NHS, National Health Service; NOPR, National Oncologic positron emission tomography (PET) Registry; OBC, outcome-based contract; PBAC, Pharmaceuticals Benefits Advisory Committee; PbR, payment by result; PBRSA, performance-based risk sharing agreement; P4P, pay for performance; UK, United Kingdom; VOI, value of information; RCT, randomised controlled trial; RET, Research Ethics Committee; RS, risk sharing; RWE, real-world evidence; ZIN, Zorginstituut Nederland (Dutch Health Care Institute); [a] *All technologies include MDs.*

Supplementary material S3.1

Table S2 Characteristics of the patients from the study on Electrochemotherapy

Gender	Age	Previous local/regional treatments	Previous systemic treatment	Systemic treatment at ECT	Location of nodules	No of ECT session	Number of nodules	Route and drug for ECT	Used electrodes
M	80	surgery, RT	No	No	armpit	2	5/5*	i.v. bleomycin	H
F	73	surgery	No	No	cheek	1	4	i.v. bleomycin	P
F	94	surgery, RT	No	No	foot	1	1	i.v. bleomycin	H
F	79	surgery, ILP	No	No	calf and tight	1	30	i.v. bleomycin	P
F	84	surgery	No	No	scalp	1	6	i.v. bleomycin	H
F	96	surgery	No	No	calf	1	34	i.v. bleomycin	H
M	62	surgery	No	No	neck	1	14	i.v. bleomycin	H + N
F	88	surgery	No	No	calf	1	1	i.t. bleomycin	P
F	83	surgery	No	No	calf and tight	3	20/19 (2*)/25(3*)	i.v. bleomycin	H/H + P/H
F	60	surgery, RT	adjuvant interferon-α	No	calf	1	18	i.t. cisplatin	P
F	82	surgery, ILP	No	No	tight	1	115	i.v. bleomycin	P
F	82	surgery	No	No	calf	1	12	i.v. bleomycin	P + N
F	79	surgery	No	No	calf	1	6	i.v. bleomycin	H/N
F	79	surgery, RT	No	No	calf	1	36	i.v. bleomycin	P
F	84	surgery, RT, ILP	No	No	calf and tight	5	18/23/27/41/20	i.v. bleomycin	P/P/HG + P/P/N + P
M	58	surgery	No	No	tight	1	10	i.v. bleomycin	P

Gender	Age	Previous local/regional treatments	Previous systemic treatment	Systemic treatment at ECT	Location of nodules	No of ECT session	Number of nodules	Route and drug for ECT	Used electrodes
F	48	surgery	pembrolizumab	pembrolizumab**	breast	1	1	i.v. bleomycin	H
M	86	surgery	No	No	shoulder	2	1	i.t. cisplatin	P
M	93	surgery, RT	No	No	cheek	1	1	i.t. bleomycin	N
M	75	surgery, RT	No	No	cheek	1	1	i.v. bleomycin	P
F	63	surgery, RT, ILP	pembrolizumab, dacarbazine	dacarbazine**	tight	2	13/3*	i.v. bleomycin	H
F	84	surgery	No	No	calf and tight	1	107	i.v. bleomycin	P
M	86	surgery, RT	No	No	calf	1	9	i.v. bleomycin	H

*ECT-electrochemotherapy, RT – radiotherapy, ILP – isolated limb perfusion, * same nodules as in previous session were treated, i.v. – intravenous, i.t. – intratumoral, ** therapy not working on skin lesions, H – hexagonal electrode, P – plate electrode, N – needle row electrode.*

