London School of Economics and Political Science

A critical appraisal of coverage and resource allocation decisions through the use of health technology assessment: evidence on orphan drugs from four countries

Elena Nicod Gallmann

A thesis submitted to the Department of Social Policy at the London School of Economics for the degree of Doctor of Philosophy

London, December, 2015

Declaration of Authorship

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it). The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 73'635 words including footnotes, excluding references and appendices.

Statement of conjoint work

I confirm that Chapters 6, 7 and 9 (Papers 1, 2 and 4) were jointly co-authored with Dr. Panos Kanavos (PK), who provided guidance on the structure of the chapters and reviewed the papers once the paper was drafted with results from the literature review, data analysis, discussion and conclusion. Chapter 10 (Paper 5) was jointly co-authored with Karen Brigham (KB), Dr Isabelle Durand-Zaleski (IDZ), and PK. I wrote the proposal and interview topic guide. KB and IDZ commented on the interview topic guide developed for the French context. Interviews were conducted on my own in Sweden, England and Scotland, and with the help of KB and IDZ in France. I drafted the introduction, methods, results, discussion and conclusions. KB, IDZ and PK commented on these. Chapter 8 (Paper 3) was done entirely on my own. I also fully drafted the introduction, conceptual background, review of the literature and conclusions (Chapters 1-5 & 11), which was reviewed by PK.

Abstract

Health Technology Assessment (HTA) relies on evidence-based medicine to inform drug coverage recommendations about the most efficient use of resources. Despite appraising the same evidence based on similar methodological approaches, HTA recommendations for the same drug differ across countries. This thesis aimed to understand the reasons for these differences, and based on cross-national comparisons, whether they are a consequence of methodological challenges in HTA. A mixed methods research design was used to develop a methodological framework that allows to breakdown these complex processes in a comparable and understandable manner, by considering: (a) the evidence appraised, (b) its interpretation, and (c) how this influenced the final decision. Ten orphan drug-indication pairs appraised in four countries (England, Scotland, Sweden and France (N=35)) were systematically analysed and compared on this basis. Results present the criteria accounted for at each stage of the process in the decisions, the reasons for differences across countries, and how HTA bodies are dealing with issues relating to orphan drugs. Quantitative analysis of these provided information about agency-specific risk and value preferences, and measured agreement in interpreting the same evidence. There was heterogeneity within and across countries in the criteria accounted for and reasons for cross-country differences. Interviews to competent authorities provided insights about these differences and implications for HTA. Although agreement was seen in the evidentiary requirements or preferences, there were subtle differences in the circumstances where uncertain evidence may be considered acceptable, partly explaining diverging HTA recommendations. The three main contributions of this thesis are: (1) the development of a methodological framework to understand what criteria feed into HTAs, which can be applied to other drugs and countries; (2) through its application, the identification of a full taxonomy of criteria considered in decision-making; and (3) the ability to understand the differences in HTA recommendations across countries. A better understanding of HTA in different settings may help advance these processes, and, ultimately, improve access to treatments.

Acknowledgments

First and foremost, I would like to thank my dear husband Francesco for his patience and encouragements and for being an important part of this whole adventure. A warm thanks also to my wonderful parents, Pascal and Rose Marie, who have always encouraged me to pursue what I believe in. I thank also my siblings and their extensions, who are my lifetime roots, wonderfully inspiring and loving: Cristina, Philippe, Emile, Isabel, Cesar, Albert, Eleonore, Edouard, Beatrice, Paul, Ana and soon to be in-law Alex. Thanks also to my dear Abuelita, for her endless love and to all my cousins, aunts and uncles, which I truly adore. Finally, a special thanks to my mother-in-law, Paola, for her kindness and valuable support.

Very importantly, I would also like to thank my PhD supervisor Panos Kanavos, for introducing me to this area many years ago, which I am very passionate about, and for supporting me all these years in so many ways. Thanks also to my supervisor Paul Dolan and LSE mentor Caroline Rudisill, who have provided valuable advice and wisdom throughout these years. I am also grateful to our Doctoral Programme Directors, Prof Anne West and Prof Jane Lee, as well as our Research Degree Programmes Administrator Sarah Helias for all their help and support. A warm thank you also to Prof Martin Knapp and Prof Johann Severens for committing to examine this PhD, and for their constructive and pertinent comments, which not only have helped improve this PhD but will also be useful for the future.

A special thanks also to my colleagues at LSE Health, the Medical Technology Research Group, and PhD fellows from the Department of Social Policy at LSE. A particular thanks also to Emilie Courtin, Alessandra Ferrario, Huseyin Naci, Laura Schlang for all those animated discussions and constructive criticisms when revising versions of these chapters.

For their constructive feedback and support throughout the years, I would like to thank all partners from the Advance-HTA project, particularly Karen Facey, Karen Berg Brigham, Isabelle Durand-Zaleski, David Taylor, Jean Mossmann, Mary Baker, Aris

Angelis, Dimitra Pantelli, Sabine Fuchs, Julio Bastida Lopez, Laia Maynou-Pujolras, Marta Trapero, Francis Ruiz, Douglas Lundin, Karl Arnberg, Jaime Espin, Jimena Luna, Anna Zawada, Kinga Grywacz, Lise Rochaix, Reinhard Busse, Domenica Tasruscio, Finn Borlum Kristensen and John Cairns. A special thanks also to the interviewees who took the time to prepare, participate and respond to the interviews, which were extremely valuable for this PhD.

Finally, I would like to thank my dear friends who have accompanied me during this London experience despite hearing the same discourse for too many years: Estelle, Anne Caroline, Sylvie, Sophie, Amy, Laurent, Raquel, Dom and many more … and last but not least, my impressive cousin Alec and all my life-time friends back home!

This research was funded by the European Commission's 7th Framework Programme and was undertaken under the auspices of Advance-HTA (Grant number 305983). I thank the Commission for the opportunity to conduct this research and for the support provided. The results presented here reflect the author's views and not the views of the European Commission. The EC is not liable for any use of the information communicated.

I dedicate this thesis to my dear husband Francesco, who now finally comes first!

Table of Contents

List of Tables

List of Figures

List of Acronyms

ANSM	[Agence Nationale de Sécurité du Médicament et des Produits de Santé] - Drug Safety Agency
ASMR	[Amelioration du Service Medical Rendu] - Incremental medical benefit
ATU	[Autorisation Temporaire d'Utilisation] - Temporary authorisation
CCG	Clinical Commissioning Group
CEA	Cost-Effectiveness Analysis
CEPS	[Comité Economique des Produits de Santé]
CMA	Cost-Minimisation Analysis
CNS	Central Nervous System
CT	[Commission de Transparence] - Transparency Committee
CUA	Cost-Utility Analysis
CVZ	Dutch Health Care Insurance Board
DNL	Do not list or negative recommendation
EMA	European Medicines Agency
EUnetHTA	European Network for Health Technology Assessment
EVIDEM	Evidence and Value: Impact on Decision Making
GP	General Practitioner
HAS	Haute Autorité de Santé
HRQOL	Health Related Quality of Life
HTA	Health Technology Assessment
ICD10	International Classification of Diseases (version 10)
ICER	Incremental Cost-Effectiveness Ratio
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
k	Kappa Score
L	List or positive recommendation
LWC	List with restrictions
MAH	Marketing Autorisation Holder
MHRA	Medicines and Healthcare products Regulatory Agency
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
QALY	Quality-Adjusted Life Years
SMC	Scottish Medicines Consortium
SMR	[Service Medical Rendu] - Treatment's medical benefit
TLV	[Tandvårds- och läkemedelsförmånsverket], HTA body in Sweden
UNCAM	[Union Nationale des Caisses d'Assurance Maladie]
WTP	Willingness-to-Pay

Notes on the structure of this thesis

This thesis conforms to the requirements of a doctoral thesis from the London School of Economics and Political Science. It is in the form of a publishable papers thesis. Guidelines state that it should consist of a minimum of three papers of publishable standard and interlinked with each other, together with an introduction and conclusion. At least one of the papers should be single authored, and any other paper should be primarily authored by the Student. A specified detailed statement on the contribution of co-authors is provided. Word count should not exceed 100,000 words.

This thesis consists of five papers published or ready for submission for publication in high quality peer review journals. An introductory chapter provides information about the context, issue, and why this research question is important. Chapters 2 and 3 provide an overview of the pharmaceutical environment and conceptual background on health technology assessment (HTA), Chapter 4 outlines the gaps in the literature, hypotheses and research questions, Chapter 5 the overall methods applied, and Chapters 6 to 10 are in the form of publishable papers. Conclusions are given in Chapter 11 alongside the contribution of this thesis, its implications for policymaking, limitations and areas for further research.

This thesis was part of a larger European project called Advance-HTA. The PhD author was able to leverage on this network (e.g. HTA bodies for the interviews) and receive feedback about her work when presenting progress at various meetings, and by circulating drafts of chapters to all partners for comment. These contributions were not sufficient to justify co-authorship, but were acknowledged within the publications. Only for the last empirical chapter did the PhD author collaborate with partners from University Paris-Est Créteil, as described in the next paragraphs.

The introduction, background, literature review, research questions and overall PhD methods (Chapters 1-5) are solely the work of the author. PK and two colleagues (Emilie Courtin and Huseyin Naci) commented on draft(s) of these introductory sections.

Chapter 6 (Paper 1)

The first paper published in 2012 in *Health Policy* is primarily the work of the PhD author:

Nicod E, Kanavos P (2012). Commonalities and differences in HTA outcomes: a comparative analysis of five countries and implications for coverage decisions. *Health Policy*, 108(2-3):167-77.

The database of HTA recommendations analysed was compiled by the PhD author and two MSc colleagues, Stacey van den Aardweg and Steve Pomedli under the supervision of PK. EN conducted the literature review, data analysis, and drafted the paper. PK commented on drafts of the paper. The paper was subject to a double-blinded peer review before its publication.

Chapter 7 (Paper 2)

The second paper has been accepted pending revisions by *Health Policy* in October 2015 and currently awaiting final decision. It is primarily the work of the PhD author:

Nicod E, Kanavos P (2015). Developing an evidence-based methodological framework to systematically compare coverage decisions across countries: a mixed methods study. *Health Policy, 120(1):35-45*.

EN devised the paper, conducted the literature review, data analysis, and drafted the paper. PK commented on drafts of the paper. The paper was subjected to a double-blinded peer review and accepted for publication in November 2015.

Chapter 8 (Paper 3)

The second paper is solely the work of the PhD author. It has undergone the peer review process for *Value in Health* by two blinded reviewers, and was accepted with minor revisions by one and rejected by the other. On this basis, the Editor did not

accept the manuscript for publication. Based on the reviewers' comments, the manuscript has been revised and submitted to the *European Journal of Health Economics.*

Nicod E (2016). Why do health technology assessment coverage recommendations for the same drugs differ across settings? Applying a mixed methods framework to systematically compare orphan drug decisions in four European countries. *European Journal of Health Economics* (in press, accepted 2 August 2016).

EN devised the paper, conducted the literature review, data analysis, and drafted the paper.

Chapter 9 (Paper 4)

The fourth paper received revisions from two blinded reviewers and was recently accepted for publication by the *International Journal for Technology Assessment in Health Care*, and is primarily the work of the PhD author:

Nicod E, Kanavos P (2016). Scientific and social value judgments for orphan drugs in health technology assessment. *International Journal for Technology Assessment in Health Care, 32(4):1-15*

EN devised the paper, conducted the literature review, data analysis, and drafted the paper. PK commented on drafts of the paper. The paper is being subjected to a double-blinded peer review. Reviewer comments were received and have been incorporated in the current version.

Chapter 10 (Paper 5)

The fifth paper was conducted in collaboration with two partners from Advance-HTA: Isabelle Durand-Zaleski and Karen Brigham (University Paris-Est Créteil). It has been submitted to Value in Health in November 2016 and is under review.

Nicod E, Berg Brigham K, Durand-Zaleski I, Kanavos P. Differences in the approaches to assessing value in four countries for orphan drugs: Evidence from four European countries. Under review, *Value in Health* (November 2016).

EN devised the paper, conducted the literature review, drafted the interview topic guide, conducted the interviews, did the data analysis, and drafted the paper. IDZ and KB commented on the interview topic guide, participated in the interview of the Haute Autorité de Santé (in France), and reviewed the final draft. PK commented on the paper.

Chapter 11

EN is the primary contributor of the concluding remarks, policy implications, contributions of this research and future research agenda, which were also reviewed by PK.

1. Introduction

Health care decision-makers are currently facing major challenges in providing equal, effective and high quality care subject to budget constraints. This is partly due to increased drug expenditure accounting for a significant and ever rising proportion of national health care budgets as a consequence of population ageing (Drummond, Evans, LeLorier, Karakiewicz, Martin, Tugwell, & MacLeod, 2009a; NIH, 2011) and rising cost of new medicines (Bach, 2009; Congressional Budget Office, 2006; Light & Kantarjian, 2013). In order to control the level of expenditures and use resources more efficiently, a variety of mechanisms have been implemented to target pharmaceutical price, volume, use, and distribution methods. The focus of this doctoral thesis is on one such mechanism: health technology assessment (HTA) used for value assessments and coverage decisions.

Since its introduction in 1967 in the US, the increasing uptake of HTA has been driven by the need to control expenditures by seeking to obtain value for money (Banta, 2003). The US Congressional Office of Technology was the first to introduce and define HTA as "a comprehensive form of policy research that examines the short- and long-term social consequences of the application or use of technology" (United States Congress, 1976). It was first implemented in Europe in the 1980s with the creation of the Swedish Council on Technology Assessment in Health Care (SBU), followed by many countries in Europe and around the world (Banta & Jonsson, 2009). With its maturation and expansion over time, the definitions of HTA are defined more broadly and go beyond the synthesis of evidence about efficacy, safety and cost-effectiveness. It has been more recently defined as "a form of policy research that examines short- and long-term consequences of the application of a health-care technology. Properties assessed include evidence of safety, efficacy, patient-reported outcomes, real world effectiveness, cost and cost-effectiveness as well as social, legal, ethical, and political impacts" (ISPOR, 2003)

The rationale for having such a mechanism is to allow for a systematic identification of treatments which have clinical and cost-effectiveness or that provide additional therapeutic benefits when compared to existing standards of care, and to set priorities

accordingly. Theoretically, this should shift resources from cost-ineffective treatments or those with relatively inferior clinical benefit, and result in a more efficient use of health care resources under which greater value for money and access are achieved (Department of Health, 2010).

Although it is assumed that HTA outcomes are likely to be similar across settings because the clinical and safety evidence considered is - in the majority of cases - similar for the same drug-indication pair, in practice evidence from the literature suggests that HTA outcomes vary greatly across settings. HTA is a complex process that operates within a multidisciplinary field with differences at each stage. These variations are a consequence of disparities in the HTA processes (e.g. evidence considered, methods used for the assessment, interpretation of the evidence), their national context (e.g. budget constraints, prioritization of disease areas), the timing of the appraisals, the level of stakeholder involvement, and their implementation in practice (e.g. advisory or regulatory role) (Cairns, 2006; Clement, Harris, Li, Yong, Lee, & Manns, 2009; Drummond et al., 2009a; Kanavos, Manning, Taylor, Schurer, & Checchi, 2010a; Kanavos, Nicod, Van den Aardweg, & Pomedli, 2010b; Morgan, McMahon, Mitton, Roughead, Kirk, Kanavos, & Menon, 2006; Nicod, 2010; Pomedli, 2010; Richards, 2010; Sorenson, 2009; Sorenson, Drummond, & Kanavos, 2008; Van den Aardweg, 2010; Velasco-Garrido, Borlum Kristensen, Palmhoj Nielsen, & Busse, 2008; Velasco-Garrido & Busse, 2005). It may also be that some of these differences are a consequence of weaknesses in HTA methods and their application, possibly resulting in resources not being used optimally. If this were the case, patients would have access to certain treatments in some countries or regions and not in others. Patients would also incur the negative consequences of resources not being used optimally if their treatment was not reimbursed because of insufficient funds spent inefficiently elsewhere.

Consequently, in order to ensure that these cross-country differences are legitimate and not a consequence of weaknesses in HTA methods and their application, to increase transparency between regulators and other stakeholders to improve the quality of the submissions, and to improve HTA methods, there is a need to understand why they occur. The objectives of this thesis are in line with this need.

This thesis begins by discussing the European pharmaceutical regulatory environment and introducing the role of HTA amongst other market interventions (Chapter 2.1). The drug development process, followed by an overview of HTA and the characteristics of HTA bodies are then explained (Chapter 2.2-2.3). The subsequent chapter (Chapter 3) synthesises the findings from selected literature reviews conducted around four topics of interest: (a) HTA as a value assessment tool, which presents the different approaches to using HTA as a value assessment tool (Chapter 3.1); (b) the HTA framework and its main limitations, which outlines the methods used to undertake HTA and their limitations (Chapter 3.2); (c) the factors that distinguish orphan drugs from other disease areas (Chapter 3.3); and (d) HTA in different settings, which summarises the existing evidence that compared HTA recommendations across more than one country and therapy area (Chapter 3.4). These were used to identify the gaps in the literature and derive the hypotheses and research questions of this thesis, outlined in Chapter 4 together with the plan of the PhD. Chapter 5 summarises the methodological approaches used, while Chapters 6-10 are the empirical chapters of the thesis. Chapter 6 sets the scene for the remainder of the thesis by quantifying the extent and contradictory nature of the differences in HTA recommendations issued across countries, highlighting the need to query why they occur. Chapter 7 develops and pilots a methodological framework enabling to systematically compare HTA decision processes as reported in the HTA reports across countries. Chapter 8 then applies this framework to a larger sample of ten orphan drugs, providing a more structured and fuller understanding of the reasons for differences across countries. Chapter 9 focuses specifically on one aspect of the decision-making process: the scientific and social value judgments. It aimed to understand how the decision-makers' value judgments influenced the decisions and to further the debate as to whether orphan drugs deserve special status. Building on the findings from the previous empirical chapters, Chapter 10 aimed to develop a broader perspective about how value is assessed for orphan drugs and how differences affect reimbursement decisions based on interviews of representatives of the four European HTA bodies. The main contributions and conclusions of this thesis, policy implications, limitations and future research agenda are presented in the final chapter (Chapter 11).

2. The Pharmaceutical Environment and Health Technology Assessment

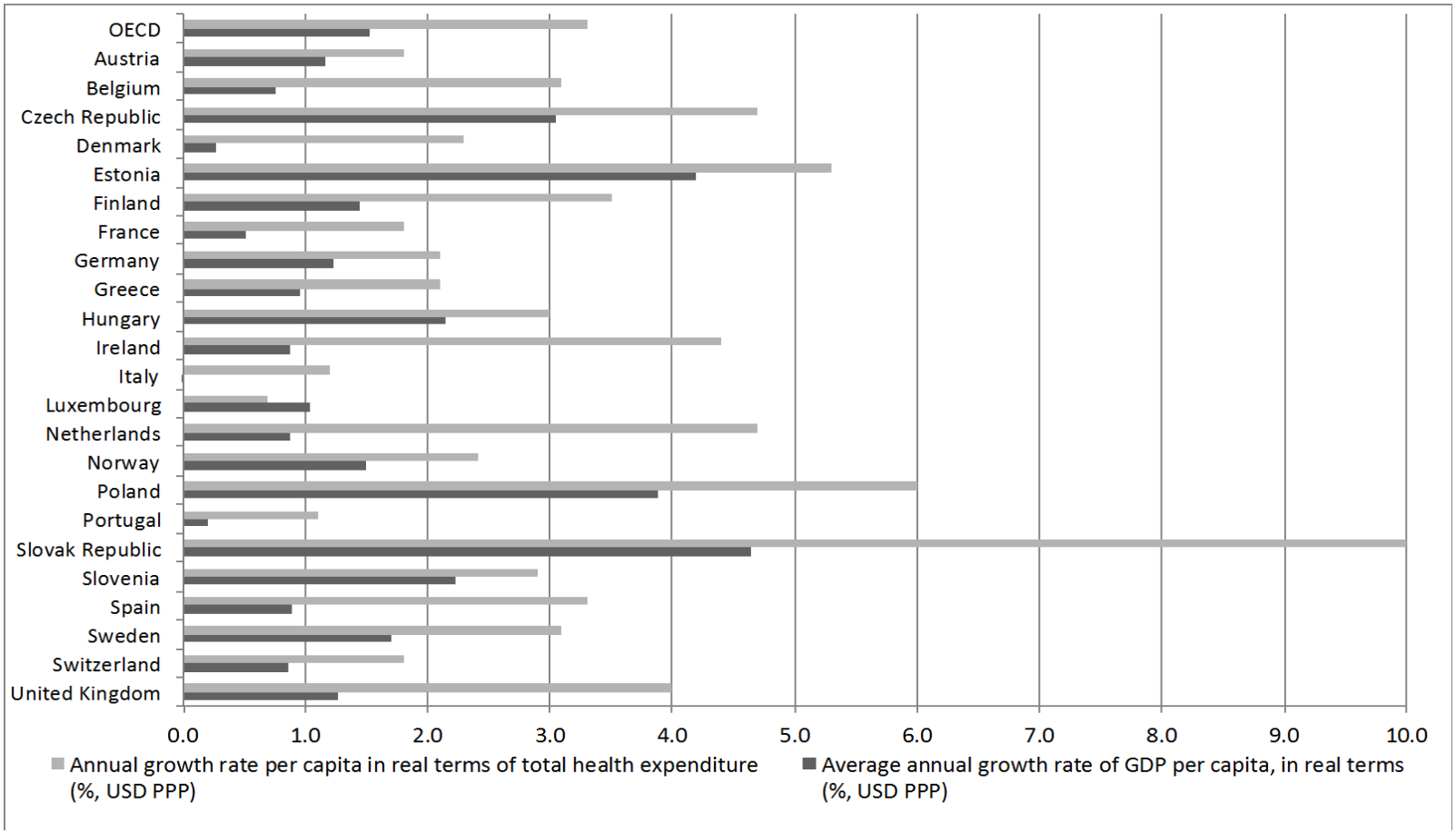
This chapter provides the conceptual background to this thesis and describes the pharmaceutical regulatory environment, the drug development process and where HTA stands in this process, and the characteristics of HTA bodies.