Supplementary material S4.1

OpenBugs code sed to estimate transition rates and probabilities from fully observed primary data on patients in the Electrochemoterapy model presented in Chapter 4.

```
model{

#estimate rates and probabilities from fully observed data
for (i in 1:3){
temp[i]<-lambda[i]*E[i]
m[i]~dpois(temp[i])
#Binomial likelihood for observed data
tr[i,1]~ dbin(condP[i],m[i])

}
G[1,1]<--G[1,2]-G[1,4]
G[1,2]<-lambda[1]*condP[1]
G[1,3]<-G[1,2]+G[1,1]+G[1,4]
G[1,4]<-lambda[1]*(1-condP[1])

G[2,1]<-G[2,2]+G[2,3]+G[2,4]
G[2,2]<--G[2,3]-G[2,4]
G[2,3]<-lambda[2]*condP[2]
G[2,4]<-lambda[2]*(1-condP[2])

G[3,1]<-G[3,2]+G[3,3]+G[3,4]
G[3,3]<--G[3,2]-G[3,4]
G[3,2]<-lambda[3]*condP[3]
G[3,4]<-lambda[3]*(1-condP[3])

# estimate QALYs and Costs
```

```
for(i in 1:QALY_A_dta_pts){QALY_A_dta[i] ~ dnorm(mu_QALY[1],
prec_QALY[1])}
for(v in 1:QALY_B_dta_pts){QALY_B_dta[v] ~ dnorm(mu_QALY[2],
prec_QALY[2])}
#priors

for (s in 1:3){
lambda[s]~dgamma(0.1,0.1)
condP[s] ~dbeta(1,1)}

for(t in 1:2){
mu_QALY[t]~dnorm(0,1.00000E-01)
prec_QALY[t] <- pow(sigma.QALY[t],-2)
sigma.QALY[t] ~ dunif(0,10)}
}
}
```

Supplementary material S4.2

Table S3 Expected Value of Information for healthcare payers and manufacturers.

No constraints			
CE Threshold	15,000	25,000	35,000
Payer	8,233	16,247	7,875
Manufacturer	17,236	12,705	8,744
Constraint	**Expected risk ratio of transplant**	**min MCD = 1.2**	
CE Threshold	15,000	25,000	35,000
Healthcare Payer	8,233	16,247	7,875
Manufacturer			
With Unacceptable risk	15,290	30,625	38,635
w/o Unacceptable risk	15,290	8,693	3,209
Constraint	**Max BIA at 24 months**	**max BIA = 80000**	
CE Threshold	15,000	25,000	35,000
Healthcare Payer	8,233	16,247	7,875
Manufacturer			
With Unacceptable risk	16,183	24,045	25,179
Without unacceptable risk	16,183	2,113	-10,247
Constraint	**MCD and BIA**		
CE Threshold	15,000	25,000	35,000
Healthcare Payer	8,233	16,247	7,875
Manufacturer			
With Unacceptable risk	14296.73	20905.11	21649.03
Without unacceptable risk	14296.73	-1,027	-13,777

For manufacturers results are reported with and without constraints

Supplementary material S5.1

Structured interview guide on CED schemes for devices

SECTION A

Name of interviewee: _____

Position: _____

Years of relevant work experience: _____

Name of organization: _____

NOTE: only summary data will be reported and individual respondents will not be identified

1. At present in your organization, do you implement CED schemes for health technologies?

Yes, for medicines only	☐	
Yes, for medical devices only	☐	
Yes, for other technologies	☐	Please specify:
Yes, for all technologies	☐	
No	☐	

<follow-up questions 2-6 if answer to Q1 is "Yes" (any technology)>

2. If CED schemes for medical devices are currently implemented, please provide the names of the existing policies or programs underpinning the schemes

3. If CED schemes for medical devices are not currently implemented, has your organization ever considered implementing them?

☐

No, never considered

Yes, in the past but it was later abandoned/discontinued ☐

Yes, in the present/ongoing ☐

4. Please explain your answer

5. Are there any guidelines available on how to design or implement CED schemes that are used in your organization?

Yes, for all technologies ☐

Yes, for medical devices only ☐

Yes, for medicines only ☐

Yes, for other technologies ☐

No, guidelines do not exist ☐

No, but guidelines are currently under development ☐

6. If guidelines exist for medical devices, please provide a reference or link to the guidelines.

<follow-up questions 7-8 if answer to Q1 is 'No'>

7. Has your organization ever considered implementing CED schemes for medical devices?

No, never considered ☐

Yes, in the past but it was later abandoned/discontinued ☐

Yes, in the present/ongoing ☐

8. Please explain your answer.

SECTION B

Based on the literature on CED schemes for medical devices, we have identified challenges that may affect a successful design and implementation of schemes. We now want to ask your thoughts on these challenges, and explore how these are addressed in your organization. When answering questions in this section, please consider CED schemes for medical devices only.

1. Please rate the following challenges on a scale of '0' (Not a challenge) to '5' (Major challenge)

(1) Deciding which medical devices are candidates for a CED scheme						
	0	1	2	3	4	5
To what extent is this a challenge for your organization? (0 = not a challenge, 5 = major challenge)	☐	☐	☐	☐	☐	☐
Please explain your answer:						
If CED schemes for MDs are currently implemented, what criteria are used to decide on which devices are candidates?						

(2) Getting stakeholder agreement on the scheme (e.g., Ministry of Health, manufacturers, hospitals, clinicians, patient associations)						
	0	1	2	3	4	5
To what extent is this a challenge for your organization? (0 = not a challenge, 5 = major challenge)	☐	☐	☐	☐	☐	☐
Please explain your answer:						

If CED schemes for MDs are currently implemented, what efforts do you make to engage with stakeholders and get an agreement on the scheme?

(3) Securing funding for the scheme						
	0	1	2	3	4	5
To what extent is this a challenge for your organization? (0 = not a challenge, 5 = major challenge)	☐	☐	☐	☐	☐	☐
Please explain your answer:						
If CED schemes for MDs are currently implemented, how is the funding secured to run the scheme?						

(4) Determining the appropriate study design for data collection (e.g., RCT, registry, audit)						
	0	1	2	3	4	5
To what extent is this a challenge for your organization? (0 = not a challenge, 5 = major challenge)	☐	☐	☐	☐	☐	☐
Please explain your answer:						
If CED schemes for MDs are currently implemented, how is the study design defined and agreed among all relevant stakeholders?						

(5) Determining the relevant outcome measure(s) to be collected in the scheme						
	0	1	2	3	4	5
To what extent is this a challenge for your organization? (0 = not a challenge, 5 = major challenge)	☐	☐	☐	☐	☐	☐
Please explain your answer:						
If CED schemes for MDs are currently implemented, how are decisions concerning the outcomes measures made? How is agreement reached among stakeholders?						

(6) Dealing with data collection and monitoring (e.g. who collects the data? Address risk of incomplete/partial data)						
	0	1	2	3	4	5
To what extent is this a challenge for your organization? (0 = not a challenge, 5 = major challenge)	☐	☐	☐	☐	☐	☐
Please explain your answer:						
If CED schemes for MDs are currently implemented, who is responsible for data collection and quality monitoring?						

(7) Dealing with analysis of the data (e.g. who performs the analysis, how is the risk of bias addressed)						
	0	1	2	3	4	5

To what extent is this a challenge for your organization? (0 = not a challenge, 5 = major challenge)	☐	☐	☐	☐	☐	☐
Please explain your answer:						
If CED schemes for MDs are currently implemented, who is responsible for the analysis of the data collected?						

(8) Determining the decision rule prior to the start of the scheme, based on the outcome of the scheme (e.g. reimbursement status, price change or refund)

	0	1	2	3	4	5
To what extent is this a challenge for your organization? (0 = not a challenge, 5 = major challenge)	☐	☐	☐	☐	☐	☐
Please explain your answer:						
If CED schemes for MDs are currently implemented, by what process is the decision rule determined (or not)? (if a decision rule is established please detail the criteria used)						

(9) Reaching an agreement on price, reimbursement or use of the device at the end of the scheme.

	0	1	2	3	4	5
To what extent is this a challenge for your organization? (0 = not a challenge, 5 = major challenge)	☐	☐	☐	☐	☐	☐
Please explain your answer:						

If CED schemes for MDs are currently implemented, how is an agreement reached on prices, reimbursement or use of a device at the end of a scheme?

(10) Withdrawing devices from the market when found not to be clinically or cost-effective.						
	0	1	2	3	4	5
To what extent is this a challenge for your organization? (0 = not a challenge, 5 = major challenge)	☐	☐	☐	☐	☐	☐
Please explain your answer:						
If CED schemes for MDs are currently implemented, what actions do you take to withdraw devices that are not found to be effective or cost-effective? How do you deal with potential reactions from physicians, patients and the general public?						