2.1. The European Pharmaceutical Regulatory Environment

Pharmaceutical expenditure accounts for a large proportion of the total health care budget estimated at an average of 16.6% of total health expenditures in European OECD countries in 2011 (OECD, 2014), and continues to rise every year (Figure 2-1). The main drivers of expenditure are the increasing volume of drugs being consumed and the high cost of the new medicines being marketed (Mousnad, Shafie, & Ibrahim, 2014). For pharmaceutical expenditures to remain sustainable, policymakers recognise that costs should be contained within budgets and used more efficiently while continuing to incentivise innovation. As a consequence, the European pharmaceutical market is heavily regulated for both on-patent and off-patent medicines (e.g. generics).

Pharmaceutical markets are imperfect and require a number of interventions. There are various factors that are responsible for this (McGuire, Drummond, & Rutten, 2002). First, patients with a given disease may receive different treatments due to variable symptoms, variable levels of tolerance to treatment, and various medical habits of patients and physicians. This results in an inconsistent demand for any particular medication. Second, the use of any particular medication is generally not decided by the consumer, but by the health provider, generating a so called “agency relationship” (Folland, Goodman, & Stano, 2013). This is due to the asymmetry of information existing between the provider and consumer, where the consumer generally trusts the choice of the provider given his lack of knowledge regarding his health status and treatment options.

Figure 2-1. Average annual growth rate of GDP per capita, in real terms, and annual growth rate per capita in real terms of total health care expenditure (%, USD PPP)



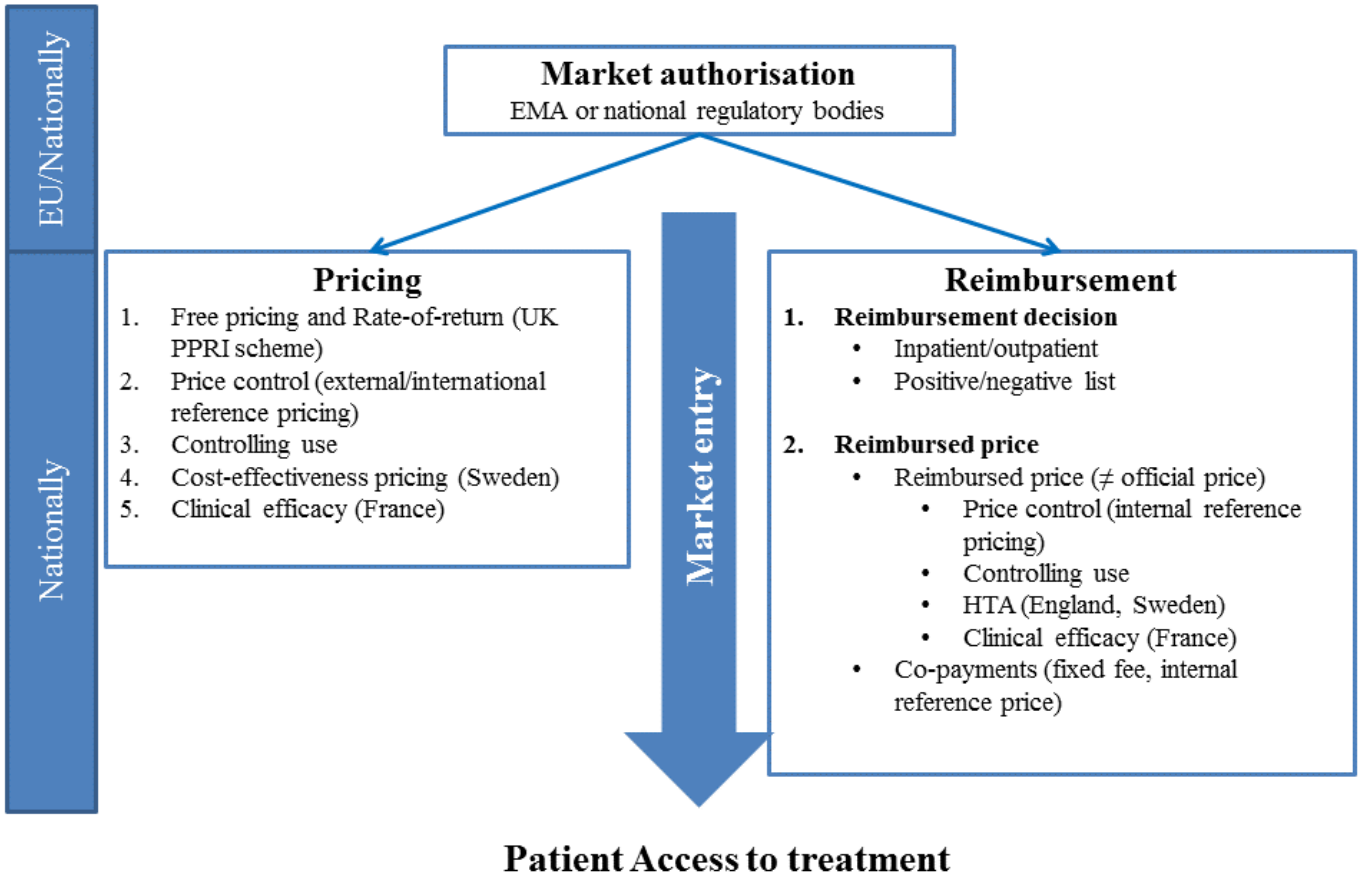
Source: (OECD Health Statistics, 2013).

Third, the existence of a third party payer, e.g. health insurance, introduces additional complications. Consumers may want to get as many and as expensive medications for themselves, if they do not cost them money (moral hazard) (Folland et al., 2013). Physicians can also make moral hazard worse, through the so-called "supplier-induced demand". On the other hand, the third party payer may restrict the use of medications by developing a list of accepted medications (positive list) or by limiting indications, among others. Fourth, the market may also be biased by the fact that some patients are willing to pay more money for a treatment that they believe will be superior. This introduces the concept of "elasticity of the demand": consumption of a given medication may be more or less affected by a change of price. If price increase results in a minor decrease in demand, the elasticity is considered low (inelastic). On the other hand, if a price increase results in a substantial change of consumption, the elasticity is considered high (elastic).

In such an imperfect market, regulatory interventions are needed to protect consumers and have been implemented both on the demand and supply-side and drive market access (Figure 2-2). On the demand-side comprising the consumers (e.g. patients) and providers (e.g. prescribers and dispensers), measures can be financial and non-financial. For the consumer, measures such as co-payment, co-insurance, a deductible or a flat fee per prescription have been implemented with positive and negative effects depending on their personal wealth. For example, high levels of co-payment can impoverish certain patients, whereas including a maximum amount to be paid by the consumer (deductibles or exoneration) may limit the amount of co-payments particularly for patients with chronic conditions. For the provider, prescription budgets, including compulsory generic prescription, can encourage physicians to prescribe generic drugs. Generic substitution and flat fee combined with regressive margins are measures encouraging pharmacists to dispense cheaper products.

On the supply-side, patent protection grants exclusive rights to a patented molecule from being used, copied or traded for a period of up to 20 years. This is to protect and reward innovation. In Europe, the Supplementary Protection Certificate (SPC) prolongs this period up to 5 years depending on the product's novelty.

Figure 2-2. Pricing and reimbursement pharmaceutical policies for market access



Source: The author.

During this time, the manufacturer benefits from a monopolistic position. This is countered by policies targeting pricing and reimbursement to ensure that excessive profits are not made from this monopolistic position. A variety of pricing mechanisms exist, where the price is based on: the expected returns (e.g. cost-plus pricing), the average price of a basket of countries (e.g. reference pricing), or the value of the treatment (e.g. value-based pricing). Prices can also be driven by negotiations between the manufacturer and the payer (e.g. price-volume agreements). Pricing is complemented by the conditions for reimbursement, which determines whether the treatment is reimbursed, at what price and with what level of co-payment. In parallel, a number of generic policies may facilitate generic penetration once patents expire (e.g. Bolar provision) (Kanavos, Costa-Font, & Seeley, 2008).

National drug policy objectives in Europe are relatively uniform and include: "universal access for all citizens, effective care for better health outcomes, efficient use of resources, high-quality services and responsiveness to patient concerns" (McGuire et al., 2002). Given the complexities of this market and its imperfections, we have seen that it is highly regulated and includes contradictory or complementary measures in order to reach a reasonable balance to achieve these goals. Despite these and due to the increasing (pharmaceutical) expenditures, we are facing issues around affordability where current budgets are insufficient to cover population needs. Therefore, decisions about which drugs national health care budgets cover have to be made. HTA is increasingly being used to support such decisions across Europe, Canada and Australia, and is increasingly used elsewhere in the world as it aims to ensure that health care resource allocation decisions are efficient. In this respect, it has recently been recognised during the World Health Assembly as a means towards universal health care (Sixty-seventh World Health Assembly, 2014).

The focus of this thesis is on HTA and its application across different Member States for new patented prescription drugs (given they are one of the main drivers of health care expenditures), with an orphan designation by the European Medicines Agency (given they are often high cost medicines and can be viewed as outliers for HTA). The next section describes the drug development process and where and why HTA fits in this process.

2.2. Drug Development Process

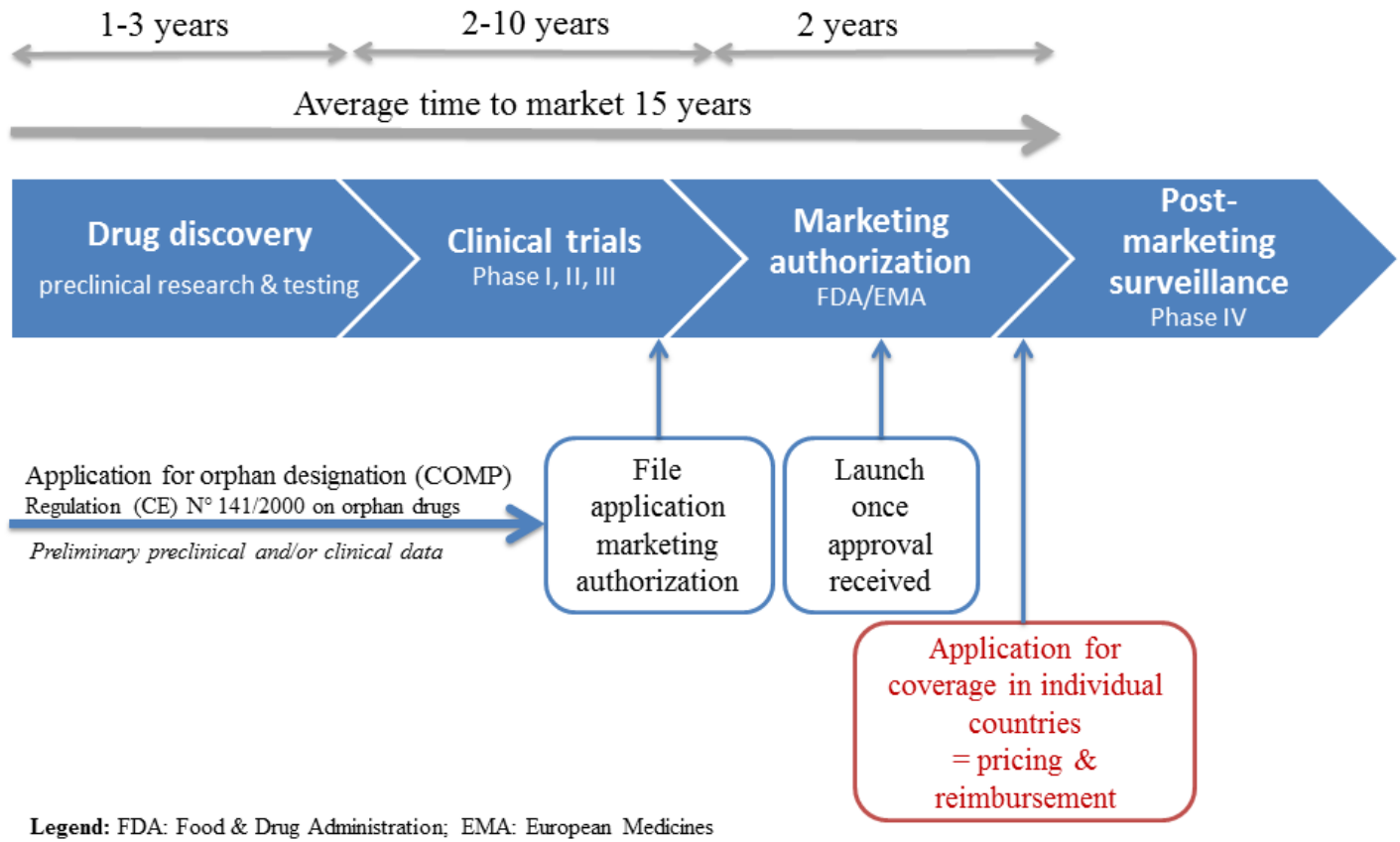
Across the drug development pipeline, very few drugs make it past clinical testing and only one in 57 drugs developed actually make it to the market (Portfolio Management Solutions). R&D costs are estimated to range from $800 million to $1.3 billion dollars, including the cost of failures (DiMasi & Grabowski, 2007; DiMasi, Hansen, & Grabowski, 2003; Vernon, Golec, & Dimasi, 2010). One recent study argues that this is an overestimate and is based on unknown factors (Light et al., 2013). This may be due to the true R&D cost that cannot be fully observed because of the cross-subsidy existing between successful and unsuccessful products and the uncertainties regarding innovative processes (McGuire et al., 2002). Nevertheless, high drug prices are often explained by the high attrition rates and the strict evidence standards required to pass regulatory hurdles. Additionally, very few of the new drugs entering the market are breakthrough or first-in-class: as little as 45% of drugs being tested in phase III trials are first-in-class, and may still potentially be unsuccessful (Long & Works, 2013). In cancer, one study showed that only twelve drugs approved in 2012 provided a survival benefit of more than two months (Light et al., 2013). Recently, the market is shifting from blockbuster to more niche and specialised markets, such as personalised medicines. One of these niche and specialized markets is for rare diseases, discussed in Chapter 3.3.

During the drug discovery, once a potential drug has been developed and appears promising, it can take up to ten years or more to perform the necessary clinical trials to generate evidence about its safety, quality and efficacy, which are requirements for marketing authorisation. In Europe, the marketing authorisation process occurs either at European level (“centralised procedure”) through the European Medicines Agency (EMA), or at national level (“decentralised procedure”). For example in the UK, the national regulatory body is the Medicines and Health care Products Regulatory Agency (MHRA).

Marketing authorization is not sufficient to ensure patient access to these new drugs. Once market authorization is granted for a new drug, the marketing authorization holder (MAH) (e.g. the manufacturer or sponsor) then files an application to obtain

coverage by the national health systems in each individual country to their local Pricing and Reimbursement authorities (Figure 2-3). Therefore, access to these treatments largely depends on the outcome of the pricing and reimbursement decisions, because such decisions drive both their affordability and availability. This consists of the reimbursement decision (e.g. yes/no), the coverage rate (e.g. % of the drug price covered), and the final drug price (e.g. cost-effectiveness pricing). In countries with HTA, these decisions are supported by whether the drug under review is considered to provide value for money, where the value of using this drug in real world settings in terms of costs and benefits is assessed, while considering the related social, ethical and legal implications.

Figure 2-3. Drug development process and orphan designation



Source: The author.

2.3. The characteristics of different HTA bodies

HTA bodies are generally in charge of deciding or making recommendations for the pricing and/or reimbursement of a technology, and are either at arm's length or integrated with other activities. The former includes England & Wales, Scotland, Sweden and France, (Appendix A), and the Dutch Health Care Insurance Board (Zorginstituut Nederland, ZIN) advising the government on the statutory health package while being responsible for risk equalisation and other activities. Another example of the latter is the AIFA in Italy, which is the Italian Medicines Agency in charge of marketing authorisation, reimbursement, and pricing of drugs through negotiations with manufacturers. HTA activities are generally funded with public monies (direct or indirect taxation, user fees, or other contributions), though this varies depending on the country.

Responsibilities and memberships depend on whether the HTA body is regulatory (with legally binding coverage recommendations), or advisory (who issues a recommendation to the final decision-maker (e.g. Health Ministry)). The general objective of the HTA will define the responsibilities in how it is conducted. The scope is generally outlined before initiation of the HTA process, and defines the aim of the review and the evidence required (e.g. stage of the HTA, perspective adopted) (Drummond, Schwartz, Jonsson, Luce, Neumann, Siebert, & Sullivan, 2008). Topic selection is a priority setting process about the technologies to be appraised, and are based on a number of criteria and types of technologies (Sorenson et al., 2008).

The burden of proof of value for money rests with the manufacturer, who is asked to submit an application outlining their product for review by the HTA bodies. Unlike marketing authorisation, HTA bodies may then conduct additional analyses in-house or by independent analysts (e.g. universities or expert groups), and focus on verifying the assumptions in the manufacturer's submission or in generating additional evidence and redoing an HTA. Each HTA body has their own HTA methodological requirements summarised in their submission guidelines, which may differ across countries. For example, the appropriate comparator may be the treatment that is intended to be replaced by the new drug or the most cost-effective standard of care.

The final outcome of HTA is in the form of a recommendation or decision about whether to reimburse, reimburse under certain conditions, or not reimburse the drug under review.

Even with a common objective in assessing whether new health technologies provide value for money, HTA bodies may vary in their responsibilities and membership, in the assessment procedures and methods, in the type and timing of the HTA, and in its dissemination and implementation (Chalkidou, Tunis, Lopert, Rochaix, Sawicki, Nasser, & Xerri, 2009; Sorenson et al., 2008). In this respect, while cross-national differences in the HTA coverage recommendations made are legitimate due to these contextual considerations, they may also be a consequence of how HTA is applied in different settings and its methodological limitations. Therefore, it is essential to understand and differentiate the reasons behind such cross-national differences, given that these have direct implications for patient access and indirect implications for society as a whole if resources are used inefficiently. This is in line with the research questions and objectives of this PhD, further discussed in Chapter 4. Beforehand, a selected review of the literature was performed in order to frame the problem and identify in a more systematic manner the gaps in the literature used to derive the research questions of this thesis. Four areas were explored and synthesised in the next chapter: (a) how HTA works as a value assessment tool, (b) the HTA framework and its limitations, (c) factors that distinguish orphan drugs from other drugs, and (d) the application of HTA in different settings.

3. Conceptual Background on HTA

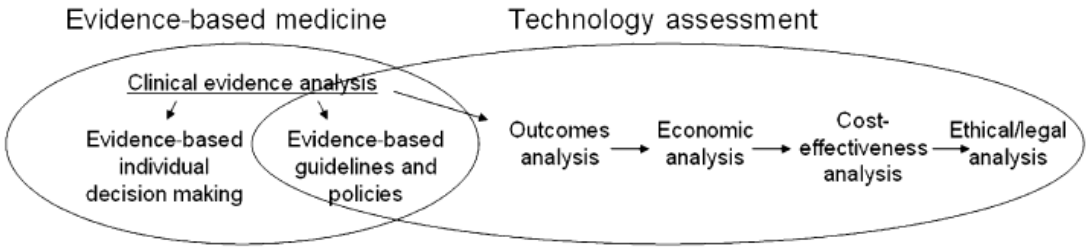
3.1. HTA as a Value Assessment Tool

This section discusses the fundamental role of HTA as a value assessment tool in the health system placing it into context based on a comprehensive review of the health services research, health policy, health economics, and medical literatures.

Given that the main goal of policymakers in the health care sector is to maximise health within a given budget, HTA aims to ensure that the price paid for a technology reflects its value and provides value for money (Hurley, 2000). Fundamental to the definition of HTA is the concept of "value" and how different stakeholders perceive it differently within and across health care systems. Value can be perceived within the context of efficiency, where only the most efficient technologies would be reimbursed within an allowable budget. This approach would not necessarily account for what really matters to society and to those being treated (Caro, 2009). Generally, there is agreement that HTA is an appropriate tool to measure value. The main issue is whether the notion of "value" captures what is valued by patients, carers and society. "Value" could then reflect specific attributes such as innovation, which may have direct implications for the patient (e.g. improved prognosis or quality of life) and society (e.g. better productivity and ability to contribute to society), but also potential spill-over effects that could benefit other patients and disease areas (e.g. new molecule or mechanism of action that could work in other disease areas). Therefore, value can be regarded as an umbrella concept that encompasses a range of facets relating to different stakeholder perspectives and capturing different attributes of interest (e.g. innovation).

The definition of HTA outlined in Chapter 1 captures well this multidisciplinary and multidimensional approach, as does the conceptualisation of a "full HTA" proposed by David Eddy, which can be completed in up to four –ordered- stages (Figure 3-1): (a) the analysis of evidence (e.g. evidence-based medicine), (b) the outcomes analysis (e.g. benefit-risk ratio), (c) the analyses of cost and cost-effectiveness, and (d) the analysis of ethical and legal implications of the technology (Eddy, 2009).

Figure 3-1. Relationship of evidence-based medicine and technology assessment



Source: (Eddy, 2009)

Evidence-based medicine is the basis of HTA. It is a way to synthesise and formulate in a clear way the best available evidence for a specific health problem based on a systematic and critical appraisal of this evidence (HTAglossary.net, 2015). This is then used to measure the benefit-risk ratio for the specific health problem of interest, typically comparing two interventions (e.g. the new treatment against standard practice). Including cost considerations allows for an estimation of the cost-effectiveness of these two treatment interventions (e.g. whether one is more cost-effective than the other). This is done using economic evaluation, which can take a number of forms. Therefore, HTA measures the incremental benefits and the incremental costs between the technology of interest with existing options, usually standard care used as a benchmark (HTAI Policy Forum Meeting, 2013), and is modelled using economic evaluation techniques to account for certain assumptions, such as extrapolating the known effects into the long-term. Deciding on whether the outcome of this economic model demonstrates adequate value for money for the technology can be deliberated while accounting for ethical and social considerations, where certain outcomes may be more or less acceptable given these contextual considerations.

The ordering of these four steps signifies that the second cannot be done without having accomplished the first step, and so forth. Further, not all HTAs are full HTAs, whereby some may stop at the assessment of incremental value (step two), others at the cost-effectiveness analysis (step three), and the most advanced will complete all four steps (Eddy, 2009). For example, the Haute Autorité de Santé (HAS) in France did not, at the time work on this thesis commenced, account for costs and cost-effectiveness but focused solely on incremental benefit translated into an ASMR

(Amélioration du Service Médical Rendu) rating. This allows for drugs across all therapy areas to be compared using this single measure which encompasses a range of attributes in order to decide on its coverage. Recently, it has implemented a requirement for economic evaluation for those drugs considered to provide additional clinical benefits (ASMR I-III) or costing more than 20 million euros to the health care system (HAS, 2012). In England, the National Institute for Health and Care Excellence (NICE) relies on cost-effectiveness and accounts for equality issues or end-of-life treatments. HTA bodies may use different attributes to valuing benefit (e.g. survival, health-related quality of life (HRQol). NICE, for example, requires that these health effects be expressed in QALYs ("quality adjusted life years"), which is a composite measure of improvement in length of life and in HRQol (NICE, 2013). This is not the case for the other HTA bodies.