(11) Agreeing the length of the scheme or stopping rule						
	0	1	2	3	4	5
To what extent is this a challenge for your organization? (0 = not a challenge, 5 = major challenge)	☐	☐	☐	☐	☐	☐
Please explain your answer:						

If CED schemes for MDs are currently implemented, how are the duration of the scheme and/or the stopping rule defined and agreed among relevant stakeholders?

(12) Adapting the scheme to deal with product modifications or the existence of a learning curve						
	0	1	2	3	4	5
To what extent is this a challenge for your organization? (0 = not a challenge, 5 = major challenge)	☐	☐	☐	☐	☐	☐
Please explain your answer:						
If CED schemes for MDs are currently implemented, is the possibility of incremental innovations and/or the existence of a learning curve addressed in the schemes? If so, how?						

(13) Dealing with similar products entering the market (e.g., should they be entered into the scheme, could they undermine the scheme?)						
	0	1	2	3	4	5
To what extent is this a challenge for your organization? (0 = not a challenge, 5 = major challenge)	☐	☐	☐	☐	☐	☐
Please explain your answer:						
If CED schemes for MDs are currently implemented, how do you deal with similar products entering the market during the scheme?						

2. Are there other factors in your view that may help or hinder the successful implementation of CED schemes?

SECTION C

NOTE: Please include only CED schemes on medical devices that are currently ongoing or were terminated in the past 5 years. If information is not available, please enter "Not Available", if information cannot be made public, please enter "Classified".

Please complete a separate form for each CED scheme for medical devices

1. Basic information on the scheme

1. 1 Year of initiation of scheme:

_____

__

1.2 ID of the scheme:

_____

_____

1.3 Description of the product(s) and/or procedure(s) involved:

1.4 Disease/Indication(s) included in the scheme:

1.5 Manufacturer(s) involved:

1.6 Actual status of the scheme (choose one of the answer options):

☐ Scheme agreed, data collection to be initiated

☐ Data collection active

☐ Data collected, re-assessment to be done

☐ Data collected, re-assessment done

1.7 How was this MD selected for the scheme:

1.8 Health care organization/insurer/payer promoting the scheme:

1.9 Agency running the scheme (if not insurer/payer):

1.10 Purpose of scheme (choose one between the suggested options or enter a new one):

☐ To confirm reimbursement status at the end of the scheme (e.g. if effectiveness or cost-effectiveness are confirmed, the technology will receive an unconditional reimbursement)

☐ To inform price decisions (e.g. if agreed clinical outcomes are not reached at the end of the scheme, price discounts are increased by a pre-agreed amount)

☐ To confirm use of the device for specific subpopulations and/or indications

☐ To ensure effective and/or cost-effective use of the device in clinical practice (e.g., payment or reimbursement linked to individual patients' outcomes)

☐ To ensure appropriateness and quality of care (e.g., reimbursement conditional on providers compliance with clinical guidelines, or with pre-defined patients selection criteria)

☐

_____

_____

2. Design of the scheme

2.1 Agreed length of scheme (years):

_____

2.2. Description of the agreement (e.g., refund in case of non-response/treatment failure):

2.3 Key sources of uncertainty at scheme initiation (choose among the suggested options or enter new ones):

☐ The efficacy/effectiveness of the technology in a tested population as compared to current standard of care

☐ The relative efficacy/effectiveness in the real target population and/or in different population subgroups

☐ The effects on long-term patient relevant outcomes (versus surrogates outcomes used in clinical studies)

☐ The effect of physicians learning curves on patient outcomes and/or costs

☐ The risk of adverse events and/or adherence problems

☐ Uncertainties related to the true budget impact of introducing the technology

☐ Uncertainties related to the true cost-effectiveness of the technology

☐ Uncertainties related to the organizational impact of the technology

☐

_____

2.4 Design of the study (choose between the suggested options or enter a new one):

☐ Randomized controlled trial

☐ Cohort study

☐ Set up of a registry

☐

_____

2.5 Primary outcomes being measured in the scheme:

1) _____

2) _____

3) _____

4) _____

5) _____

2.6 Secondary outcomes being measured in the scheme:

1) _____

2) _____

3) _____

4) _____

5) _____

2.7 Source of funding for scheme:

3. Outcomes of the scheme

3.1 Public source of evidence about the scheme, if any (e.g. peer-reviewed articles, manufacturer's website, public registries):

3.2 Results (if scheme completed):

a) evidence generated by scheme

b) decision(s) made as a result of the scheme

3.3 Overall impression on successful aspects of the scheme:

3.4 Overall impression on failure aspects of the scheme:

3.5 Is there anything you would do differently when designing/applying a CED scheme for this type of MD in the future?

3.6 Other observations about the scheme:

Supplementary material S5.2

Table S4 List of challenges included in the questionnaire to decision-makers as modified in chapter

#	Chapter 2.6	#	Challenges included in the questionnaire
1	Deciding on whether a CED scheme is required	1	Deciding which medical devices are candidates for CED schemes
2	Information asymmetry between payers and manufacturers about the potential real-world performance of a technology		
3	Lengthy and complex negotiations	2	Obtaining stakeholder agreement on the scheme
4	Lack of transparency		
5	Lack of governance		
6	Stakeholder involvement		
7	Obtaining funding	3	Securing funding for the scheme
8	Understanding the relevant uncertainties and risks	4	Determining the appropriate study design for data collection
9	Defining the decision problem		
10	Data requirements		
11	Obtaining informed consent		
12	Ethical issues		
13	Identifying meaningful outcomes	5	Determining the relevant outcome measure(s) on which data are collected
14	Quality of the data	6	Dealing with data collection and monitoring
		7	Dealing with data analysis
		8	Adapting the scheme to account for product modifications or a learning curve
15	Ex-ante definition of a final decision rule	9	Ex-ante definition of decision rule, based on possible outcomes of the scheme
16	Deciding on when a CED is considered successful	10	Reaching an agreement on price, reimbursement or use of the device at the end of the scheme
17	Withdrawing a technology	11	Withdrawing a device from the market when evidence indicates the device is not (cost-) effective

18	Defining an adequate duration for a scheme	12	Obtaining agreements about the duration of the scheme and the stopping rule
19	Market entry of new technologies	13	Dealing with the market entry of similar devices
20	Economies of scale in the management of CED schemes and the difficulties small countries may have in applying CED schemes because of the associated costs and monitoring mechanisms		This challenge was considered transversal to all included challenges and covered in section A of the questionnaire.

In order to reduce the participants' burden when attending the interview, the list of challenges reported in Section 2.6 was reduced to 13 challenges by grouping some of the reported items into broader overarching challenges. For example, the challenges regarding the lengthy and complex negotiations to reach consensus on a scheme, the lack of transparency and lack of governance, as well as the challenges related to involving relevant stakeholders were grouped into the broader item "Obtaining stakeholder agreement on the scheme". Nonetheless, the challenge regarding the quality of the data collected in a scheme was expanded to explicitly address three separate issues: dealing with data collection and monitoring, dealing with data analysis, and dealing with data issues when accounting for product modifications or the existence of a learning curve effect.

Supplementary material S5.3

Table S5 Overview of the countries, jurisdictions and institutions of the participants to the survey