Many different methods, both quantitative and qualitative, are being used to devise these measures of value on the basis of evidence-based medicine. The outcomes of interest for the health benefit include measures of therapeutic effect (e.g. effectiveness, safety) and/or of HRQol. Quantitative methods include, for example, the analysis of survival data or the calculation of average costs. In terms of qualitative methods, outcomes could be measured through interviews, focus groups, surveys, etc. and aim to elicit preferences, priorities or broader aspects not covered by quantitative methods (HTAI Policy Forum Meeting, 2013).

The "three-step HTA" can be summarised by the cost-effectiveness formula (Box 1). The ratio of incremental costs and incremental benefits demonstrates the incremental cost-effectiveness ratio (ICER), providing information about the extra cost to be paid for an extra unit of effect. This represents a measure of value of this new treatment compared to standard care. It is worth paying for (or is cost-effective) if it is within the decision-maker's maximum willingness-to-pay (WTP) threshold. The fourth step (e.g. analysis of ethical and legal implications of the technology) would be accounted for when deliberating about the ICER.

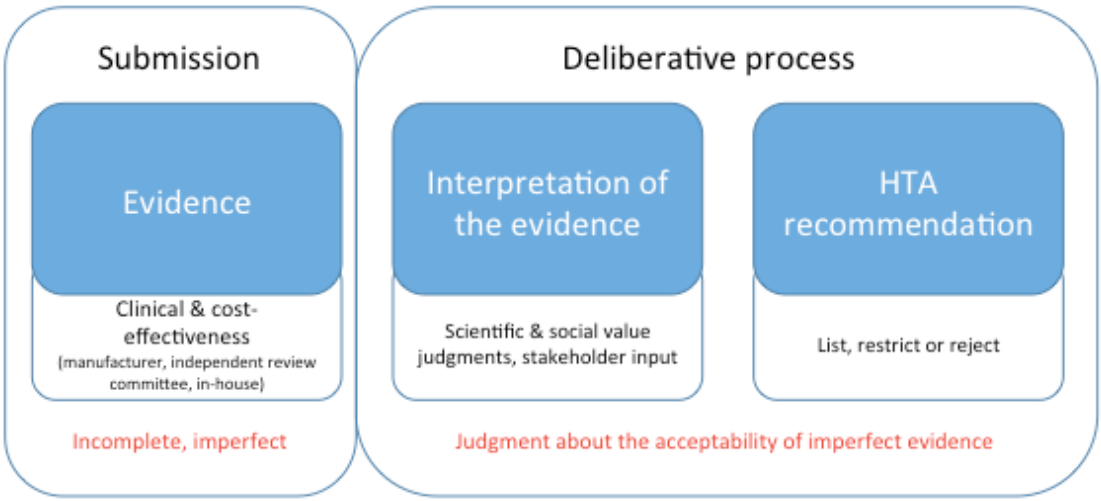
Box 1: The cost-effectiveness formula (≈ 3 step HTA)

$$\Delta cost/\Delta effect = ICER < WTP$$

Legend: Δcost = difference between the cost of a new treatment and its comparator; Δeffect = difference between the effect of a new treatment and its comparator; ICER = incremental cost effectiveness ratio; WTP = willingness-to-pay threshold.

Despite its fundamental role in value assessment, HTA is not without its controversies. In addition to the issues discussed previously around defining "value", these decision-making processes most often rely on incomplete or imperfect evidence, referred to as "uncertainty". Value judgments are being made about the acceptability of this uncertainty. Different methods exist to deal with uncertainty, as discussed in the next section, but the decision about its acceptability relies on the decision-maker Committee's own (scientific) judgment. These decisions may also be influenced by the Committee Members' social value judgments about considerations related to living with a disease and taking a course of treatment, which may also influence their judgment based on their own experience or on what they believe society would prefer. This is illustrated in Figure 3-2.

Figure 3-2. Visual model of the decision making process



Source: The author.

Generally, the uncertain estimates of clinical benefit and costs accounted for in the cost-effectiveness model result in uncertain ICER estimates. The acceptability of this uncertainty depends on the decision-makers' value judgments during the deliberative process. However, the higher and more uncertain, the more likely they are to lead to a negative coverage recommendation. In such cases, patients are being told they will not receive a treatment that could provide some benefit to them because it is not "cost-effective". This is often misunderstood and interpreted as putting a monetary value on life. Some decisions have led to strong reactions from different stakeholders and the media. For example, NICE's decision against the use of Arbiterone for advanced prostate cancer before chemotherapy was heavily criticised: "NICE decision on prostate cancer drug is a 'kick in the teeth' for patients" say UK Prostate Cancer (Prostate Cancer UK, 2014). Similarly was their decision to deny access to Herceptin to treat breast cancer due to a high cost per QALY: "It has given me back my normal life. You just can't put a price on it" (Orr, 2014).

Several important questions arise in light of these controversies around whether current HTA approaches sufficiently capture the attributes of value that matter most to patients, their carers and society. These include, for example, patient preferences (e.g. preference for certain side effects over others), considerations of additional attributes of value (e.g. innovation), or whether benefits beyond health gains are to be considered (e.g. productivity losses). There is no consensus around how these are accounted for across HTA bodies. These issues reflect the previous discussion about capturing value. The next section discusses some of these key issues pertaining to the use of HTA in assessing the value of a drug.

3.2. The HTA Framework and its Main Limitations

Despite the extensive and continuously increasing use of HTA as a decision-making tool, its methods and processes present significant challenges; these are often a consequence of issues around evidence generation, the outcomes to measure effectiveness, costs or cost-effectiveness, the extent to which willingness-to-pay thresholds (WTP) are used explicitly, and how they are interpreted. This section

portrays the main challenges and their implications on the final HTA recommendation, and focuses on each of the components of a "full HTA".

3.2.1. Evidence-based medicine & comparative effectiveness

HTA relies on evidence-based medicine to measure the comparative effectiveness of a set of treatment interventions in real world settings, and assess the clinical benefit of the treatment under consideration. This differs from the requirements for marketing authorisation, where the focus is solely on the treatment's efficacy, quality and safety compared to placebo in a controlled environment. At the time of an HTA, information about the treatment's effectiveness - in real world settings – is rarely available since HTA is often conducted soon after market authorisation (Figure 2-3). Therefore, HTA relies mainly on imperfect or incomplete – or uncertain - evidence from clinical trials conducted in controlled environments, which most likely do not capture the full effects of the treatment since the trial period is often inferior to the time horizon during which the treatment produces its effects (Sculpher, Claxton, Drummond, & McCabe, 2006). Uncertainty refers to the fact that we can never know for certain what the mean (expected) costs and effects would be if the treatment is provided for a particular population of patients, even if they have the same observed characteristics" (Claxton 2008). Decision-makers make scientific value judgments about the extent to which this uncertain evidence is acceptable. This includes judgments about whether the evidence presented captures the effect of the intervention, whether it is generalizable to the local context of the decision, whether quality of life changes are accurately captured, or whether it is appropriate to impose restrictions to population subgroups (Rawlins 2014). Assessing uncertainty is important in terms of using a correct estimate of costs and effects, of determining if existing evidence is sufficient, and to assess the consequences of an uncertain decision (Claxton 2008). Part of this uncertainty relates to the estimates of costs and effect used, where no two different individuals will have the same response to treatment or way of living the disease. It may be greater in some cases (e.g. cancer) or less in others (e.g. cardio-vascular), and the evidence produced will aim to capture this effect in the best way possible but will always have some degree of uncertainty.

NICE defines three types of uncertainty: structural, parameter and uncertainty around the choice of data sources (NICE 2009a): structural uncertainty relates to the assumptions made to construct a model (e.g. treatment pathway), parameter uncertainty relates to the mean values of the parameters considered (e.g. clinical endpoints, utilities), whereas uncertainty about the choice of data sources that provide the values for the key parameters (e.g. sufficient to capture the full effects of the treatment). The type of uncertainty discussed throughout this thesis relates to both parameter uncertainty and uncertainty about the sources of data, however they were not distinguished as such but similarly to how they were reported or discussed during the interviews.

Consequently, uncertainties are inevitable and making a decision failing to account for them can be misleading (Rawlins, 2008). Different statistical mechanisms exist to address this issue, such as probabilistic or deterministic sensitivity analysis (Claxton, 2008). The main question continues to be around appropriate methods to tackle these uncertainties, and whether the estimates or the methods used capture sufficient information about the expected benefits and costs. These decisions therefore rely on the decision-makers scientific and social value judgments about the acceptability of this imperfect or incomplete evidence and about certain treatment and disease considerations that may not have been captured in the evidence presented (Rawlins 2014).

The debate around uncertainty also opens further discussions about when requirements for the generation of additional evidence should be made to manufacturers after a positive or “only in research” HTA recommendation (e.g. marketing authorisation under exceptional circumstances) (Claxton, 2008). From a more ethical perspective, there are on-going debates about whether patients should be given early access to a treatment that is being evaluated within a clinical trial and shows positive interim results at the cost of producing lower quality evidence. For example, interim results for sorafenib (2nd line treatment for renal cell cancer) were positive and the decision was taken to unblind the trial participants and provide treatment to all. This resulted in uncertain evidence generation since the control group was unblinded before the end of

the trial, which impacted negatively on the reimbursement decision (Drummond et al., 2009a).

Many different types of studies contribute to demonstrating effectiveness (e.g. RCTs, observational studies). Systematic reviews or meta-analyses are often required for HTA submissions in order to account for all existing evidence and avoid bias in selecting only a subset of studies when considering relevant information for a decision problem. Direct comparisons are often considered the gold standard or preferred type of evidence for HTA bodies; though indirect comparisons or placebo-controlled trials are usually also accepted. Evidence suggests that direct comparisons also have their limitations (e.g. often carried out as open trials), and can be addressed by combining all levels of evidence (e.g. direct and indirect comparative evidence) (Lu & Ades, 2004; Madan, Stevenson, Cooper, Whyte, & Akehurst, 2011; Sutton, Ades, Cooper, & Abrams, 2008); though this may not be recognized across the board.

Hierarchy of evidence is common practice, where RCTs are considered the gold standard despite their limitations. For example, multi-drug interactions may not be identified in RCTs since patients with multiple morbidities are often excluded. Some argue that non-controlled studies enable to capture elements that are not identified in RCTs (e.g. less common adverse effects) and consequently all levels of evidence are crucial for the evaluation of a technology (Rawlins, 2008). Many questions about what constitutes evidence of sufficient quality remain, and as such HTA bodies may have different perceptions about the acceptable levels of evidence.

The estimates of effectiveness capture two aspects of the treatment: 1) how well it works (e.g. life years gained), and 2) what impact it has on the patient's HRQol. "HRQol includes physical and mental health perceptions (e.g. energy level, mood) and their correlates—including health risks and conditions, functional status, social support, and socioeconomic status" (CDC, 2016). Symptoms alleviation would be considered a HRQol improvement. A set of generic and disease-specific tools exist to measure the impact on HRQol, the most common and also often preferred method is the generic EQ-5D instrument enabling comparisons across disease areas. This is summarized in a measure referred to as "utility" to the patient. Current debate exists

around whether these HRQol measures really capture patient preferences and compare HRQol across disease areas accurately, considering that they are measured from healthy individuals (Ghislandi, Apolone, Garattini, & Ghislandi, 2002). In some countries, the preferred metric for expressing both the duration and health-related quality of life is the QALY ("quality adjusted life years"), where one QALY represents one year of life gained in full health. In other countries, the clinical effect is considered as a hard endpoint and HRQol as a soft endpoint. Consequently and similarly to the above, many different methods, each with their own recognised advantages and disadvantages, exist to measure HRQol and effectiveness, and in the key question for HTA remains whether they accurately reflect the impact of a disease or treatment on a patient and their carers (Brazier, 2008).

In summary, differences are seen in the way common HTA methods to estimate the treatment's benefit are applied (e.g. QALY or clinical benefit of the treatment), which are likely to result in differing HTA outcomes. Indeed, methods can take many different forms, and there is no clear understanding about what constitutes evidence of sufficient quality. These decisions therefore rely on the evidence presented and methods used for its interpretation, as well as the decision makers' judgments about the acceptability and plausibility of this evidence made during the deliberative process and influenced by their own experience or what they believe society would prefer, as well as additional contextual considerations (e.g. political context).

3.2.2. Costs (direct and indirect) and cost-effectiveness

A similar scenario is seen for costs and cost-effectiveness. Indeed, HTA often includes the economic consequences of different treatment alternatives ("3-step HTA"), referred to as comparative costs (Chalkidou et al., 2009). The costs collected are either direct medical or non-medical costs (e.g. hospitalisation or carer services, respectively), or indirect costs to the health care system (e.g. productivity losses or informal carers). Generally, costs vary depending on the context. The cost of GP visits or hospitalization differs in different settings. The perspective adopted will also determine which costs are included in the analysis. For example, direct costs to the NHS and Personal Social Services are considered in England and Wales and in only

exceptional cases will costs outside this scope be considered (Sculpher, 2008). In contrast, TLV in Sweden adopts a societal approach and considers both direct and indirect costs. Cost can be measured using different methodological approaches from average per diem costs to micro-costing. Consequently, cost estimates are also likely to differ across countries depending on the approach used (Hughes, Tilson, & Drummond, 2009).

Once the costs and benefits of a new treatment and its comparator(s) are estimated, the next step is to determine whether the new treatment provides value for money; this is often done through an economic evaluation. Economic evaluation uses decision analysis to model the different treatment paths from taking this new treatment or its alternative comparator, and account for the different events occurring (e.g. full health, death) and their probability of occurring. These models allow for a number of assumptions to be made, such as to extrapolate short-term costs and effects over the long run, or under a set of assumptions (e.g. treatment is taken for 2 years, natural risk of death within different age groups is accounted for, probability of the event depends on the health state, etc.).

Many different methods of economic evaluation exist, the most common being: cost-utility analysis (CUA), which compares QALYs and costs of alternative treatments; cost-effectiveness analysis, which compares health outcomes (e.g. life years gained) and costs of alternative treatments, and cost-minimisation analysis, which compares the cost associated with two "identical" treatment alternatives. Each HTA body has different preferences and requirements for economic evaluation modelling, which may also be one of the explanatory factors for different HTA outcomes, discussed later. Results from economic modelling are the incremental cost-effectiveness ratio (ICER), which provides information about the cost per extra unit of effect. Whether or not it is cost-effective depends on whether it is within the limits of what is considered to provide value for money.

3.2.3. Willingness-to-pay and "Other considerations"

The willingness-to-pay (WTP) is the maximum amount a payer is willing to pay for an extra unit of health benefit, which is specific to each country depending on national preferences and affordability. High ICER estimates, which represent a high cost per unit of health benefit, are more likely to be rejected than lower ICER estimates. Evidence shows that the WTP threshold ranges between £20,000 and £30,000 per QALY in England and Wales, and increases with disease severity and need in Sweden (McCabe, Claxton, & Culyer, 2008; Pearson & Rawlins, 2005; Webb, 2009). When a treatment's ICER is less than the threshold, it is expected to receive a positive reimbursement recommendation, and when higher, it is expected to receive a negative recommendation. Despite this, evidence suggests that decisions are not always consistent with WTP thresholds and may vary depending on the disease characteristics (e.g. disease severity) and level of progression, as well as on the treatment's characteristics (e.g. curative or symptomatic) (Dolan, Shaw, Tsuchiya, & Williams, 2005). This may suggest that "other considerations" beyond ICER estimates are likely to have an impact on the final HTA recommendation.

This is illustrated in a number of studies that identified variable WTP levels depending on the therapy areas being appraised. Dupont and van Wilder compared the Belgian HTA recommendations for a sample of orphan and non-orphan drugs considered equally severe and innovative. Orphan drugs received a higher rate of positive recommendations despite the less robust nature of the evidence considered for orphan conditions (Dupont & Van Wilder, 2011). Another study showed that 43% of the orphan drugs appraised in Scotland had an ICER greater than the £30,000 informal WTP threshold and almost all were rejected for reimbursement; suggesting that more flexibility in accepting high and uncertain ICERS may not be given to orphan drugs (Vegter, Rozenbaum, Postema, Tolley, & Postma, 2010). Simoens and colleagues (2011) demonstrated that SMC and NICE have higher WTP levels for cancer drugs (including orphan indications) than for other disease areas (Simoens & Dooms, 2011).

Such disparities in accepting different levels of ICER may reflect social value judgments or preferences, where the decision makers prioritise certain treatments because of considerations around living with the disease or taking a treatment. Such social values may either be elicited from a representative sample of society (e.g.

disease severity in Sweden, where it is recognised that higher ICERs are accepted for patients with high disease severity), or non-elicited and relies on the personal judgment of the decision-maker. There is a need to better understand the value judgments made throughout these processes and the weight they have on decisions. The lack of consistency and transparency in accounting for these "other considerations" were highlighted by Earnshaw and colleagues (Earnshaw & Lewis, 2008).

3.3. Factors that distinguish orphan drugs from other drugs

Rare diseases affect small patient numbers, less than five in 10,000 in Europe and less than 200,000 people in the US. There are currently between 5,000 and 8,000 rare diseases; put together they affect six to eight per cent of the European population. These conditions are often life-threatening, debilitating and frequently genetic (European Commission, 2015b).

In the past, R&D for the development of treatments against rare diseases was not sufficiently attractive, given the modest returns from a small number of patients. Based on the principle of equality, where all patients are entitled the same quality of care, incentives were implemented at the marketing authorisation level to stimulate research and development (European Union, 1999). According to recent regulations, all medicinal products treating a rare disease are eligible to receive an orphan designation and benefit from these incentives. In Europe, incentives are granted at the marketing authorisation level by the EMA and include free regulatory advice for clinical development, accelerated marketing authorization review procedures, a 50% fee reduction on the regulatory procedure and marketing exclusivity for ten years after marketing approval. Orphan recognition and similar incentives are implemented in the US, Japan, Canada, and Australia. Another characteristic is the willingness to grant early access to these treatments through the exceptional circumstances or conditional approval processes, depending on the Member State and under specific circumstances (e.g. treatment of a serious or life-threatening disease, no alternative treatment available, treatment being tested in clinical trials and in an active phase of marketing approval) (Orphanet, 2015). These incentives have successfully stimulated the development of orphan drugs, as demonstrated by the increasing number of medicinal

products receiving orphan designations over the past decade compared to before (Kesselheim, 2011; US Food and Drug Administration, 2015).

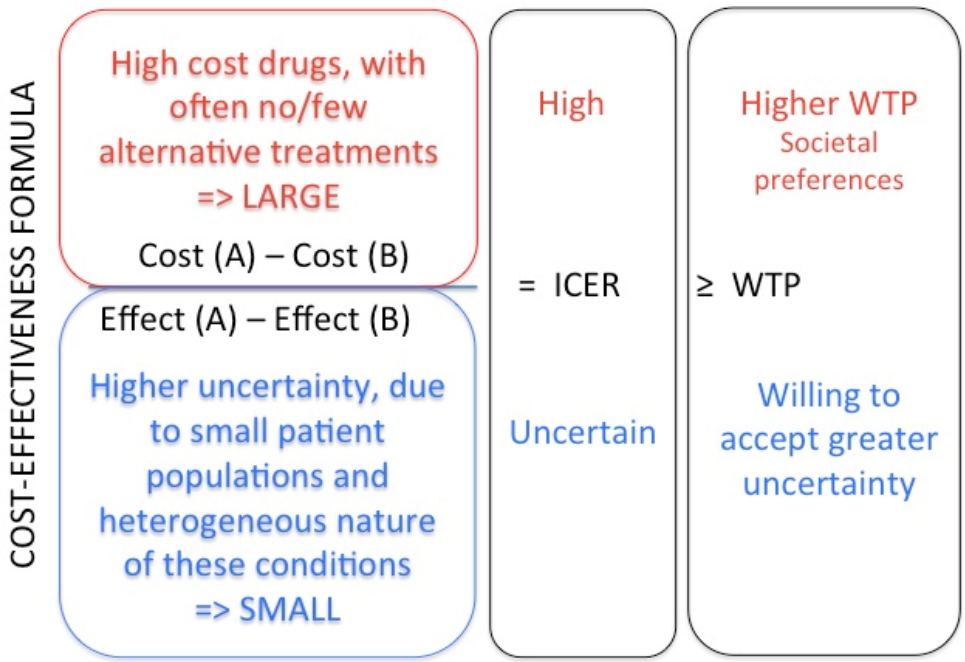
Despite the unique nature of marketing authorization procedures for orphan drugs, HTA processes are similar to other drugs. Once these orphan drugs are granted the authorisation to be marketed, the MAH seeks coverage in each individual country in the same manner as those drugs for more prevalent conditions. Given the substantial uptake of HTA in Europe and around the world, there is a high probability that the decision is evidence-based and informed by HTA.

However, unlike other (non-rare) conditions, challenges exist in producing robust evidence for rare conditions. Rarity implies that there is a lack of knowledge about the natural history of the disease, and a scarcity of scientific literature and of experienced clinical investigators around the world. This is even more accentuated for those diseases with no existing treatments, where there is no knowledge about the clinical development path (Vickers, 2013). The lack of knowledge about these diseases and their heterogeneous nature creates issues in designing, recruiting and conducting appropriate clinical trials. Issues around trial design pertain to selecting the appropriate endpoints, treatment pathway(s) and trial duration; during the trial period, there may be difficulties in recruiting sufficient patient numbers from small patient populations and in generating meaningful trial results from scarce clinical experts; trials are often conducted on multi-sites requiring high fixed costs for each site and delays (Kesselheim, Myers, & Avorn, 2011; Vickers, 2013). Additionally, trials for rare diseases are often smaller than those for more common conditions (Kesselheim et al., 2011), whereby treatment effects should be much greater to attain statistical significance, which may not be representative of the added benefit of some of these treatments for very rare conditions (Boudes, 2013). This was shown in one study that compared the evidence for orphan and non-orphan cancer drugs, where trials for rare cancers were less likely to be randomized and double-blinded, and more likely to assess a surrogate outcome rather than a hard endpoint compared to trials for more common cancers (Kesselheim et al., 2011). A number of innovative trial designs exist to deal with small patient trial populations, which minimize the number of patients included in the trial (e.g. with longer trials or innovative trial designs such as factorial

designs), or maximize on-treatment participants (e.g. N of 1 cross-over trial design) (Gagne, Thompson, O'Keefe, & Kesselheim, 2014), but their use is not often yet seen in practice (Kesselheim et al., 2011).

This has implications for HTA, as the evidence for a treatment for a rare disease is likely to be subject to a high degree of uncertainty compared to the more common conditions; additionally, the price of orphan drugs are usually very high in order to recoup R&D investments from small patient populations. As a consequence, an orphan drug is hardly cost-effective (Clarke, 2006; Denis, Mergaert, Fostier, Cleemput, & Simoens, 2010a; Drummond, Wilson, Kanavos, Ubel, & Rovira, 2007), as illustrated in Figure 3-3. The main question therefore relates to whether we are willing-to-pay more for these conditions or whether we are willing to accept greater uncertainty in the evidence. This is further discussed in the next section.

Figure 3-3. HTA and cost-effectiveness



Source: The author.

3.4. HTA in Different Settings

Even if the reasons for using HTA are similar across countries, David Eddy argues that differences across countries are inevitable because of the methodological approaches

used and the stages of the HTA carried out, often relying on the social and political circumstances in that particular setting (Eddy, 2009). This was further emphasised in the last section, which highlighted the variety of ways to conduct HTA, even when similar methodological approaches are being used. This section discusses evidence that compared HTA recommendations across countries and identified differences (Section 3.4.1), including cases when they related to decisions for orphan drugs (Section 3.4.2).

3.4.1. Differences in HTA coverage recommendations across countries

Cross-national differences in the HTA coverage recommendations made exist and are generally recognised, mainly because of the complexity of these processes and the context in which they operate (Banta, 2003). Each country sets its own objectives for conducting HTA reflecting its values, preferences (e.g. population disease profile) and constraints (e.g. budget constraints, structure of the health care system), and consequently differences across countries in the HTA outcomes are inevitable (Banta, 2003; Busse, Orvain, Velasco, Perleth, Drummond, Gurtner, Jorgensen, Jovell, Malone, Ruther, & Wild, 2002). This is further emphasized in a recent systematic review of the literature that aimed develop a framework describing and comparing features of HTA bodies. Substantial differences in the stages of an HTA process were identified: only 38% of the scope of HTA were similar across the four study countries (Germany, France, England, Sweden), 26% in the process, 29% in the methods, 40% in the dissemination, and in only 19% of the decisions (Schwarzer & Siebert, 2009).

Variations across settings may be a consequence of national considerations, reflecting different priorities or preferences, but also a consequence of differences in HTA processes or other factors considered in the appraisals, possibly also reflecting weaknesses in HTA methods. In the latter case, implications are that value for money may not be obtained for some of the treatments covered or that access to these treatments across countries may also differ, especially considering that HTA recommendations have shown to be fairly consistent with the final reimbursement decision (McMahon, Morgan, & Mitton, 2006; Wonder, Neville, & Parsons, 2006).

As a result, the need to understand the reasons for these differences and increase transparency is recognized. A number of initiatives exist to increase collaboration and establish methods for the harmonization of methodological approaches to HTA (e.g. EUnetHTA, Pharmaceutical Forum). The European Network for Health Technology Assessment (EUnetHTA), for example, aims to improve HTA methods and avoid duplication of efforts. It has developed an HTA framework (the "HTA Core Model") that enables the sharing of information across countries, and is piloting EU-level HTAs. It has also developed a common methods for HTA through collaborative work between Member States, focusing on the clinical effectiveness of a treatment (EUnetHTA, 2015). This research complements such initiatives in generating a better understanding of the causes for variation and areas where HTA methods may be improved and potentially result in a common agreement about their appropriateness.

At the time this thesis began, eight studies comparing HTA recommendations across more than one country were identified. They identified important variations (Table 3-1) (Clement et al., 2009; Kanavos et al., 2010b; Lexchin & Mintzes, 2008; Morgan et al., 2006; Nicod, 2010; Pomedli, 2010; Shah, Mestre-Ferrandiz, Towse, & Smyth, 2013; Van den Aardweg, 2010). Their research designs were generally in the form of retrospective descriptive or cohort analyses. The level of agreement in the HTA recommendations issued across countries was measured in three of these studies and ranged between poor and moderate, further emphasising the extent of these differences (Clement et al., 2009; Lexchin et al., 2008). Countries included in the comparisons were Canada, Australia, England, Scotland, France, and New Zealand.

Morgan and colleagues compared coverage recommendations for the seventeen top selling drugs in four countries with a centralised drug review process (England, Australia, Canada, New Zealand), and evaluated their impact on use and spending. They showed that listing decisions differed substantially and may indicate differences in appraising and interpreting the evidence, in the alternatives covered, or in price negotiations (Morgan et al., 2006). The manner in which these potential explanations for differences were identified was not described nor did it appear to have been explored systematically. The main objective of the research was to discuss challenges relating to the centralised drug review process rather than identify cross-national

differences in a systematic manner, and – through these - derive recommendations so that these processes become more transparent and rigorous.

Lexchin and colleagues also conducted a retrospective analysis of all HTA recommendations issued in Canada, Australia and Scotland until September 2006. They compared the HTA recommendations made for 51 drugs and analysed the clinical and pharmacoeconomic assessments made for individual drugs with diverging recommendations. Results showed poor to moderate agreement in the recommendations issued across agencies ($0.19 < k < 0.44$), and differences in the clinical and pharmacoeconomic assessments in nine cases with diverging recommendations (Lexchin et al., 2008). While the reasons for differences were explored in more depth compared to previous studies, they were considered meaningful and described in only nine cases, which potentially suggests that the depth of the analysis was insufficient to systematically capture the reasons for differences. This may be due to the nature of the comparison that focused on the clinical and pharmacoeconomic assessment, which may not be sufficient to capture the subtleties of how different agencies are dealing with uncertainty and how their concerns are dealt with across settings (e.g. means through a concern was deemed acceptable). This would allow for a more extensive understanding of the application of HTA across settings, which could contribute to furthering the debate around how HTA bodies are dealing with uncertainty resulting from limitations in the HTA approach (which is one of the objectives of this thesis).

Clement and colleagues conducted a retrospective analysis of all HTA recommendations issued in Canada, Australia, and England and Wales until 2008. Similar to the previous study, poor to moderate agreement ($0.13 < \kappa < 0.55$) in the recommendations issued across agencies was also demonstrated. They also identified key issues known in these assessments, such as clinical and economic uncertainty, and to what extent such characteristics were identified in each of the study countries. For example, clinical uncertainty was present in more than 40% of the submissions, but was lower for NICE reflecting its willingness to narrow down the indication to more effective and cost-effective niche groups of patients. Three case studies were selected to illustrate some of the differences and similarities identified across countries. They

concluded that listing variations were more likely to be a consequence of HTA processes than of the subjective interpretation of the evidence by each agency. The study limitations recognise that differences exist in the way HTA bodies use clinical and/or cost-effectiveness, but that more research is needed to understand the reasons for these differences (Clement et al., 2009). While this study identified a number of key issues, and illustrated similarities and differences through three case studies, they recognise that further research is needed to identify the reasons for these cross-national differences. This would require a more systematic and in depth approach in the analysis made across a greater sample of compounds.

A more recent study by Shah compared HTA decisions for seventy-six decisions in two oncologic areas: breast and colorectal cancer, and highlight important variations (Shah et al., 2013). They then conducted a thorough review of those drugs that received diverging recommendations in order to understand the reasons for differences. They identified three key recurring issues around surrogate endpoints, patient voice and comparator selection. They also recognise other aspects influencing these processes, such as the different ways agencies are using the available evidence or dealing with uncertainty, but these were not further explored (Shah et al., 2013). While a more thorough and in depth analysis was undertaken compared to the previously reported studies, the level of detail in the methodological approach used (thematic analysis) was not sufficient for its transferability, nor was it clear to what extent the analysis was systematic, encompassing all factors captured in the decision reports influencing these decisions (including during the deliberative process), or allows for cross-country comparisons. The study also lists a number of potential reasons explaining differences. A more systematic and in-depth approach could constitute a stronger case to understand more comprehensively the limitations in the application of HTA and how these are dealt with across settings.

Additionally, there were a number of studies that investigated the drivers of HTA decisions in a more systematic manner by focusing on one specific country (Devlin & Parker, 2004; Dakin, Devlin, & Odeyemi, 2006). These have not been covered in the literature review but are reflected upon in the Conclusion chapter together with any additional comparative studies published until 2016.

Table 3-1. Summary of existing evidence focusing on cross-national comparisons of HTA recommendations

Author Year	Objectives	Method	Countries	Main findings	Reasons for cross-country differences in the HTA recommendations	How these studies differ from this thesis
Morgan 2006	To explore features of the centralised review processes across countries, and compare the HTA processes and recommendations, and the impact of the decisions on cost and use	Comparison of 17 top selling drugs (2003) and reimbursement processes in the four countries	Australia, Canada, New Zealand, UK	Different HTA processes, and HTA recommendations across countries	Differences in: - what is appraised; - interpretation of the evidence; - coverage of therapeutic alternatives; - negotiation with suppliers	The approach used to identify the reasons for different HTA recommendations was not described. The differences were presented as potential reasons. They are considered not to have been examined systematically. This is because the main objective was to discuss challenges relating to the centralised drug review process rather than systematically examining cross-national differences
Lexchin 2008	To compare HTA recommendations across countries	Retrospective cohort analysis. To compare HTA recommendations across countries and explore the reasons when these were different	Canada, Scotland, Australia	Different recommendations: PBAC-CDR: moderate agreement CDR-SMC: poor agreement (Kappa score $0.19<k<0.44$)	Potential reasons for differences in: - proposed price and effectiveness of competing products in the national markets; - hospitalisation and physician visit costs; other considerations: disease prevalence and severity, perceived need, composition of the panel making the recommendation, the scientific rigor and relevance of the evidence for comparative safety and effectiveness	The reasons for differences were explored in more depth compared to previous studies, however, they were considered meaningful and described for only nine cases. This may suggest that the depth of the analysis was insufficient to systematically capture the reasons for differences. This may be due to the nature of the comparison that focused on the clinical and pharmacoeconomic assessment, which may not be sufficient to capture the subtleties of how different

						agencies are dealing with uncertainty and how their concerns are dealt with across settings (e.g. means through a concern was deemed acceptable)
Clement 2009	To describe how clinical and cost-effectiveness are used in coverage decisions, and identify common issues in the study countries	Descriptive analysis of retrospective data	England and Wales, Australia, Canada	Different rates of positive recommendations: poor to moderate agreement (Kappa score 0.13<k<0.55)	Most common causes of variation: - narrowing down of indication (NICE); - price negotiations (PBAC); - different attitude towards drugs with little or no therapeutic benefit (CDR and PBAC); - cost-effectiveness data; - approaches to assessing low quality evidence (CDR, PBAC, NICE)	While this study identified a number of key issues, and illustrated similarities and differences through three case studies, they recognise that further research is needed to identify the reasons for these cross-national differences. This would require a more systematic and in depth approach in the analysis made across a greater sample of compounds
Shah 2013	The aims are to identify key drivers of decisions and to understand the similarities and differences in the requirements of different agencies	Qualitative thematic analysis	Australia, Canada, England and Wales, Scotland, France	Different recommendations across countries for cancer drugs, and reasons for these	Main causes of variation : - interpretation of clinical endpoints ; - differing levels of patient input ; - issues around appropriate comparators. Other aspects listed but not explored in depth (e.g. ways agencies are using the available evidence or dealing with uncertainty)	While a more thorough and in depth analysis was undertaken compared to the previously reported studies, the level of detail in the methodological approach used (thematic analysis) was not sufficient for its transferability, nor was it clear to what extent the analysis was systematic, encompassing all factors influencing these decisions (including during the deliberative process), or allows for cross-country comparisons

Sources: (Clement et al., 2009; Lexchin et al., 2008; Morgan et al., 2006; Shah et al., 2013)

In summary, inter-country variability in HTA recommendations are well known to scientists preoccupied with comparative research because they may reflect weaknesses in the application of HTA. This problem, together with its implications, has been identified and possible explanations examined. The above mentioned studies all have in common that they highlight the extent of these differences, by comparing the HTA coverage recommendations across a sample of drugs and countries. They differ in that some of the countries included were not common, nor were the therapy areas being compared.

The reasons for cross-national differences were also explored, but with varying levels of thoroughness. Morgan and colleagues focus more on the process, transparency and rigour of these processes, rather than on case-specific reasons for diverging recommendations (Morgan et al., 2006). In contrast, the three other studies investigate these differences in a more thorough manner (Clement et al., 2009; Lexchin et al., 2008; Shah et al., 2013). Nevertheless, the reasons, or similarities and differences, across countries relied on a few cases or potential reasons, and may not encompass the full picture and subtleties of these processes. First, no clear picture of the key determinants and of the structure of the decision-making problem is outlined in any of these studies, where consequently, the reasons set forth may not constitute the full picture. Second, issues relating to the clinical and pharmacoeconomic assessments also referred to as clinical and economic uncertainty or key issues, were identified. However, the level of detail provided did not differentiate the type of concern raised (e.g. trial duration or uncertain magnitude of the benefit?), how these were dealt with across countries (e.g. acceptable by some and not by others?), and what were the factors influencing these processes (e.g. “other considerations” accounted for during the deliberative process?). Third, the methodological approaches used, even if thorough, were not sufficiently detailed for these approaches to be transferable. One exception may be the paper from Lexchin and colleagues, which does describe in detail how the variables were categorised (Lexchin et al., 2008). This was accounted for when setting up the coding scheme in this thesis. Given that these decision processes are complex and understanding what happened for one same drug in different countries may be challenging, a more systematic, structured, and comprehensive approach to identifying and comparing differences would be required.

Additionally, understanding how similar scenarios are dealt with across settings may also constitute a way forward to identifying some of the limitations in the applications of HTA and for cross-country learning about how these were dealt with across settings.

In 2010, a series of papers was published in the EuroObserver, which looked at the impact of HTA on an international level (Kanavos et al., 2010b; Nicod, 2010; Pomedli, 2010; Van den Aardweg, 2010). Having been part of this project and collaborated with the three co-authors, this project formed the motivation for this thesis, which resulted in the recognition about the need to further explore the differences across countries. The database compiled during this preliminary stage was further leveraged and developed, and contributed to the first empirical paper of this thesis (Chapter 6).

3.4.2. Differences in access to orphan drugs

The literature focusing on orphan drugs is more limited in scope and only a few studies evaluated the level of variability in access to orphan drugs across different countries (Blankart, Stargardt, & Schreyogg, 2011; Kanters, Hakkaart, Rutten-van Molken, & Redekop, 2015; Michel & Toumi, 2012; Stolk, Heemstra, Leufkens, Bloechl-Daum, & Heerdink, 2009). Two studies identified such differences, where disease and treatment characteristics have shown to play a significant role (Denis et al., 2010a; Vegter et al., 2010). Indeed, rarity translates into difficulties to produce evidence of sufficient quality because of challenges in recruiting an adequate number of patients into RCTs (Clarke, 2006; Denis et al., 2010a; Denis, Mergaert, Fostier, Cleemput, & Simoens, 2010b; Drummond et al., 2007; Dupont et al., 2011; Joppi, Bertele, & Garattini, 2009; Kanavos & Nicod, 2012; Simoens et al., 2011; Vegter et al., 2010), together with the nature of the disease that can pursue highly variable clinical courses (Clarke, 2006), as discussed in section 3.3.

Different studies demonstrate that orphan drugs are more likely to receive the same or a higher level of acceptance for reimbursement compared to other, more common disease areas (Dupont et al., 2011; Simoens et al., 2011; Stolk et al., 2009). This suggests that coverage decisions for orphan drugs are based on the decision-makers' willingness to accept high and uncertain cost-effective drugs while taking into account

other considerations related to the disease and treatment characteristics, such as disease severity or the availability of treatment alternatives.

There is ongoing debate as to whether greater flexibility should be given to orphan drugs compared to drugs for more common diseases, based on the principle of equality and on the fact that these drugs by nature are likely not cost-effective (Drummond et al., 2007; McCabe, 2010; McCabe, Stafinski, & Menon, 2010; McCabe, Tsuchiya, Claxton, & Raftery, 2006, 2007; Simoens, 2011; Simoens et al., 2011). Both positive and negative arguments are put forward, and both parties agree that societal preferences (about whether or not society is willing to pay more for rarity) need to be elicited. The arguments for having greater flexibility for orphan drugs are the following: recoupment of R&D costs, equity in access, lack of existing treatment alternatives, potentially catastrophic levels of out-of-pocket costs to the patient, and societal preferences. The latter is based on the assumption that society puts more value on rare diseases due to the severe, life-threatening and disabling nature of these diseases.

Currently, with the exception of SMC in Scotland and HAS in France, HTA bodies in other countries do not make a differentiation in their processes for orphan drugs (Garau & Mestre-Ferrandiz, 2009a; Panju & Bell, 2010). The SMC accounts for additional treatment and disease characteristics referred to as the "SMC modifiers" (e.g. life-threatening, curative, life-extending, quality of life, no alternative treatments). In France, orphan drugs and rare diseases are recognised as a national priority under the 2004 Public Health Act (Ministere des Affaires Sociales et de Sante, 2004), though it is yet unclear how and whether this has any influence in the HTA coverage recommendation. Orphan drugs are generally subject to the same HTA evaluation processes, with some exceptions where they may receive preferential status under certain conditions, possibly explaining some of the variation between countries.

4. Research Questions and Plan of the PhD

4.1. Gaps in the Literature and Hypotheses

It is widely recognized that a coverage recommendation issued for the same drug and indication by several HTA bodies may differ, for many reasons as stated earlier. Eight studies identified the extent of these differences and possible reasons for having issued different HTA recommendations (Table 3-1) (Clement et al., 2009; Kanavos et al., 2010b; Lexchin et al., 2008; Morgan et al., 2006; Nicod, 2010; Pomedli, 2010; Shah et al., 2013; Van den Aardweg, 2010). No attempt, however, has been made to scrutinize these variations and query why they occur in a systematic manner. More specifically, the extent to which variations in coverage recommendations across agencies are due to the different type of evidence considered and methods used, or the manner in which the evidence was interpreted (e.g. how uncertainty was dealt with) are not clear.

In addition, there is a gap in the literature between what we know about these inter-country differences, how they differ across settings, and how this information can be useful to generate a better understanding of, and increase transparency in HTA processes, as well as improve HTA methods, and further down the line, improve access the clinically cost-effective treatments.

The key issue when determining whether a technology provides value for money is that many different approaches to appraise a technology exist and little is known about HTA body expectations in these “grey areas” (e.g. the appropriate methods to be used, the weight “other considerations” have in the final decision, or the value judgments that are made as part of the deliberative process).

Based on the above and considering attempts at approximating HTA structures in Europe, e.g. through joint actions such as the EUNetHTA, it is necessary to better understand and identify: (a) the evidence and methods used in HTA appraisals; (b) whether the evidence and methods were deemed acceptable or not by the evaluators when uncertain or incomplete, e.g. scientific value judgments; (c) the reasons behind the final recommendation; (d) whether “unacceptable” cases could have been

avoidable; and (e) to what extent “other considerations” influenced the final decision, e.g. social value judgments.

There is a need for more clarity and transparency about the appropriate methods to be used in the different contexts and the acceptability criteria for uncertain evidence. This would also include how disease or treatment characteristics are taken into account in the HTA processes, their weight on the decisions and whether they are accounted for consistently across cases (Earnshaw et al., 2008). This would ensure that the decisions being made are fair, robust and transparent, and that health care resources are being used efficiently and deliver the best outcomes for society.

4.2. Hypotheses

In this context, this thesis tests a number of hypotheses:

- ✓ Given that there are differences in HTA recommendations across countries, these differences are important and where they occur, they have substantial consequences for patient access to health technologies;

- ✓ These differences may be a consequence of:
    - o context-specific considerations, such as different willingness-to-pay thresholds, costs or national priorities;
    - o the methodological approach used, whereby different evidence may be appraised alongside different methodological approaches;
    - o agency-specific risk preferences, which consists of those risks or types of uncertainty that are more of a concern for certain decision-makers compared to others, resulting in different levels in accepting uncertain evidence;
    - o potential subjectivity in the interpretation of evidence in terms of what constitutes acceptable levels of uncertainty, accounting for additional evidence or not;
    - o social value preferences, which consist of “other considerations” beyond clinical benefit and cost-effectiveness (e.g. disease severity), that are

relatively more valued by some decision-makers compared to others and therefore carry more weight in the decisions in the former than in the latter;

- ✓ Value judgments are routinely made to inform decisions as part of the deliberative process of HTA, during which experts and key stakeholders are consulted and the evidence is discussed until a decision is taken. These value judgments originate from the individual decision-makers based on their experience and expectation about the preferences of society, including patients.

- ✓ Scientific and social value judgments that are routinely made to inform decisions are not used consistently within a setting across drugs, but should be. These judgments are not used consistently, but because of their importance, they should be. These are identifiable and could be accounted for more consistently.

- ✓ Decision-makers are willing to accept a higher ICER or uncertain clinical benefit in the case of orphan drugs to treat rare diseases by accounting for other considerations beyond routine HTA outcomes.

- ✓ Issues related to the rarity of conditions (e.g. trial design and conduct) are reflected in the HTA submissions in the quality and level of the evidence presented, and the concerns that arise from the assessors. These issues are not dealt with consistently and depend on the judgment of the decision-making committee members.

This thesis focused on HTA coverage decisions for branded prescription-only orphan medicinal products in four European countries: England and Wales, Sweden, Scotland and France[1]. More information about the structure and processes of these four HTA bodies and the reasons for their selection are provided in Chapter 5.