Country	Jurisdiction	Type of institution	Institution of the participant
Austria		National decision body	Ludwig Boltzmann Institute for Health Technology Assessment (LBI-HTA)
Belgium		National/regional decision body	National Institute of Health and Disability Insurance (RIZIV)
Bulgaria		University	Medical University Sofia
Czech Republic		University	Czech Technical University in Prague
Denmark		Hospital	Odense University Hospital
England		National decision body	National Institute of Health and Care Excellence (NICE)
Finland		National decision body	Finnish National Institute for Health and Welfare[a].
France		National decision body	French National Authority for Health (HAS);
		National decision body	French Medicine Pricing Committee (CEPS)
Germany		National decision body	The Federal Joint Committee (G-BA)
Greece		Hospital	Onassis Cardiac Surgery Center (OSCS)
Hungary		Health insurance	National Health Insurance Fund of Hungary
Ireland		National decision body	Health Information and Quality Authority (HIQA)
Italy	National level	National/regional decision body	Italian Medicine Agency (AIFA)
Italy	Emilia Romagna Region	Regional decision body	Clinical Governance area - Emilia Romagna Region
The Netherlands		National decision body	National Health Care Institute (ZIN)
Norway		Hospital	Norwegian Hospital Procurement Trust, Division Pharmaceuticals (LIS)
Poland		National decision body	National Health Fund
Portugal		National decision body	National Authority of Medicines and Health Products (Infarmed)
Scotland		National decision body	Health Improvement Scotland
Slovakia		Health insurance	Union Health Insurance

Spain	Basque Region	Regional decision body	Basque Office for Health Technology Assessment (Osteba)
Sweden		University	Linköping University
Switzerland		National decision body	Federal Office of Public Health (FOPH)

Supplementary material S5.4

Table S6 Factors with positive and negative influence on challenges with CED schemes for devices

	Challenge	Factors with positive influence	Factors with negative influence
1	Deciding which medical devices are candidates for CED schemes	There is a structured process leading to the identification of potential candidates for CED schemes. Prioritization and inclusion of technologies into a scheme is made according to explicit and shared criteria. The suitability of the proposed study protocol is a pre-condition to inclusion of a technology in a scheme. The request to provide additional data is applied to all technologies for which relevant evidence gaps are identified during an assessment and the main responsibility of data collection falls on the manufacturers/applicants.	HTA processes for devices are less formalized, commissioning mainly occurs at the local level. A high number of devices and lack of horizon scanning processes to inform candidates for CED schemes of medical devices. Optimal allocation of the funds for CED schemes is hampered by the fact that proposals are evaluated at different times over the year. It is not easy to establish whether the available evidence is sufficient to initiate CED scheme or whether it is too early for reimbursement.
2	Obtaining stakeholder agreement on the scheme	There exists a well-defined and structured processes for stakeholder engagement. All details of the scheme, including the roles and obligations of the stakeholders involved are defined in a contractual agreement before scheme initiation. Relationships with clinicians and manufacturers are facilitated if CED schemes are perceived as the only means to use the technology. The responsibility to collect the data (and coordinate with participating centres and other stakeholders) fall	The complexity of CED schemes and the different expectations of the stakeholders involved require a strong and time-consuming coordinating effort. For devices, it is more difficult to find patients to participate in public consultations during the scheme (e.g., compared to pharmaceuticals). In countries with small markets manufacturers may have a high bargaining power when discussing the conditions for the schemes.

		on manufacturers/applicants.	
3	Securing funding for the scheme	Fixed budgets for CED schemes are granted on a periodic basis. The additional costs of running a scheme fall upon the manufacturers/applicants.	Lack of *ad hoc* funds and/or human resources to run the schemes
4	Determining the appropriate study design for data collection	The health authority can explicitly or implicitly mandate the type of study to be conducted. Study design is defined by a third-party research institution. CED schemes are mostly relying on routinely collected data. A registry on the disease/device is already in place and suitable to answer the research questions.	Setting up the research governance is usually complex, with several organizations involved and many practical questions to answer There may be disagreement on study design between the government, the manufacturers and the providers Selecting the centres that will collect data for the schemes may be problematic and time consuming. Original patients' informed consent for registries may not allow subsequent analyses of data.
5	Determining the relevant outcome measure(s) on which data are collected	The health authority defines the primary and secondary outcomes. Those responsible for carrying out the research must justify if they do not follow the indication. Clinicians and experts are involved from the onset in the definition of the outcomes. Previous evidence from the literature or international collaborations (e.g., EunetHTA reports) already outlined the most relevant outcomes.	Relevant safety and effectiveness issues are more difficult to identify for devices compared to drugs at the time of the evaluation. Patient Reported Outcomes data are generally difficult to collect. A balance is required between what outcomes would be desirable and what can be pragmatically collected by the participating centres Different stakeholders may disagree on the relative importance of the outcomes to be collected (e.g. surrogate