[1] HTA bodies issuing the HTA recommendations are the following:
- In England : the National Institute for Health and Care Excellence (NICE)
- In France : the Transparency Committee of the Haute Autorité de Santé (HAS)
- In Scotland : the Scottish Medicines Consortium (SMC)

4.3. Research questions

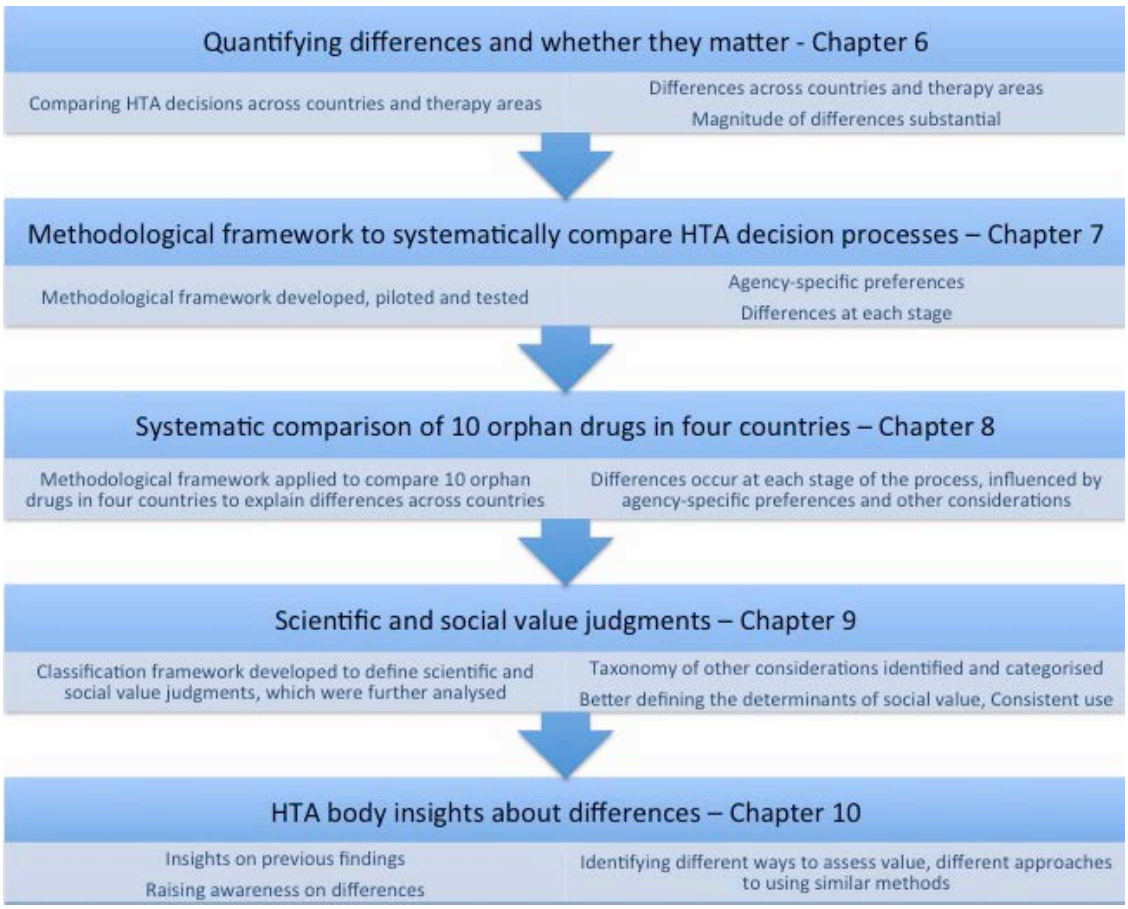
The above gaps in the literature identified emphasise the need to understand, in a more systematic manner, the reasons for differences in HTA recommendations across countries, which formed the main research question of this thesis, as follows:

-Main research question-

Why are there differences in HTA recommendations across countries for a same drug and indication pair? Is there consistency within and across countries?

This question was addressed in five inter-connected steps, which correspond to the five publishable papers that constitute this thesis (Figure 4-1).

Figure 4-1. Thesis structure



Source: The author.

4.3.1. Paper 1

The first paper (Chapter 6, Paper 1) further explores differences in HTA coverage recommendations. In addition to measuring the magnitude of these differences and potential reasons, the reasons and implications to understanding these differences were also emphasised, further contributing to existing literature.

What is more, it aimed to identify the scope and extent of differences across five countries (Australia, Canada, England, Scotland, Sweden), and query whether similar patterns across therapy areas are seen in these countries. The countries included were not the same as those analysed throughout the remainder of the thesis; results are nevertheless considered to be consistent and generalizable with the trends seen in previous studies that examined the same countries (Clement et al., 2009; Lexchin et al., 2008; Morgan et al., 2006; Shah et al., 2013). The research question was the following:

-Sub-research question 1-

What are the commonalities and differences in HTA recommendations made across five countries and three therapy areas? On this basis, are these differences meaningful?

Objectives were two-fold: 1) to examine the commonalities and differences in the HTA recommendations issued by the study countries and across three therapy areas between 2007 and 2009, and 2) to identify possible reasons for differences through a number of case studies in order to understand whether it is worth further exploring. A comparative analysis of HTA recommendations for 287 drug-indication pairs appraised was undertaken, as well as an in-depth analysis of two case studies.

Results showed significant inter-country variability in the HTA recommendations, where 46% of the 122 drug-indication pairs appraised by at least two agencies received diverging recommendations across countries. Agreement levels in the recommendations issued amongst pairs of countries measured using Cohen's kappa scores were poor to moderate, confirming the trend seen in previous research. The

contribution of this study compared to the existing evidence is the comparison undertaken across therapy area, where the relative proportions of positive and negative recommendations for each therapy area would be assumed to be similar across countries. Results showed the contrary, where associations between HTA recommendations issued by each HTA body per therapy area (cancer, orphan, CNS) differed from the general pattern observed across the complete sample. These results suggest that expectations from HTA bodies in terms of relative effectiveness differ depending on the drug and disease characteristics, although agency-specific guidelines are homogeneous for all treatments. Findings from this first chapter provided the rationale for the remainder of the thesis, emphasising that these differences are important and matter, and should be further explored.

4.3.2. Paper 2

In order to analyse these decisions in a more systematic, comprehensive and comparative manner, a methodological framework was developed and piloted (Chapter 7, Paper 2). In this respect, the research question for this second component of this thesis was the following:

-Sub-research question 2-

How can we identify and compare in a systematic manner similarities and differences in HTA recommendations? Developing and testing a methodological framework.

This framework was developed using a mixed methods research design that enabled to capture the depth and complexity of these decision processes, while making them comparable across countries and drugs, and quantitatively analysable. The framework enabled to examine the different stages of the HTA process across the countries for each drug and indication included, in order to identify when causes of variations were a consequence of national considerations or of differences in the methods used, the interpretation of the evidence, or the influence of the “other considerations”. This approach was used to derive this instrument-based model referred to as the methodological framework, which was used as a basis for the analyses in the

remainder of the PhD. Care was given to detail each step of the process undertaken to develop this framework to allow its transferability to third parties.

4.3.3. Paper 3

Once piloted, this framework was applied to a larger sample of orphan drug-indication pairs (Chapter 8, Paper 3), in order to answer the following research question:

-Sub-research question 3-

Why are there differences in HTA recommendations for orphan drugs in four countries? Can we learn from these differences?

Results showed differences at each step of the decision-making process. The same pivotal trials were generally appraised, but with varying levels in reporting the clinical outcomes, partly explaining some of the differences. Agency-specific risk and value preferences were also identified, where one agency was relatively more concerned about an issue, or valued to a greater extent a specific criteria compared to the other agencies, which also contributed to explaining varying HTA recommendations. When comparing the issues raised (e.g. uncertainty) by each agency, only 14.5% (Nu = 124) were common across countries, the remainder having been raised by only one or a few agencies. Quantifying these differences using kappa scores showed that agreement was poor to moderate in interpreting the same evidence or in dealing with the same concerns. Differences were also seen in the extent to which stakeholders, or how considerations relating to disease and treatment characteristics, influenced the decisions. This study has led to a better understanding of how value is assessed by different HTA bodies and the reasons for differences differentiating for whether they relate to the processes or methods adopted in each jurisdiction, or whether they reflect weaknesses in the HTA methodological approaches used.

4.3.4. Paper 4

The next step was to focus on one component of the drivers of decisions, the value judgments made, using a more qualitative in-depth analytical approach (Chapter 9,

Paper 4). The aims were (1) to explore in more depth how broader aspects of a treatment's value and the impact of the condition on the patient (referred to as "other considerations"), not captured by routine HTA methods, influence HTA processes; and on this basis, (2) to explore whether orphan drugs have a "special status".

-Sub-research question 4-

How do scientific and social value judgments influence HTA decisions? And on this basis, do orphan drugs have a "special status"?

Results identified in total 125 different "other considerations" or value judgments, which were grouped into 16 subcategories based on the information provided. Between 19% and 100% of these, depending on the agency, were put forward as one of the main reasons for the final decision. A classification framework was developed defining and dividing these into scientific or social value judgments. This was then used to identify needs for further research and areas where more consistency in their use across cases is needed. On this basis, different issues were addressed around better defining the determinants of social value or how to improve the lack of accountability for reasonableness particularly in cases when it is not clear how the "other considerations" identified influenced the decisions. It also provided a way forward to eliciting whether these orphan drugs deserve a special status by eliciting preferences around some of the social value judgments made which are more likely to pertain to orphan drugs compared to normal condition, rather than focusing on the opportunity cost of these. Given the challenges in producing robust evidence for orphan drugs due to the small patient numbers and heterogeneity of the diseases, scientific and social value judgments are unavoidably part of the decision process. Identifying and understanding the scientific and social value judgments made provides a way forward to improving their transparency and consistency across decisions.

These last two chapters showcase the added value of applying this framework, which enables to understand in a simple and comparable manner these complex decision processes, and learn from cases when differences across countries were seen. This contributed to filling some of the gaps identified in the literature, which included 1) the need to scrutinize these variations and query why they occur in a systematic manner;

2) the need for a more consistent and transparent approach in the manner HTA agencies interpret the “other considerations”; and 3) the need to increase transparency in what the appropriate methods are in the different contexts.

4.3.5. Paper 5

The last paper (Chapter 10, Paper 5) builds on the key findings from previous chapters to steer a number of open-ended interviews administered to HTA body representatives to obtain their insights about the results and their meaning when dealing with orphan drugs.

-Sub-research question 5-

How is the value of orphan drugs assessed across different settings and how do differences affect coverage decisions?

Semi-structured interviews with HTA body representatives in the study countries were conducted. An interview topic guide was developed based on findings from a systematic comparison of HTA decisions for ten orphan drugs. The interview questions were divided into four general themes: (a) Evidentiary requirements; (2) Stakeholder involvement; (3) Other considerations; (4) Orphan drugs. Each theme discussed a number of issues seen for orphan drugs and derived from cases when differences were seen across countries or when it was unclear how certain of the criteria identified influenced the decisions. Qualitative thematic analysis was applied to the interview transcripts using the Framework Approach. These interviews were a way forward to furthering the debate about a number of HTA methodological issues, while simultaneously raising awareness about the types of differences that are seen across countries, which may also be applicable within countries when more than one decision-making body exists. Results show that although agreement was generally seen in the evidentiary requirements or preferences, there were subtle differences in the circumstances where uncertain evidence may be considered acceptable, possibly explaining differences in HTA recommendations.

5. Overall PhD Methods

Given that this thesis is in the form of a publishable paper thesis, Chapters 6-10 corresponding to the five empirical papers summarise the methods adopted in greater detail. This section provides an overview of the methods used throughout the thesis.

The first paper (**Chapter 6**) sets the scene as to why it is important to understand the reasons for differences across countries. The study provided a secondary analysis of the HTA recommendations issued in five countries (England, Scotland, Sweden, Canada, Australia) between January 2009 and December 2012. The selection of the study countries was based on whether they had long-established HTA agencies and processes, the availability of HTA reports with the HTA recommendations, the criteria used to produce the recommendations (clinical and/or cost-effectiveness), and a geographical spread encompassing agencies in Europe, Canada, and Australia. Materials for the study were either publicly available or were requested from the study agencies in direct communication. A number of HTA bodies fulfil these criteria, of which five of the more well-established were selected. Countries like France were excluded because the final HTA recommendation (SMR and ASMR) is not directly comparable with the recommendations compared across the study countries (e.g. list, restrict, reject). Other countries such as Germany or the Netherlands were excluded because of language barriers and/or the availability of the data. Data for 297 drug-indication pairs were compiled into a database together with two other research assistants prior to the start of my PhD and with my MSc supervisor. The data compiled included the generic names of all drugs appraised within that period, their indication and ICD code, the date and the HTA recommendation – for each of the five countries. Before using the data collated for my first empirical chapter, I crosschecked most of the data compiled by my colleagues to ensure their exactitude. This database was used for the comparative analysis of cross-country differences in HTA outcomes, as well as of differences in the acceptability of three therapy areas (orphan, cancer, and central nervous system treatments). Descriptive statistics were conducted to understand the frequencies of types of HTA recommendations, per HTA body and per therapy area. Agreement levels and associations between HTA recommendations,

therapy areas and countries were measured quantitatively through correspondence analysis and Cohen's kappa scores. The evidentiary requirements of each HTA body's guidelines were compared, and contrasts were highlighted. From the cancer, orphan and central nervous system drug-indication pairs receiving non-homogeneous recommendations identified, two were randomly selected and further analysed to better understand the rationale for decision-making. The criteria accounted for in their selection was that they should have received non-homogeneous recommendations and were from different therapy areas. The selection was a convenience one. Results show the magnitude and contradictory nature of agreement across countries, suggesting that differences matter and are worth further investigating, which was the overarching objective of the remainder of this PhD.

The second paper (**Chapter 7**) aimed to set up and test a methodological framework, which was accepted for publication in a peer reviewed journal[2], further tested to a greater sample of orphan drugs in the subsequent chapters (Chapters 8-10), where the validity of the framework was explored, as outlined in the next paragraphs. The first paper (Chapter 6) was used as a basis to set up a first outline of this framework. The aim of the framework is to allow for a systematic comparison of HTA-based recommendations as reflected in their HTA reports across drugs and countries. An exploratory sequential mixed methods research design in the form of an instrumental-based model was adopted. A mixed methods design was considered appropriate in order to capture the depth and complexity of these HTA decision processes, and the breadth of decisions across cases (generalisation). The exploratory sequential design was used because (a) of the direct interaction between the qualitative and quantitative strands, (b) priority was given to the qualitative strand given that the research question was more qualitative by nature, (c) the qualitative strand was conducted before the quantitative one (e.g. sequential), (d) the mixing of both strands was done during data collection and analysis, and (e) the approach was exploratory given that the variables of interest were unknown and that no taxonomy of criteria existed. The qualitative strand was used to explain, illustrate quantitative findings, and quantitative findings were used to enhance the qualitative ones (Creswell et al. 2011c).

---

[2] This chapter has been published in Health Policy (2016)

Purposeful sampling was used to select the study countries based on predefined criteria (Creswell & Plano Clark, 2011b), notably whether, (a) they had well-established HTA agencies and processes, (b) similar decision-making criteria (clinical and/or cost-effectiveness) were used, (c) the different approaches used in HTA were represented (e.g. clinical benefit assessment versus clinical and cost effectiveness, health service versus societal approach), (d) the HTA reports were publicly available and included the rationales for the recommendations, and (e) they were based in Europe. These criteria were meant to ensure that the decision processes were comparable for the analysis. On this basis, England, Scotland, Sweden and France were included and constituted the study countries for the remainder of this thesis. France was included in the remaining empirical chapters given that the comparison focused on what went on during the decision process, therefore on the assessment and appraisal of clinical benefit, rather than solely on the final decision. Other countries, such as Germany or the Netherlands were excluded because of language barriers and/or availability of the data.

The framework was developed in three phases. First, case study analyses were piloted in order to identify the range of criteria accounted for and the structure of the decision-processes, such that they were comparable across countries. The two case studies used to develop the proposed methodological framework were selected from all European Medicines Agency approved drug-indication pairs - until December 2012 - that had an orphan designation and were appraised in the four study countries. Drug-indication pairs were excluded if (a) they did not undergo the single technology assessment process at NICE, or the full submission process at SMC where the full HTA process was conducted and documented, and (b) they did not receive diverging coverage recommendations. This resulted in a selection of an oncologic and non-oncologic drug, which are likely to be valued differently by HTA bodies given that they differ in terms of disease and treatment characteristics. Second and on this basis, the decision-making process was decomposed into a more comprehensible manner by looking at the evidence accounted for, its interpretation and how this influenced the final recommendation. This structure of analysis was used to establish the coding manual and case study template (Figure 7-3 & Appendix B. Case Study), which form the basis

of the methodological framework, used for the thematic coding of each HTA report. It also enabled for the coding to be homogeneous, systematic and flexible given the iterative process adopted. Third, the qualitative data were quantitatively analysed in order to understand and measure the level of agreement at each stage of the decision-process and reasons for differences, as well as identify agency-specific preferences across the sample (risk and value preferences). Data sources consisted in the publicly available documents containing the specific HTA recommendation, which will be referred to as the "HTA report". Throughout the thesis, reference to analysing HTA processes refers to the HTA processes and the drivers of the HTA recommendations as they are reflected in the HTA reports. Based on the assumption of transparency, the information reported in the HTA reports were considered to reflect the key reasons around the final HTA recommendation.

The validity of this methodological framework in terms of its accuracy, absence of bias and ensuring the interpretation is traceable and justified was verified through various means. First, through data triangulation with various sources of data, from HTA reports, input from HTA stakeholders and interviews with HTA bodies, where the interpretation of the results was presented and comments were collected. Second, data analysis was undertaken by adopting a more in-depth approach with the case study analyses, while conducting thematic analysis across a greater sample. The advantage of this mixed methods approach allowed for the qualitative findings to explain and illustrate quantitative ones, ensuring that the interpretation of the results was accurate. Third, an audit trail was recorded and included all the thoughts, queries, uncertainties, clarifications and progress of the researcher whilst conducting the research. This allowed for documentation and traceability of all the steps conducted and reasons for having interpreted an outcome in one way of another.

The focus of Chapter 7 (Paper 2) was to describe the approach used to develop and pilot this framework, in order for it to be transferable. An iterative process was adopted to develop the framework, where the case study template, coding manual and quantitative data extraction matrixes were re-worked and re-adjusted several times before arriving at the current framework. The framework has now been applied to

other countries and drugs, by other researchers showcasing its feasibility and reproducibility.

Its application to analysing a sample of orphan drugs in several countries facilitated the identification of a taxonomy of criteria that may have contributed to the decision process, as well as of similarities and differences across countries (Chapter 8, Paper 3). Ten orphan drug-indication pairs in the four study countries were analysed following the framework. Orphan drugs were selected because they are characterised by high and uncertain cost-effectiveness ratios due to the small patient numbers and heterogeneous nature of the conditions they treat (Drummond et al., 2007; Dupont et al., 2011; Kanavos et al., 2012; McCabe, Claxton, & Tsuchiya, 2005). For each drug, a case study report was compiled regrouping all the information of interest. This information was then coded directly on the case study reports using NVivo 10 (QSR International Pty Ltd., 2012). The advantages of having these case studies were to ensure the comparability of the information being related and to facilitate the interpretation of the results (during the next stage). The data was extracted through various thematic matrixes from NVivo 10 into Excel, which allowed to categorise the variables by types and frequencies, into two-way contingency tables. Descriptive statistics were used to quantitatively analyse the data coded.

Correspondence analysis was conducted to measure the associations between the variables of interest using the chi-squared statistic test of independence, facilitating the understanding of these complex relationships in a simple bi-dimensional representation (Dickenson, 2010; Bartholomew, Steele, Moustaki, Galbraith, 2008b). Correspondence analysis was considered appropriate for the purposes of this thesis as it applies to categorical data, unlike principle component analysis that applies to continuous data. It is a descriptive technique used to explore the associations between variables in a contingency table, which are considered dependent if the null hypothesis of independence is rejected. The row and column profiles also allow us to explore the relative positions of the rows and columns to each other, and get a better understanding of the structure of the data.

It does so by looking at the proportions for each coordinate in the corresponding rows and columns (row/column profiles), and the overall proportion for each row or column (row/column mass or average row/column profile). Each row and column profile is compared to the average row profile to determine the amount of scatter or variation, also referred to as inertia. In the biplot, those points with the largest coordinates (profile masses) explain a larger amount of the inertia in that dimension. Those points close to each other have similar profiles and are characterised by a stronger relative association. When interpreting the two-way correspondence analysis biplots (e.g. Figure 6-2, Figure 6-3, Figure 7-4, Figure 7-5, Figure 8-2, Figure 8-3), the coordinates show the extent to which they contribute to the inertia in that dimension. The two-way biplot captures two dimensions: the dimension on the vertical axis that captures a percent of the inertia, and the dimension on the horizontal axis that captures another percent. The points on the origin of the biplot do not contribute to any inertia, whereas those farthest from the centroid contribute to the most inertia. By transposing all points perpendicularly onto the vertical axis, we can identify those points farthest from the origin. The variables that sit the closest together have the strongest associations. Those points on the positive and those on the negative section of the axis represent the contrasting associations amongst variables. Doing the same with the horizontal axis gives additional insights on the existing associations between the variables.

Cohen's kappa scores were used to check the robustness of the results obtained by the primary metric measuring the frequency of common interpretation. These scores were used to measure agreement between two HTA bodies and their interpretation of the evidence, by comparing agreement observed to agreement expected by chance. The scores are calculated by estimating the amount of agreement by which the observed agreement exceeds that of expected by chance alone, divided by the maximum that this difference could be. Agreement may range from poor (kappa = 0) to perfect agreement (kappa = 1). Negative values of kappa correspond to cases when agreement was less than expected by chance.

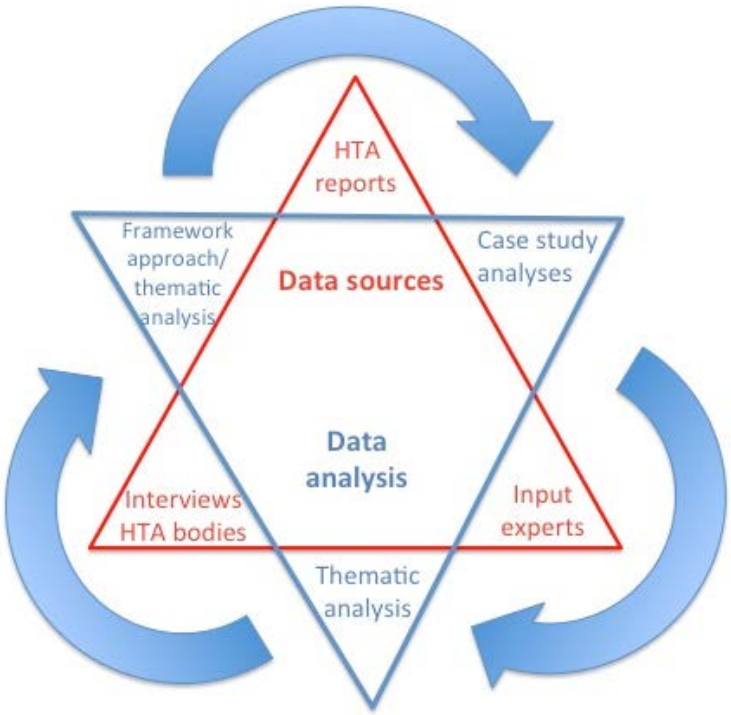
These results from Chapter 8 contributed to raising awareness about how value assessments are conducted in different settings: by shedding light on the four most

common reasons for differences, generating a taxonomy of criteria influencing these decisions, and showing that differences are often the result of a combination of circumstances. The implications for orphan drugs were also discussed.

The depth of the analysis enabled to capture certain aspects of what goes on during the deliberative processes. The fourth paper (Chapter 9) focused specifically on the scientific and social value judgments made during processes. It also aimed to further the debate as to whether orphan drugs deserve special status. A classification framework was developed based on what is known from the literature and was applied to categorise the "other considerations" identified. The willingness-to-pay thresholds in each country were compared to the actual ICERs accounted for in the decision, in order to understand whether greater flexibility was given to high and uncertain ICERs. The variables categorised as "other considerations" previously coded through thematic analysis were then extracted and grouped into clusters depending on the information provided (e.g. synonyms or common themes). These were then analysed to understand what type of value judgment they pertained to (e.g. scientific or social), whether it pertained to disease or treatment characteristics, and how it influenced the final HTA recommendation, including in accepting the high and uncertain ICERs previously earmarked.

The classification framework was used to identify needs for further research and to improve consistency in their use across cases. This was then used to address different issues around identifying and better defining the determinants of social value or how to improve the lack of accountability for reasonableness. It also provided a way forward to eliciting whether these orphan drugs deserve a special status by eliciting preferences around some of the social value judgments made which are more likely to pertain to orphan compared with nonorphan drugs, rather than focusing on the opportunity cost of these.

Figure 5-1. Triangulation of data and analysis



Source: The author.

The final paper (**Chapter 10**) consisted in conducting a number of semi-structured face-to-face interviews with HTA body representatives. The objectives were to ensure that the interpretation of the researcher when analysing the decisions throughout the PhD was accurate, particularly in cases when differences across countries were one of the explanatory factors for different HTA outcomes across countries, and simultaneously get further insights on the drivers of these decisions in each agency. HTA body representatives from each study country were identified by partners of a European research consortium Advance-HTA. All interviewees accepted the invitation. One interview per agency was conducted in order to capture discussions and reach common agreement amongst interviewees. The interviewees included four men and four women, and occupied senior roles in their organisations, e.g. Head of the Technology Appraisal Programme, Head Economist or Pharmacist, Chair of the Appraisal Committee. An interview topic guide was developed. It included open-ended questions derived from actual scenarios that arose in the context of our cross-national comparison of 10 orphan drugs in Chapter 8. The questions were open-ended,

which allowed some flexibility in the discussions such that the interviewees were free to offer additional insights and interviewers could ask spontaneous questions. The topic guide received several rounds of comments from the co-authors of that chapter to ensure that the questions were meaningful. Prior to the interview, it was circulated to interviewees. One interview per agency was conducted, which included one to three interviewees and lasted between 1-3.5 hours. Interviews were recorded and transcribed and sent to the interviewees for comment and validation. Following the interviews, the topics or issues that emerged as relevant or different across countries were compiled and analysed together with a summary of the views of the interviewer, circulated amongst co-authors and accounted for during the analysis. Qualitative thematic data analysis was undertaken using the Framework Approach (Gale NK, Heath G, Cameron E, et al. 2013). Subthemes within each general theme were identified and inductively coded, and a matrix was created to facilitate comparison of each subtheme across the four HTA bodies. The key findings from each of these subthemes were summarised in tables that incorporated illustrative quotes. A list of follow-up questions was developed to complement the interviews where information was unclear or incomplete. These additional questions were sent to each interviewee along with the summary of findings for their particular HTA body for confirmation. Results focus on the contrasts across countries identified within each theme. These interviews were a way forward to furthering the debate about a number of HTA methodological issues, while simultaneously raising awareness about the types of differences that are seen across countries, which may also be applicable within countries when more than one decision-making body exists.

Triangulation of data collection was applied throughout this thesis (Figure 5-1). Data collection consisted in the database compiled of HTA recommendations issued by the study countries between 2007 and 2009, case study analyses and coding of HTA reports through thematic analysis, and interviews of competent HTA authorities. Findings derived from the methodological framework and from the interviews were integrated in the final step and are summarised in Paper 5 (Chapter 10). A more detailed description of methods can be found in each paper (Chapters 6-10).

6. Commonalities and Differences in HTA Outcomes: A Comparative Analysis of Five Countries and Implications for Coverage Decisions[3]

6.1. Abstract

This paper aimed to identify diverging HTA recommendations across five countries, understand the rationale for decision-making in specific therapeutic categories, and suggest ways forward to minimize these inter-country differences. A comparative analysis of HTA recommendations for 287 drug-indication pairs appraised by five countries (England, Scotland, Sweden, Canada, and Australia) between 2007 and 2009 was undertaken, including an in-depth analysis of two case studies. Agreement levels were measured using kappa scores. Associations were explored through correspondence analysis. Results show that significant inter-country variability in the HTA recommendations exists: 46% of the drug-indication pairs studied received diverging recommendations across countries. The level of agreement between agencies was poor to moderate. Associations between HTA recommendations issued by each HTA body per therapy area (cancer, orphan, CNS) differed from the general pattern observed across the complete sample. Expectations from HTA bodies in terms of relative effectiveness differ depending on the drug's and disease's characteristics, although agency-specific guidelines are homogeneous for all treatments. Distinguishing and accounting for the specifics underpinning individual conditions and their characteristics in HTA processes may constitute a way forward to improved HTA methods, while increasing transparency in the expectations that HTA bodies have in terms of relative effectiveness of the drug depending on these characteristics.

6.2. Introduction and Background

Pharmaceutical costs account for an ever increasing proportion of health care costs (Kristensen, Makela, Neikter, Rehnqvist, Haheim, Morland, Milne, Nielson, & Busse,

[3] A version of this chapter was published in *Health Policy* (Nicod E, Kanavos P, 2012)

2009), and it is anticipated that they will only continue to grow in significance with the increasing need for and cost of developing new treatments (Bach, 2009; Congressional Budget Office, 2006). A general consensus at national level exists on the need to control these costs and use resources more efficiently in order to ensure system sustainability. At the same time, the provision of equitable, effective and high quality care remains very challenging in many OECD countries because of both supply and demand-side constraints (e.g. ageing, increased patient expectations) (United Nations Population Fund, 2011; Wanless, 2002), as well as stringent budget constraints. Efforts have been converging towards controlling the level of expenditures through a number of regulatory mechanisms for an affordable, efficient and sustainable health care system (Scherer, 2000).

A key operational contribution of health technology assessment (HTA) in informing decision-making is how to obtain value for money. HTA relies on evidence-based medicine (EBM) to determine the relative effectiveness of a new technology compared to current (best) practice (Drummond et al., 2008), as well as the technology's cost-effectiveness (Eddy, 2009). This mechanism enables a detailed identification of treatments that are cost-effective or provide additional (clinical) benefit compared to existing standard of care, in principle enabling rational decisions for resource allocation (Claxton, Briggs, Buxton, Culyer, McCabe, Walker, & Sculpher, 2008).

Considering that HTA processes are based on internationally recognized methods, it can be assumed that the same drug-indication pair based on the same evidence and relying on –broadly– the same assessment techniques and processes would obtain a similar or comparable recommendation across countries. However, significant disparities have been observed to date in the recommendations made by HTA agencies (Cairns, 2006; Clement et al., 2009; Drummond, 2009; Kanavos et al., 2010b; Kristensen & Gerhardus, 2010; Morgan et al., 2006; Nicod, 2010; Pomedli, 2010; Richards, 2010; Sorenson, 2009; Sorenson, Drummond, & Bhuiyan Khan, 2013; Sorenson et al., 2008; Van den Aardweg, 2010). These may be due, among others, to the quality and type of evidence submitted (Clement et al., 2009; Sorenson et al., 2008; Velasco-Garrido et al., 2005), the appraisal processes (Clement et al., 2009; Sorenson, 2009; Sorenson et al., 2008; Velasco-Garrido et al., 2005), country-specific

considerations (e.g. willingness to pay thresholds) (Clement et al., 2009; Drummond, 2009; Kanavos et al., 2010b; Pomedli, 2010; Sorenson et al., 2008; Van den Aardweg, 2010), or societal preferences (Clement et al., 2009; Sorenson, 2009; Sorenson et al., 2008). Yet, the identification of the key factors driving these diverging decisions is relatively under-studied.

The implication of diverging HTA recommendations is that a particular treatment may be covered in one jurisdiction but not in another, based on individual country combinations of scientific and social value judgments. While these decisions may be justifiable from a policy perspective as they are based on well-established processes, from a patient or societal perspective, having differential access to a particular treatment across –often neighbouring- countries may be difficult to comprehend, particularly in the context of divergent decisions being made by countries with similar levels of income and comparable levels of health care spending. Addressing such disparities across countries has become both a concern and a priority internationally and a number of initiatives exist in this direction (DG Research and Innovation. European Commission, 2011; Hailey, 2009; HTAI, 2011; Kristensen et al., 2010; Kristensen et al., 2009). A better understanding of the critical factors that drive decisions is an essential component in attempts to determine whether and how HTA processes across countries can be approximated.

By pooling together all HTA recommendations across five HTA agencies and as many countries over the 2007-2009 period, the objective of this paper is threefold: first, to provide a post-hoc analysis of those cases where divergences in recommendations have been observed; second, to identify why such differences exist and analyse the critical factors leading to coverage decisions or rejections; and third, to analyse how these recommendations differ across HTA agencies with emphasis on cancer, orphan and central nervous system (CNS) treatments.

The following section outlines the methods used for this analysis, followed by a presentation of the results subdivided into three parts: first, the overall results across all drug-indication pairs included in the study are presented; second, the specific outcomes across three broad therapy areas are studied (cancer, orphan and CNS

treatments); finally, two case studies illustrate some of the possible differences in HTA processes and outcomes by further investigating the appraisal process. A discussion section subsequently identifies the policy implications. The last section draws the main conclusions.

6.3. Materials and Methods

The study provides a secondary analysis of HTA processes and recommendations in five countries: England (National Institute for Health and Clinical Excellence – NICE), Canada (Common Drug Review – CDR, and Committee to Evaluate Drugs for oncology products appraised [since March 2007] - CED), Australia (Pharmaceutical Benefits Advisory Committee - PBAC), Sweden (Dental and Pharmaceutical Benefits Board - TLV), Scotland (Scottish Medicines Consortium - SMC) over a three-year period (January 2007-December 2009). The selection of the study countries was based on whether or not they had long-established HTA agencies and processes, the criteria used to produce recommendation (clinical and/or cost-effectiveness), and a geographical spread encompassing agencies in Europe, Canada, and Australia. Materials for the study were either publicly available or were requested from the study agencies in direct communication.

HTA recommendations over the study period were collected and compiled into a central database, together with their corresponding appraisal reports. Recommendations have been extracted from the appraisal reports issued by each agency. Unique drug-indication pairs have been recorded, according to their international non-proprietary name, ICD10 WHO classification of indications (ICD10, 2007), and HTA recommendation. Only Single Technology Appraisal (STA) has been included in the case of England. The focus of the study is limited to the HTA recommendation issued by the referent agency and not the final coverage decision. HTA recommendations provide a good indication about the level of access to medicines, since the decisions made are generally consistent with the recommendations issued by the HTA bodies (McMahon et al., 2006; Wonder et al., 2006).

Recommendations were classified into three categories, namely: a) “list” (L), where the technology has been accepted as applied for; b) “list with criteria” (LWC), where the technology has been accepted with restrictions (e.g. by limiting the label indication); and c) “do no list” (DNL), where the application has been rejected based on a negative HTA appraisal. In Canada, drugs that received a “list in a similar manner as drugs in the same class” were given a “L/LWC” rating for comparability purposes.

Based on the above classification, cross-country comparisons were made to identify the differences and commonalities in HTA recommendations (homogeneous across the board, or mix of positive *and* negative), the number of technologies appraised, in general, and by therapeutic area. Compounds from three therapy areas (cancer, orphan and CNS treatments) were subsequently extracted. Reasons for their selection include the treatment and disease characteristics that define these three therapy areas. These are very similar for cancer and rare diseases, which often represent severe and disabling diseases and where there is a high unmet medical need. Although these two therapy areas are very similar, evidence shows that different levels of acceptance rates in their coverage decisions apply (Dupont et al., 2011). This also links to the on-going debate as to whether orphan drugs deserve special status because of the small number of patients affected (Drummond et al., 2007). CNS diseases differ from the latter, whereby treatments are often symptomatic and numerous alternatives are available (e.g. schizophrenia) (World Health Organisation, 2012); nevertheless, a strong need for additional treatments remains because of patient compliance issues (Huskamp, 2006). As such, these three therapy areas provide a good basis for comparison through their diversity in disease and treatment characteristics and acceptance rates for reimbursement. The three therapy areas were defined as follows: a) orphan indications based on the EMA classification (N=26) (European Medicines Agency); b) cancer indications include all neoplasms classified under the ICD10 class C excluding orphan indications (N=51) (ICD10, 2007); c) CNS treatments include all treatments from ICD10 classes F (Mental and behavioural disorders), G (nervous system), and R (symptoms, signs and abnormal clinical and laboratory findings) (N=56) (ICD10, 2007).

The level of agreement across agencies was measured using kappa score indicators (Cohen, 1960). Correspondence analyses were performed to describe the associations between the HTA recommendations and the HTA body issuing the recommendation (Bartholomew et al., 2008b). Comparisons of the associations across the three therapy areas were performed to determine whether coverage recommendations for specific therapy areas were likely to differ from the general pattern identified previously. In addition, drug-indication pairs receiving non-homogeneous recommendations were identified, two of which were randomly selected and further analysed to better understand the rationale for decision-making. The focus of these two case studies was on comparing the type of evidence submitted for HTA appraisal across countries, including the type of clinical trials, efficacy and safety endpoints, economic evaluation, the level of stakeholder involvement, and how all these factors were perceived by decision-making operating at national context.

6.4. Results

HTA agency-specific guidelines do not adopt the same perspective or approach in the way they appraise a technology across all study countries. Table 6-1 summarizes the clinical and economic requirements for submission that are stated in agency guidelines (CDR, 2006, 2008a; NICE, 2008; PBAC, 2008; SMC, 2011b; TLV, 2003, 2012). While most agencies consider only direct costs and consequences of an intervention, TLV adopts a societal perspective including both direct and indirect costs and consequences (TLV, 2003).

Table 6-1. Clinical and economic agency-specific requirements/preferences

		CDR/CED Canada	NICE England	PBAC Australia	TLV Sweden	SMC Scotland
CLINICAL EVIDENCE	RCT – H2H	❖	❖	❖	❖	❖
	RCT – indirect	✓	✓	✓	✓ *	✓
	RCT – placebo	✓	✓			✓
	RCT – non-blinded	✓				
	Non-RCT – experimental and observational		✓ **	✓		✓
	Systematic reviews	✓	✓	✓	✓ *	✓
ECONOMIC MODEL	Cost-utility	✓ Qol	❖	❖	❖	❖
	Cost-effectiveness	✓ ***		✓	❖	✓ ***
	Cost-minimisation	✓ Eq		✓ Eq	✓ Eq	✓ Eq
	Cost-benefit	✓ Sec		✓ Sec	✓ *** ✓ ****	
	Cost-consequence	✓ *** ✓ ****		✓ *****		
	Perspective	Publicly-funded health care system	NHS and PSS	PBS	Societal	NHS and social work

Source: (Nicod & Kanavos, 2012), based on individual agency guidelines.

❖ Preferred type of evidence;

✓ Accepted;

*systematic review of indirect comparisons, if no direct comparative evidence available;

** if limitations with RCT and to supplement information from RCTs when they are available;

*** if cost-utility analysis inappropriate;

**** if cost-effectiveness analysis inappropriate;

***** if disaggregation of outcomes would be helpful;

✓ Qol: when there are differences in Health-Related Quality of Life;

✓ Eq: when equivalence demonstrated;

✓ Sec: as secondary analysis;

Legend: RCT: randomized controlled trial; H2H: direct or head-to-head comparison; RCT-indirect: indirect comparison with a placebo-controlled trial; RCT-placebo: placebo-randomized controlled trial; Non-RCT: non randomized controlled trial; CDR/CED: Common Drug Review/Committee to Evaluate Drugs; NICE: National Institute for Health and Clinical Excellence; PBAC: Pharmaceutical Benefits Advisory Committee; TLV: Dental and Pharmaceutical Benefits Board; SMC: Scottish Medicines Consortium; NS: not stated; NHS: National Health Services; PSS: Personal Social Services.

Looking at the type of evidence required in a submission, PBAC has a clear preference for head-to-head RCT (if available) and indirect comparisons (PBAC, 2008), where a submission relying exclusively on direct comparisons of a treatment to placebo will most likely be considered as being weak. In contrast, placebo comparisons are listed as acceptable evidence for CDR, NICE and SMC, although in the case of NICE indirect comparisons are routinely used (CDR, 2008a; NICE, 2008; SMC, 2011b). A similar picture emerges for the use of economic models; NICE typically accepts cost-utility models, whereas CDR accepts cost-utility, cost-effectiveness, cost-minimisation, and cost-consequence analysis depending on the outcome measure (CDR, 2008a). At this level already, the evidence considered across the board is most likely to vary, potentially resulting in diverging HTA recommendations.

6.4.1. Coverage decisions and differential access

Between January 2007 and December 2009, 287 unique drug-indication pairs were appraised by at least one of the five agencies and received either a positive, restricted or negative coverage recommendation, 226 of which were appraised by at least 2 agencies. The former (287 drug-indication pairs appraised by at least one HTA agency) were used to study the general trends of the drugs appraised, whereas the latter (226 drug-indication pairs appraised by at least 2 HTA agencies) were used for comparative purposes.

When comparing HTA outcomes across agencies, results show that only 54.0% (122 out of 226) of the study drug-indication pairs appraised by at least 2 HTA agencies were homogeneously appraised (either all list and list with restrictions, or all do not list), the remaining having received a mix of positive and negative recommendations. In 46.0% of cases (104 out of 226), compounds may be available in one or some of the study countries, but not available in others. This is further emphasized through the level of agreement across agencies measured by the kappa scores (Cohen, 1960) (Table 6-2). Poor agreement exists between all agencies (-0.023 < kappa score < 0.178), except for NICE and TLV, where moderate agreement can be seen (kappa score = 0.228). Considering that the study countries are very similar in terms of GDP levels (average GDP per capita in 2009 = US$ 40,870 (SD US$ 4,317) (IMF, 2010)), it

is most likely that the reasons for different recommendations can be attributed to value judgments and preferences at national level, consequently it is important to determine what shapes these.

Table 6-2. Level of agreement in HTA outcomes across agencies, measured by kappa scores

	CDR/CED Canada	NICE England	PBAC Australia	TLV Sweden	SMC Scotland
CDR	-	0.038	0.165	-0.001	0.062
NICE		-	0.178	0.228	0.105
PBAC			-	-0.023	0.132
TLV				-	0.066
SMC					-

Source: (Nicod et al., 2012).

Legend: CDR/CED: Common Drug Review/Committee to Evaluate Drugs; NICE: National Institute for Health and Clinical Excellence; PBAC: Pharmaceutical Benefits Advisory Committee; TLV: Dental and Pharmaceutical Benefits Board; SMC: Scottish Medicines Consortium

Focusing on the 287 drug-indication pairs appraised by at least one HTA agency, Table 6-3 illustrates the number of compounds appraised by each agency, and the proportion of drugs accepted, restricted, or not recommended for reimbursement. Substantial differences exist in the number of drugs appraised by each agency; for example, 111 drugs were appraised in Sweden compared to 211 in Australia. One way of explaining these variations across agencies is through country differences in terms of what drugs are required to undergo an HTA appraisal in each country. Indeed, all drugs in Canada and Australia, and all newly licensed indications in Scotland can undergo an HTA process (ISPOR, 2012; PBAC, 2008; SMC, 2011b). In contrast, the lower number of appraisals in Sweden can be explained partly by the fact that only outpatient drugs are appraised by TLV, compared to the other agencies that examine both inpatient and outpatient drugs (ISPOR, 2012; Sorenson et al., 2008).