			versus patient relevant outcomes).
6	Dealing with data collection and monitoring	Data collection is based on routinely collected data from electronic sources (e.g., electronic health records). Feasibility of the data collection burden is discussed and agreed among all actors involved at the beginning of the scheme. There is interoperability of data across data sources and research centres/providers. Continuous follow-up is done to check the quality and validity of data submitted and to ensure meaningful data is being collected.	There is less availability of routinely collected outcomes data for devices compared to pharmaceuticals. Uncertainties on devices may require the collection of long-term outcome data, incompatible with the length of the scheme. Having to deal with many low-volume centres with different experience may affect data quality, and increase the collection effort. Hospitals/participating centres may lack incentives to provide timely and high quality data if they do not receive specific funding for this task. Recruitment may be slower than expected affecting the time when the scheme reports its results.
7	Dealing with data analysis	An independent research body is appointed for data analysis, including quality and risk of bias assessment. There is an established experience with data analysis.	Difficult to find adequate controls with observational studies. Getting the analysis done and timely delivered may be difficult if no additional funds are provided for this task.
8	Ex-ante definition of decision rule, based on possible outcomes of the scheme	Schemes are only about collection of new data. Decision rules, including stopping rules during the schemes, and management of specific cases at the end of the data collection (e.g., insufficient quality of data, technology not effective) are defined in a contract agreed among all parties involved.	Fixed decision rules at the onset may be affected by unforeseen changes in the devices or market dynamics.
9	Reaching an agreement on price, reimbursement or use of the device at the	At the end of the scheme, technologies are re-evaluated according to business as usual evaluation procedures.	The scheme may not have collected the planned data by the time of the reassessment, or data may be un-conclusory.

	end of the scheme		Relevant differences in (cost) effectiveness less clear among similar devices compared to pharmaceuticals
10	Withdrawing a device from the market when evidence indicates the device is not (cost-) effective	An exit strategy in case the technology is not (cost) effective is defined at the onset in a contract agreed between all stakeholders involved. Having a well-designed scheme which produces scientifically robust results.	Patients and manufacturers may challenge the withdrawal decision and take actions against it. The management of explants for implantable devices in case the study outlines safety issues is complex.
11	Obtaining agreements about the duration of the scheme and the stopping rule	The duration of the scheme is agreed based on the time that is needed to collect the required data and the characteristics of the disease/technology. Continuation of the scheme is linked to periodic monitoring on its progresses.	Adopting the stopping rules defined at the onset of the scheme may be difficult when the scheme is ongoing. There is a tension between the short life-cycle of devices and the need for long-term outcomes. Different perspectives among involved stakeholders (e.g. clinicians, manufacturers, NHS and HTA bodies). Slow recruiting may impact on the time when the study reports its results.
12	Adapting the scheme to account for product modifications or a learning curve	The time frame of the scheme is relatively short to avoid product modifications. Considerations on the eligibility of a device to a scheme also consider if newer generations of the same devices are expected in the short-term. The company shares in advance available information on potential evolutions of the device and these are considered when discussing the study protocol. Data on the effect of the learning curve is publicly available.	There is little policy experience with how to deal with product modifications and/or learning curves. Interpretation of results are confounded by product modifications that occur during the time-frame of the study. Existence of a learning curve may complicate the selection process of participating centres in the scheme.
13	Dealing with the market	Schemes evaluate the class of devices or the procedure, not individual devices.	Identifying similar devices entering the market is hampered by

	entry of similar devices	A scheme can involve multiple devices from different manufacturers Schemes are not comparative in nature. Any similar product entering the market may be requested to provide additional data or not based on their level of evidence. Manufacturers of similar devices entering the market after scheme is initiated may be required to provide data to the same nationally-wide registry.	the difficulty to do horizon scanning for devices. More rapid changes in clinical practice with devices compared to pharmaceuticals. Inclusion of a new device entering the market when the scheme is ongoing may be more difficult than including it from the beginning.

CED, coverage with evidence development