Table 6-3. Total number of appraisals per country, and the proportion of drugs accepted, restricted, or not recommended for reimbursement

	CDR Canada	NICE England & Wales	PBAC Australia	TLV Sweden	SMC Scotland
Total # appraisals	129	110	211	111	193
List/L	3.1%	19.1%	21.8%	71.2%	28.0%
(+/- 95% CI)	(0.1%; 6.1%)	(11.7%; 26.4%)	(16.2%; 27.4%)	(62.7%; 79.6%)	(21.6%; 34.3%)
List with restrictions/LWC	47.3%	63.6%	53.1%	23.4%	40.4%
(+/- 95% CI)	(38.7%; 55.9%)	(54.6%; 72.6%)	(46.3%; 59.8%)	(15.5%; 31.3%)	(33.5%; 47.3%)
Do not list/DNL	49.6%	17.3%	25.1%	5.4%	31.6%
(+/- 95% CI)	(41.0%; 58.2%)	(10.2%; 24.3%)	(19.3%; 31.0%)	(1.2%; 9.6%)	(25.0%; 38.2%)

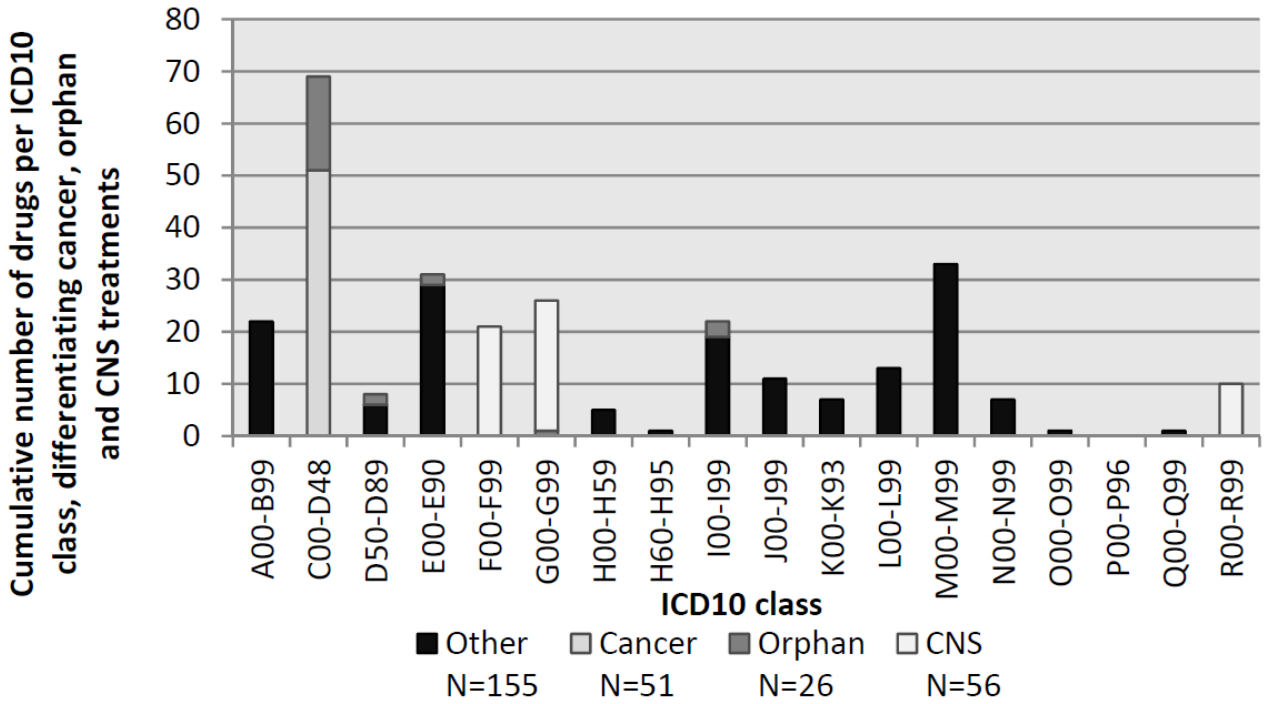
Source: (Nicod et al., 2012).

Legend: CDR/CED: Common Drug Review/Committee to Evaluate Drugs; NICE: National Institute for Health and Clinical Excellence; PBAC: Pharmaceutical Benefits Advisory Committee; TLV: Dental and Pharmaceutical Benefits Board; SMC: Scottish Medicines Consortium; CI: 95% confidence intervals for proportions.

In Canada, 49.6% (64 out of 129) of compounds have received a negative recommendation, whereas this is the case in only 5.4% (6 out of 111) of cases in Sweden. Few or no indication restrictions have been issued in Sweden (23.4% of 111 respectively) compared to Australia or England, where more recommendations with restrictions were issued than either outright positive recommendations or rejections (53.1% of 211 and 63.6% of 110 respectively).

A similar exercise focusing on therapeutic areas also suggests significant disparities in the cumulative number of drugs appraised per therapy area across the 5 agencies (Figure 6-1). A considerably higher number of drug-indication pairs from ICD 10 class C00-D48 (23.7%, 68 out of 287 compounds) representing anti-neoplastic treatments, have been appraised by all agencies, followed by class M00-M99 (musculoskeletal system and connective tissue and endocrine disorders) representing 11.5% of the total, and E00-E99 (nutritional and metabolic diseases) with 10.8% of the total. TLV has the lowest proportion of neoplasm drug-indication pairs appraised (18.0%, 20 out of 111) since it only appraises outpatient drugs, whereas many drugs in this category are used in inpatient settings.

Figure 6-1. Cumulative number of drugs per ICD10 class appraised across the study agencies, differentiating cancer, orphan and CNS drugs.



Source: (Nicod et al., 2012).

This figure represents the number of drugs within each ICD10 class that have been included in the study, and appraised by at least one of the study agencies. Cancer, orphan and CNS treatments have been differentiated.

Legend: ICD Categories	
Neoplasms	C00-D48
Musculoskeletal system and connective tissue	M00-M99
Endocrine, nutritional and metabolic diseases	E00-E90
Nervous system	G00-G99
Circulatory system	I00-I99
Infectious and parasitic diseases	A00-B99
Mental and behavioural disorders	F00-F99
Skin and subcutaneous tissue	L00-L99
Respiratory system	J00-J99
Symptoms, signs and abnormal clinical and laboratory findings	R00-R99
Blood-related	D50-D89
Digestive system	K00-K93
Genitourinary system	N00-N99
Eye and adnexa	H00-H59
Ear and mastoid process	H60-H95
Pregnancy, childbirth and the puerperium	O00-O99
Congenital malformations, deformations and chromosomal abnormalities	Q00-Q99
Certain conditions originating in the perinatal period	P00-P96

6.4.2. Orphan, cancer and CNS treatments and HTA

Narrowing down the study sample to three broad categories of products, those with an orphan status, cancer drugs and CNS treatments, similar results to the complete sample were obtained. There were no marked differences in the proportion of drugs appraised by category submitted in each country, except for NICE and TLV where a small variance is seen in each. Cancer drugs represent the highest proportion of appraised compounds amongst the three categories, on average 51.1%, ranging between 40.1% (29 out of 71) in Sweden and 67.1% (45 out of 67) in England of all compounds. NICE has the highest proportion of cancer drug appraisals most likely because it appraises drugs with the highest need for guidance, which are likely to have significant impact on NHS resources (Dranitsaris, Truter, Lubbe, Amir, & Evans, 2011), whereas TLV currently does not appraise inpatient drugs that are very often cancer therapies (but may be likely to do so in the future).

The HTA recommendations made in each of the three categories suggest that there are substantial disparities in the recommendations made across categories and agencies. In Sweden, all orphan drugs have been recommended with or without restrictions (100% of n=13), whereas in Scotland approximately 65% (15 out of 23) of orphan drugs appraised received a negative recommendation. Focusing more specifically on the common 13 compounds appraised both in Sweden and Scotland, only 8% (one out of 13) received a uniform assessment, 46% (6 out of 13) received diverging assessment (L or LWC in Sweden versus DNL in Scotland), and 46% (6 out of 13) were listed in Sweden and listed with restrictions in Scotland. This demonstrates that there are differences in the key drivers when assessing an orphan drug in these two countries and their acceptability or not of high cost-effectiveness ratios. Because of the nature of rare diseases (e.g. small patient population often associated with severe disability), it is more difficult to collect sufficient data to demonstrate efficacy and safety in very small patient populations, and as such, orphan drugs are often more prone to significant uncertainty in cost-effectiveness. A number of studies focus specifically on orphan drug characteristics and emphasize their clinical uncertainty and high cost-effectiveness (Drummond et al., 2007; Dupont et al., 2011; Simoens, 2011; Vegter et

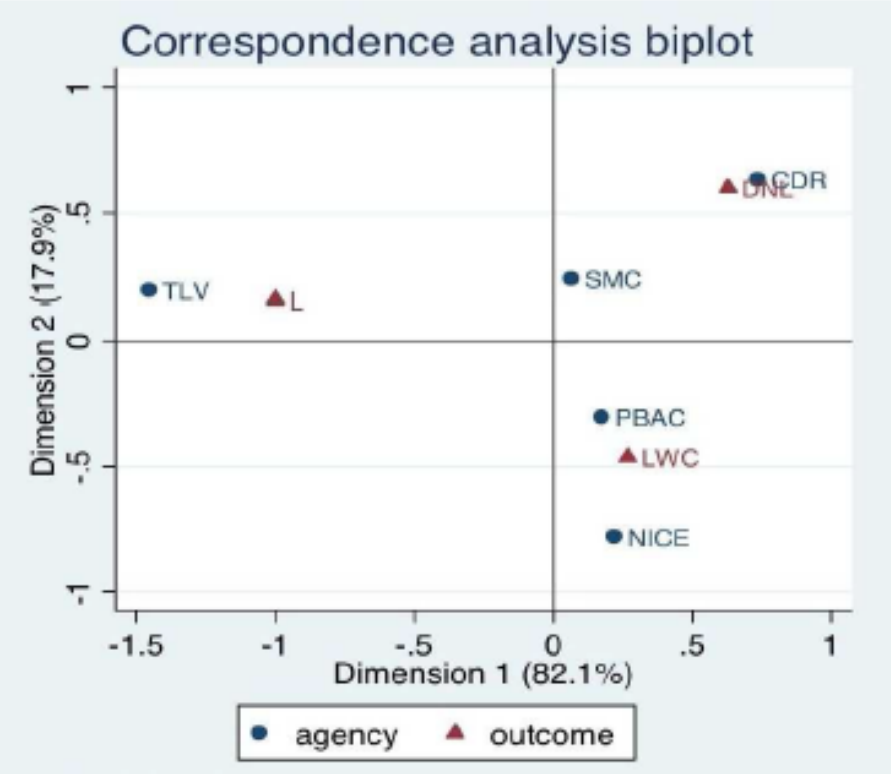
al., 2010), making HTA processes more difficult and subject to interpretation in each setting.

With regards to CNS drugs, 71% (22 out of 31) were rejected in Canada, whereas fewer negative recommendations took place elsewhere (34.1% in Scotland [14 out of 41], 0% in Sweden [out of 29], 19.5% in Australia [8 out of 41], and 15.4% in England [2 out of 13]). This demonstrates that CDR is possibly less inclined to provide a positive recommendation on drugs with marginal benefits (me-too drugs) than other agencies.

Associations between the HTA agencies and their recommendations are summarised in Figure 6-2; further stratified in the three therapy areas in Figure 6-3. In the first case, the Null Hypothesis of independence from the correspondence analysis was rejected ($\chi 2=163.92$; $p<0.000$), demonstrating that an association exists between the recommendations issued and the HTA body. These associations are described in the correspondence analysis biplot, where TLV is relatively more likely to issue a positive recommendation ("L") than the other study agencies; similarly, PBAC and NICE are relatively more likely to issue a positive recommendation with restriction ("LWC"), and CDR a negative recommendation ("DNL") compared to the other study agencies.

In the second case, the Null Hypothesis of independence was rejected ($\chi 2=187.5$; $p<0.000$), demonstrating that associations between the HTA recommendation issued and the HTA body making the recommendation exist. The correspondence analysis biplot (Figure 6-3) suggests that, on aggregate, TLV is relatively more likely to issue a positive recommendation for all therapy areas than the other bodies; NICE is also relatively more likely to issue a positive recommendation for CNS treatments; PBAC is relatively more likely to issue a positive recommendation with restrictions for all therapy areas; and CDR and SMC are more relatively more likely to reject orphan indications.

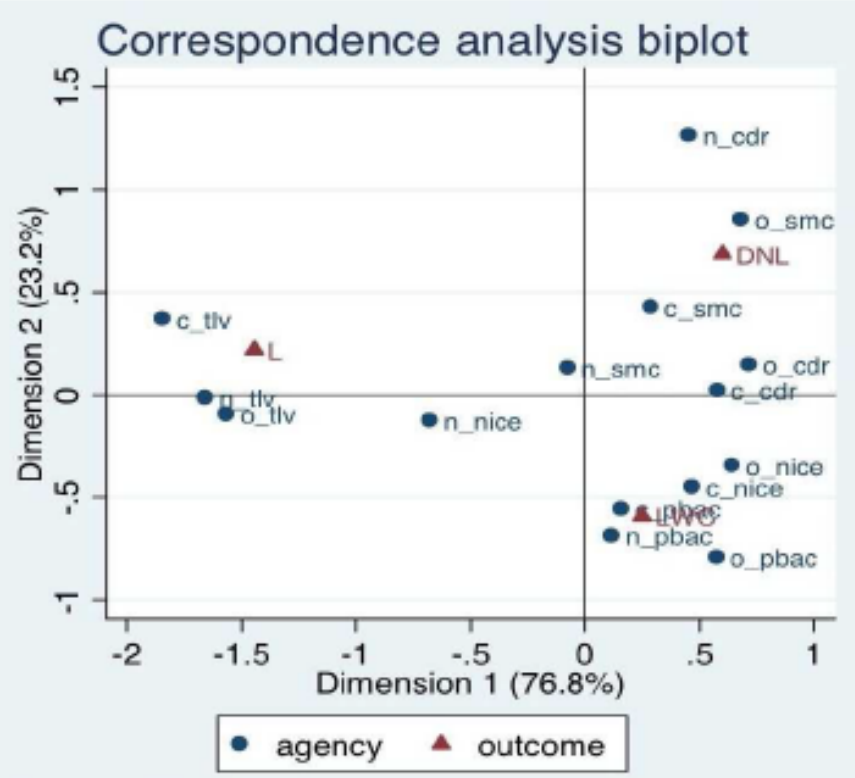
Figure 6-2. Correspondence analysis biplot representing associations between all HTA recommendations and the HTA body issuing the recommendation



Source: (Nicod et al., 2012).

Figure 6-2 summarizes the associations that exist between the HTA agencies and their recommendations issued. The Null Hypothesis of independence was rejected ($\chi 2=163.92$; $p<0.000$), demonstrating that associations exist. The correspondence analysis biplot illustrates these associations on two dimensions. Dimension 1 explains most of the inertia (82.1%), and dimension 2 the remaining 17.9%. Results suggest TLV is relatively more likely to issue a positive recommendation than the other bodies, NICE and PBAC are relatively more likely to issue a positive recommendation with restrictions, and CDR is relatively more likely to negatively recommend when compared to the other bodies.

Figure 6-3. Correspondence analysis biplot representing associations between all HTA recommendations and across three the therapy areas



Source: (Nicod et al., 2012).

Figure 6-3 represents the associations between the recommendations issued in three therapy areas (cancer, orphan and central nervous system indications), and the HTA body issuing the recommendation. The Null Hypothesis of independence was rejected ($\chi 2=187.50$; $p<0.000$), demonstrating that associations exist. The correspondence analysis biplot illustrates these associations on two dimensions. Dimension 1 explains most of the inertia (76.8%), and dimension 2 the remaining 23.2%. Results suggest TLV is relatively more likely to issue a positive recommendation for all therapy areas than the other bodies; NICE is relatively more likely to issue a positive recommendation for central nervous system treatments. PBAC is relatively more likely to issue a positive recommendation with restrictions for all therapy areas; and CDR and SMC are relatively more likely to reject orphan indications

Legend: L: list; LWC: list with criteria; DNL: do not list; CDR: Common Drug Review/Committee to Evaluate Drugs; NICE: National Institute for Health and Clinical Excellence; PBAC: Pharmaceutical Benefits Advisory Committee; TLV: Dental and Pharmaceutical Benefits Board; SMC: Scottish Medicines Consortium; c: cancer treatment; o: orphan treatment; n: central nervous system treatment.

NICE is generally relatively more likely to issue restrictions, but is relatively more likely to recommend CNS treatments. Similarly, CDR is relatively more likely to reject CNS treatments and reject or restrict orphan and cancer indications; in contrast, the SMC is relatively more likely to reject orphan indications and restrict or reject cancer and CNS treatments. The above suggests that HTA bodies may have different stance vis-à-vis incremental versus higher levels of innovation.

Further insights about the nature of decisions can be obtained by examining in greater detail two cases of drug-indication pairs that have received variable recommendations. The results obtained by agency and the rationale are summarised in Table 6-4.

The first case relates to Paliperidone (Invega©) for the treatment of schizophrenia. It has been appraised by all HTA agencies except NICE and has received a mix of positive and negative recommendations (CDR, 2008b; PBAC, 2007; SMC, 2008; TLV, 2008). Although submissions to HTA agencies were made during the same period (2007-2008), different clinical trials were presented. SMC and CDR based their assessments mainly on placebo comparisons, whereas PBAC and TLV on indirect comparisons with quetiapine/olanzapine and risperidone respectively. For SMC and CDR, the lack of comparative data with other anti-psychotics was raised as an issue and one of the main reasons for rejection.

A further issue that was important in agencies' deliberations related to the drug's safety profile. All trials considered, both placebo and indirect comparisons, had a 6-week duration. For CDR and SMC, this duration was considered insufficient to demonstrate the drug's positive toxicity profile. This issue was also raised by PBAC, because the 6-week period characterises only the acute phase of the disease, but did not seem to negatively impact the final decision, suggesting that the data provided were sufficient to demonstrate the drug's safety profile.

The economic assessments presented included a cost-minimisation analysis to all except for SMC, which considered a cost-utility analysis. For the latter, although the cost/QALY estimate was within acceptable range (lower than £30,000/QALY), the drug received a negative recommendation because of the uncertainty in the drug's

clinical benefit. The other three cost-minimisation analyses did not present the same comparators, or results (Table 6-4). In Sweden and Australia, the cost comparisons with risperidone and olanzapine, respectively, were deemed acceptable. In Australia, a flat pricing structure was also proposed and a risk sharing agreement was recommended. In contrast, in Canada, paliperidone had a substantially higher price than generic risperidone, mainly because the generic version was used, which led to an unacceptable outcome. Because of the lack of direct comparative data presented in the submission, this price difference was deemed not justifiable.

In the case of Cetuximab (Erbitux©), indicated for the treatment of metastatic colorectal cancer (mCRC), the drug was appraised by three of the study agencies, positively by NICE and negatively by PBAC and SMC (NICE, 2009b; PBAC, 2010; SMC, 2009). No assessments were found for TLV or CDR (Table 6-4). In this context, different clinical trials were considered by the agencies. NICE and SMC considered the same two placebo-controlled trials, whereas PBAC examined indirect comparisons of the same placebo-controlled trials with bevacizumab as comparator.

Resistance to treatment by patients with mutations of the KRAS gene was demonstrated after the start of the two placebo-controlled trials. Therefore, post-hoc analyses were performed on patients with wild-type KRAS gene and the drug's indication was narrowed down to this patient population; results from this analysis were considered in all assessments, including indirect comparisons. Efficacy was based on the primary endpoints of the two trials, notably "progression-free survival" and "response rate", which were statistically significant in the placebo controlled trials but not in the indirect comparisons. Due to the nature of the sample considered from the post-hoc analysis (retrospective nature, patient representativeness, and sample size), the primary endpoints were deemed uncertain. This was one of the main causes of query and negative evaluation in Scotland and Australia. In England, where clinical experts were solicited and confirmed that these results were "biologically plausible given the current understanding of the pathology of metastatic colorectal cancer", cetuximab received a positive recommendation (NICE, 2009b).

Table 6-4. HTA recommendation and criteria in the evaluation of paliperidone and cetuximab by the different HTA agencies

		PALIPERIDONE				CETUXIMAB		
		CDR/CED Canada	PBAC Australia	TLV Sweden	SMC Scotland	NICE England	PBAC Australia	SMC Scotland
	Date of appraisal	05.2008	11.2007	09.2008	03.2008	08.2009	03.2010	03.2009
	Recommendation	DNL	L	L	DNL	L	DNL	DNL
CLINICAL EVIDENCE	Clinical trials*	Syst-R P-RCT, IND	IND 9 trials	IND	P-RCT IND	P-RCT (post hoc)	IND	P-RCT (post hoc)
	Clinical comparators	Placebo Quetiapine	Quetiapine Olanzapine	Risperidone	Placebo Quetiapine	Placebo	Bevacizumab	Placebo
	Relative efficacy	*			*		X*	*
	Relative safety	*			*		X*	
ECONOMIC MODEL	Economic model	CMA	CMA	CMA	CUA	CUA	CMA, CEA	CUA
	Economic comparator	Generic risperidone	Paliperidone Olanzapine	Risperidone	Olanzapine Quetiapine	Chemo	Chemo	Chemo
	Cost-effective?	*			Aripiprazole		X	X
	RAS/PAS				X			

* Uncertain

Source: (Nicod et al., 2012).

Legend: P-RCT: placebo randomized controlled trial; IND: indirect comparison; NA: not applicable; CMA: cost-minimisation analysis; CUA: cost-utility analysis; Syst-R: systematic review; Chemo: chemotherapy; PAS: patient access scheme; RAS: risk sharing agreement; CDR/CED: Common Drug Review/Committee to Evaluate Drugs; NICE: National Institute for Health and Clinical Excellence; PBAC: Pharmaceutical Benefits Advisory Committee; TLV: Dental and Pharmaceutical Benefits Board; SMC: Scottish Medicines Consortium.

Based on results from the placebo comparisons, the drug's toxicity profile compared to placebo was deemed similar to the Summary for Product Characteristics (SPC) and acceptable. For PBAC however, the nature of the indirect comparison with bevacizumab made it difficult to generate robust data about safety, though this had limited impact on the final assessment.

In England and Scotland, cost-utility analyses were submitted. The cost/QALY of cetuximab compared to chemotherapy was lower in England than in Scotland, because NICE used a provisional lower price. In Scotland, where cetuximab received a negative recommendation, the model was deemed not robust enough because the cost/QALY estimate was above the £30,000 national threshold (ranging between £28,000 and £38,000), together with the uncertainties in the clinical claim. In England, clinical uncertainty was addressed by expert input and an additional model integrating a Patient Access Scheme with an acceptable cost/QALY estimate (< £30,000), based on which a positive recommendation was given. In Australia, based on the clinical claim that cetuximab was not different from bevacizumab, a cost-minimisation analysis was submitted and the drug's cost-effectiveness was rejected mainly because of uncertainty surrounding the clinical claim that the drug is not different from its comparator.

6.5. Discussion and Policy Implications

This study confirms that significant disparities in coverage recommendations exist across the entire range of new pharmaceutical technologies appraised between January 2007 and December 2009. It also demonstrates that these differences are significant in numbers: forty-six per cent of diverging recommendations across countries and mainly a poor level of agreement across countries as indicated by the kappa scores, implying that access to these medicines could vary considerably in individual countries. This may also reflect that HTA processes are influenced by different priorities in individual settings, different perception of benefit and value, and use different tools of addressing uncertainty within their HTA appraisal process. It also reveals different preferences based on settings and the indication under consideration.

The comparisons of the two correspondence analyses exploring the associations between the HTA recommendations for all drugs across the three therapy areas (cancer, orphan and CNS treatments) and the HTA body issuing the recommendation has provided significant insights into some of the key factors driving the appraisal process. SMC, for example, seems to be less likely to approve drugs with high and uncertain clinical cost-effectiveness than other agencies. The latter may put more emphasis on other considerations such as the severity of the condition or the need for treatment alternatives; this is well illustrated in the case of orphan drugs, characterized by high and uncertain cost-effectiveness, which in the majority of cases have been rejected in Scotland but are all listed in Sweden. Several studies have identified factors driving access to orphan drugs in different countries, where cost-effectiveness in some countries was complemented by other considerations such as disease severity, or the availability of other treatment alternatives (Denis et al., 2010a; Drummond et al., 2007; Dupont et al., 2011; Simoens, 2011; Vegter et al., 2010). The analysis has also demonstrated that there are different perspectives in the assessment of me-too drugs. For CNS treatments, CDR appears to put more emphasis on the relative effectiveness of a new treatment and does not appear to encourage marginal benefits, whereas other agencies seem less strict to be placing greater emphasis on other aspects such as patient characteristics or preferences. As such, patient preferences and characteristics seem to weigh more heavily in certain disease areas than others, as also highlighted in the case of psychotropic drugs (Huskamp, 2006).

The design of HTA submissions, including the evidence submitted, appears likely to be tailored by the manufacturers to match agency-specific preferences and requirements. This can be resource-intensive and complex for manufacturers, while at the same time it may provide a window of opportunity to game the system. The complexity of HTAs may itself lead to expectations from agencies being far from clear, and may lead to initial rejections and subsequent re-submissions. This is likely to be time and resource consuming for both the manufacturer and HTA body. Greater transparency, communication and collaboration at earlier stages of the processes could minimize such outcomes, also resulting in a potentially improved access to a treatment following an appraisal.

The case of paliperidone is a good example of different evidence submitted driven by agency-specific requirements. In this case, the manufacturer intended to demonstrate the drug's effectiveness based on placebo-comparisons in Canada and Scotland, and on indirect comparisons in Australia and Sweden, mainly because of the type of evidence required by each agency. The drug's relative effectiveness compared to placebo was deemed insufficient and as a result paliperidone was rejected in Canada and Scotland. Cetuximab also had a similar outlook, where all agencies except PBAC were presented with placebo-controlled trials. This was mainly because PBAC has a clear preference for comparative evidence (head-to-head or indirect comparisons). In the case of PBAC, the nature of the comparison was questionable and uncertain and as a result the drug's relative effectiveness was not demonstrated.

In a number of cases, the interpretation of the evidence could lead to contradictory recommendations. For instance, the 6-week trial duration was considered to provide limited data on paliperidone's toxicity profile and was one of the main reasons for a negative recommendation in Canada and Scotland, whereas this was neither a necessary nor a sufficient reason for a negative recommendation in Australia and Sweden. This was also illustrated for cetuximab that was rejected by all (PBAC and SMC) because of clinical uncertainty, except for NICE that addressed this through expert opinion. This case highlights the importance of the ability and willingness of different agencies to leverage the different types of evidence available. It further demonstrates that uncertain outcomes can be addressed in part if no explicit hierarchies in evidence exist and if the opinion of relevant stakeholders is accounted for (Pearson et al., 2005; Rawlins, 2008).

Differences also exist in the agency-specific requirements regarding economic models and choice of comparator(s), which can partly explain diverging recommendations. The case of paliperidone is an example where the drug was rejected in Canada because of the significantly higher price compared to the price of a generic comparator as shown in the cost-minimisation analysis. This was mainly because the cheapest available alternative should explicitly be included in the Canadian model (CDR, 2006), whereas this is not an explicit requirement elsewhere.

Finally, the possibility to set up risk-sharing agreements may also impact the final HTA recommendation. In the case of paliperidone, a risk sharing agreement was proposed in the submission to PBAC; similarly, in the case of cetuximab, a PAS was included in the economic model and as a result, the drug was deemed cost-effective by NICE. Had these not been suggested, it is likely that both drugs could have received a negative recommendation.

6.6. Conclusion

This study emphasizes the substantial level of disparity in the HTA recommendations issued for pharmaceuticals across five countries, implying that HTA methods may be influenced by different priorities in individual settings, different preferences based on individual settings and therapeutic area, levels of hierarchies in evidence, perceptions of value, tools used to address uncertainty, and the ability and willingness or not to consider and implement risk sharing agreements. Adapting HTA submission requirements the specifics underpinning individual condition and their characteristics to disease areas, by being more explicit about expectations in terms of whether a manufacturer should demonstrate the drug's relative effectiveness compared to placebo or to other comparators may avoid unnecessary rejections and resubmissions. This could also be improved through greater communication about what expectations HTA agencies may have at early stages. A better understanding of agency-specific expectations could also improve the knowledge on why HTA agencies issue different recommendations for a same drug-indication pair, based on which solutions to harmonize guidelines across borders may be proposed. Further research is needed in this direction, in order to better understand and quantify how the evidence submitted may impact the assessment within different therapeutic areas. Ultimately, improved HTA processes may lead to better access and use of health care resources, resulting in better health to the population.

7. Developing an Evidence-based Methodological Framework to Systematically Compare HTA Coverage Decisions across Countries: A Mixed Methods Study[4]

7.1. Abstract

Health Technology Assessment (HTA) often results in different coverage recommendations across countries for a same drug despite similar methodological approaches. This paper develops and pilots a methodological framework that systematically identifies the reasons for these differences using an exploratory sequential mixed methods research design. The study countries were England, Scotland, Sweden and France. The methodological framework was built around three stages of the HTA process: (a) evidence, (b) its interpretation, and (c) its influence on the final recommendation; and was applied to two orphan drugs. The criteria accounted for at each stage were qualitatively analysed through thematic analysis. Piloting the framework for two drugs, 8 trials, 43 clinical endpoints and 7 economic models were coded 155 times. Eighteen different uncertainties about this evidence were coded 28 times, 56% of which pertained to evidence commonly appraised and 44% to evidence considered by only some agencies. The poor agreement in interpreting this evidence (κ=0.183) was partly explained by stakeholder input (n_s=48 times) or by the agency-specific risk (n_u=28 uncertainties) and value preferences (n_{oc}=62 "other considerations"), derived through correspondence analysis. Accounting for variability at each stage of the process can be achieved by codifying its existence and quantifying its impact through the application of this framework. The transferability of this framework to other disease areas, drugs and countries is ensured by its iterative and flexible nature, and detailed description.

[4] A version of this chapter has been accepted for publication in *Health Policy* (Nicod & Kanavos, 2015)

7.2. Introduction

Health technology assessment (HTA) is widely used to provide recommendations on whether health care systems should reimburse a particular drug or health technology. These recommendations rely on information on the comparative effectiveness of different treatment alternatives in a particular clinical setting and aim to ensure that the health technologies covered provide value for money (or are cost-effective) (Luce, Drummond, Jonsson, Neumann, Schwartz, Siebert, & Sullivan, 2010), ultimately, improving access to medicines. In reality, differences in HTA coverage recommendations across countries are seen when appraising the same drug using similar methodological approaches and the same body of clinical evidence. These differences are inevitable due to the complexity of the HTA processes and the context within which they operate, given that each country sets its own objectives for conducting HTA reflecting its values, preferences and constraints (Banta, 2003; Busse et al., 2002; Schwarzer et al., 2009).

Better understanding the application of HTA in different settings together with the reasons for these differences through cross-country learning and sharing of expertise is high on European and supra-national agendas, and may contribute to identify ways to minimize these differences (European Commission, 2011, 2014) or understand how innovation was rewarded by HTA (Bouvy & Vogler, 2013; Pharmaceutical Forum, 2013). This is all the more important given the recent appreciation of HTA as a means towards universal health care (World Health Organisation, 2014) and the commitment of European Member States in implementing cross-border HTA collaboration through the EUnetHTA Joint Action 2 and to further pilot this collaboration in Joint Action 3.

Nine studies (Clement et al., 2009; Kanavos et al., 2010b; Lexchin et al., 2008; Morgan et al., 2006; Nicod, 2010; Nicod et al., 2012; Pomedli, 2010; Shah et al., 2013; Van den Aardweg, 2010) compared HTA coverage recommendations for drugs in more than one country and identified important variations. Three of these quantified these differences (Clement et al., 2009; Lexchin et al., 2008; Nicod et al., 2012), where agreement in HTA recommendations ranged from poor to moderate. The countries compared included Canada, Australia, England, Scotland, France, New Zealand, and

other European countries. One study concluded that the most common reasons for differing recommendations related to the HTA process and context (Clement et al., 2009). Another study highlighted cross-country variations for seventeen of the most expensive drugs, but the extent of and reasons for these differences were not explored (Morgan et al., 2006). A more recent study compared HTA decisions in oncology concluding that negative recommendations were largely driven by the high relative costs in comparison to the marginal benefits (Shah et al., 2013). Possible reasons for diverging recommendations included differences in interpreting the clinical endpoints or in levels of patient input, or issues around appropriate comparators (Shah et al., 2013). Further research on understanding these differences showed that preferences varied according to the therapy area being appraised (Nicod et al., 2012)[5]. The authors highlighted the need for greater transparency around expectations in terms of what constitutes evidence of sufficient quality, how uncertainties are addressed, to what extent disease and treatment characteristics influence the assessment, and whether these vary depending on the therapy area being appraised (Nicod et al., 2012).

These studies have in common the qualitative approach adopted (retrospective descriptive or cohort analyses) to identify these cross-country variations, highlighting possible reasons for these through single case study analyses. None, however, have attempted to scrutinize these variations and query why they occur in a systematic manner. This is likely due to decision-making processes being complex with many factors being accounted for, which may also be inter-related and thus challenging to compare. Comparing these decision processes systematically could contribute to a better understanding of a more comprehensive range of factors accounted for and determine the extent to which they explain differences in coverage recommendations. This could be done by distinguishing between factors relating to context-specific considerations from those that relate to the complexities of HTA processes or the nature and quality of the evidence base of the technologies submitted. Doing so would require a methodological approach that decomposes HTA processes in such way that they can be compared across settings and would enable the identification of the key drivers contributing to decision-making in a systematic way. While this approach may

[5] A version of this publication relates to Chapter 6.

not necessarily eliminate the variation observed in the criteria used to arrive at decisions, reducing it considerably would also be beneficial.

The aim of this study is to develop and pilot such a methodological framework that allows for a comprehensive and systematic identification and comparison of the key factors that influence coverage decisions at different stages of the HTA process. A better understanding of these value assessment processes may help to address some of the methodological challenges in conducting HTA and, potentially, minimize cross-country differences when these are a consequence of the review or interpretation of the evidence.

The framework proposed in this study is informed by evidence from drugs with a European Medicines Agency (EMA) orphan designation (EMA, 2013) and which have undergone an HTA in a variety of settings in Europe. Orphan drugs are often characterized by significant inequalities in access (Le Cam, 2010) and are not always cost-effective (Dupont et al., 2011). In this context, a broader range of factors are likely to be accounted for during the HTA process, which are to be captured by the proposed framework.

7.3. Methods

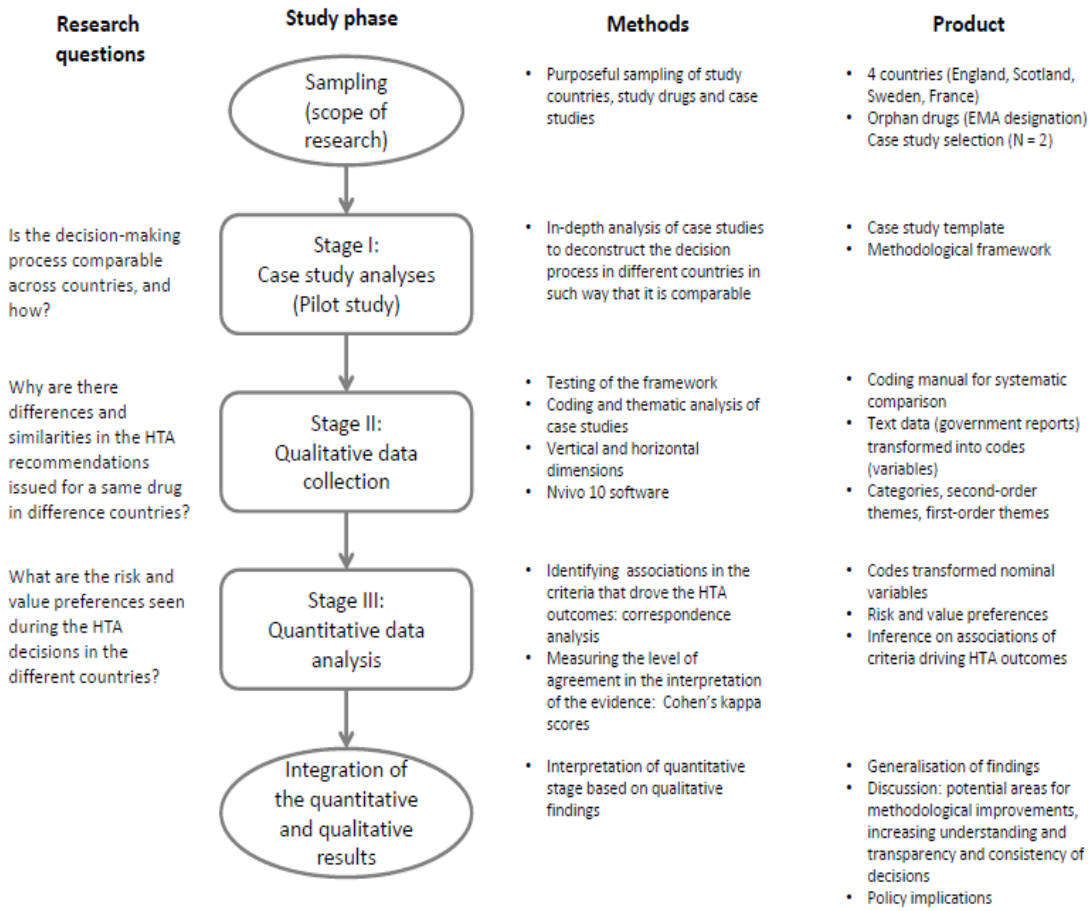
7.3.1. Study design

A sequential exploratory mixed methods research approach was used to develop and pilot the methodological framework in the form of an instrument development design (Creswell & Plano Clark, 2011c). Both the depth and breadth of the HTA decision process were captured within the qualitative (stages I and II) and quantitative strands (stage III) (Creswell, 2003; Creswell et al., 2011c), where findings from the former were built on designing and interpreting the latter aiming at their generalization. A key characteristic of mixed methods design is the "iterative and cyclic approach used in the research" (Teddlie & Tashakkori, 2010), where an inductive logic was used in the qualitative strand in exploring and identifying the decision-making criteria, and a deductive position was used to test the hypothesis made by means of this framework in order to draw inferences from the findings in the qualitative strand (Bryman, 2004).

Priority was given to outline specifically the steps achieved in designing and piloting this methodological framework, while showcasing how the data collected can be analysed quantitatively without drawing any conclusions due to the small sample size.

Figure 7-1 illustrates the stages of this mixed methods study aiming to "simplify the complex interrelationships among elements inherent in HTA processes" (Teddlie et al., 2010), by distinguishing between the different methods and data sources used to address the research questions and their integration in the final stage (Merten, 2011). Within the qualitative strand, stage I consisted in understanding whether the HTA decision-making process was comparable, and, if so, how; this was achieved through two in-depth case study analyses. The aim of stage II was to determine the similarities and differences in the HTA decision processes for the same drug across a number of countries. For this purpose, qualitative thematic analysis was undertaken, where all criteria identified during the HTA process were coded such that they were comparable across case study drugs and countries. Within the quantitative strand (stage III), codes were quantitatively analysed in order to understand HTA agency-specific risk and value preferences, as well as agreement levels across countries.

Figure 7-1. Visual model of the mixed methods research design used



Source: (Nicod & Kanavos, 2015)

Figure 7-1 illustrates the stages of this mixed methods study aiming to simplify the complex interrelationships by distinguishing between the different methods and data sources used to address the research questions and their integration in the final stage.

7.3.2. Sampling

Purposeful sampling was used to select the study countries based on predefined criteria (Creswell & Plano Clark, 2011b), notably whether, (a) they had well-established HTA agencies and processes, (b) similar decision-making criteria (clinical and/or cost-effectiveness) were used, (c) all the different approaches adopted in HTA were represented (e.g. clinical benefit assessment vs clinical and cost effectiveness, health service vs. societal approach), and (d) the HTA reports were publicly available. On this basis, England, Scotland, Sweden and France were included and are referred to as the "study countries". Appendix A summarises some of the key operating features of HTA in these countries.

The subject matter of the analysis was orphan drugs based on EMA orphan designation (EMA, 2013). Their coverage decisions are often characterized by controversy due to high degrees of uncertainty about cost and outcome (Kanavos et al., 2012). Drug and indication pairs were the unit of analysis. The two case studies used to develop the proposed methodological framework were selected from all EMA approved drug-indication pairs - until December 2012 - that had an orphan designation and were appraised in the four study countries. Drug-indication pairs were excluded if (a) they did not undergo the single technology assessment process at NICE, or the full submission process at SMC where the full HTA process was conducted and documented, and (b) they did not receive diverging coverage recommendations. Coverage recommendations were either to list, restrict or reject the drug under review, or in the case of France, to issue a ranking of clinical benefit (Service Medical Rendu, SMR) defining the coverage decision and rate, or one of improvement in clinical benefit (Amelioration de Service Medical Rendu, ASMR) providing a basis for the price fixing regime applicable, ranging from major to insufficient. For example, a medicine receiving an ASMR V is considered not to provide any additional benefit and is covered only if its price is inferior or equal to the other treatments.

Two drug-indication pairs remained that formed the basis for analysis: eltrombopag (REVOLADE©) for the treatment idiopathic thrombocytopenic purpura (ITP) and everolimus (AFINITOR©) as a second line treatment for advanced renal-cell

carcinoma (RCC) after failure of alternative therapies. These technologies have evidential characteristics that are broadly similar to other oncology and non-oncology products and are likely to be valued differently by HTA bodies given that they differ in terms of disease and treatment characteristics.

7.3.3. Data sources and data analysis

Stage I of the qualitative strand, involved an analysis of two drug-indication pairs in order to obtain an in-depth understanding of the decision-making processes, the evidence that informs these and determine whether they are comparable across countries. Previous research aiming at better understanding the HTA decision-making process in each study country was used to outline the structure of the process (Nicod et al., 2012). The case study analyses were used as "an intensive study of a single unit for the purpose of defining a larger class of similar units" (Gerring, 2007), where the decision-making processes for each drug were deconstructed in such way that the criteria identified were comparable across countries. The structure was derived by decomposing the process into three components: (a) the clinical and cost-effectiveness evidence appraised; (b) the interpretation of this evidence in terms of whether it was deemed uncertain, the "other considerations" and stakeholder input accounted for; and (c) the impact of the evidence and its interpretation on the final HTA recommendation. Within each of these components, all of the related criteria recorded in the HTA were extracted into the case study template and coded. Data sources comprised HTA recommendation reports and other relevant material published in the study countries and accessed by the authors. Although these materials adopt similar structures and outline the rationale for the decision, their purposes may differ (e.g. legal document and memo in Sweden, summary of advice to the NHS in Scotland).

The end-product of stage I was the development of the methodological framework, composed of the case study template and the coding manual used to inform the second stage of the research. The case study template provides a structure for the information to be extracted and analysed. It includes tables regrouping all the data identified and extracted for each of the components of the decision-process. Each line item represents one criteria and whether it was accounted for by the individual HTA bodies.

It also includes identifiers to ensure cross-country comparability. These are, for example, the number of trial participants to ensure that the trial being coded was the same in the other countries. This facilitated the understanding and comparison of whether the criteria identified was accounted for in the different countries, and how; as showcased in our results. The coding manual consists of an exhaustive list of all of the criteria identified and included in the case study templates, which were coded. These codes were organized into hierarchical levels clustered into common themes according to the information provided, and are referred to as first-order, second-order (clustering first-order themes) and third-order (clustering second-order themes) themes.

In stage II of the qualitative strand, the objective was to pilot the methodological framework developed in the previous stage by comparing the HTA recommendations and the evidence used for this purpose across countries in order to understand the reasons for differences for a same drug and indication pair. This was done in a systematic manner involving thematic analysis in order to identify and study patterns in the data that describe the decision-making process (Bryman, 2004). Bottom-up coding was undertaken (e.g. adopting an inductive approach, from specific to general) (Onwuegbuzie & Combs, 2010), where codes were created while examining the data to summarize and categorize the criteria and variables included in the decision-making process and identified during the case study analyses. The unit of coding, which is the section of text coded that represents one criteria, was clearly defined for each theme (e.g. first-order, second-order, or third order theme) and was illustrated with examples, for consistency across codes to avoid confusion or duplication in the results (Boyatzis, 1998). Double-coding was performed to capture additional information such as those cases where differences across countries were seen, how these were dealt with, and whether this influenced the final decision. For example, each uncertainty was double-coded with: a) those agencies that raised the same concern, b) whether the uncertainty was addressed and by what means (e.g. stakeholder input), and c) whether it was one of the main reasons for the final outcome. Similar double-coding was performed for the "other considerations" identified.

Coding and clustering of codes were performed by the lead author. Intra-coding reliability was tested to minimize coding bias. Reliability of the clustering was tested

by an academic colleague, who re-categorised each individual code into one of these. Where differences were observed, adjustments were made. An iterative approach was adopted throughout the coding process to ensure that the identified criteria captured the numerous dimensions of the decision-making process. At the end of the coding period, all the information coded was reviewed to ensure that the codes reflect what was meant to be coded (within case-comparison) across all codes (cross-case comparison). Primary data collection took place by means of the HTA reports summarizing the recommendations and soliciting input from HTA experts and HTA body representatives to obtain additional insights about the decisions and ensure that the criteria were coded accurately. The analysis was performed using QSR International's NVivo10 (QSR International Pty Ltd, 2012).

In stage III, codes were quantitatively analysed both vertically and horizontally through descriptive exploratory analyses (Onwuegbuzie et al., 2010) to study their interrelationships. Thematic-matrixes summarizing the codes per medicine and country were exported from NVivo10 into Excel and transformed into nominal variables. The statistical software STATA13 was used for the analysis (StataCorp, 2013). The vertical dimension provided findings about agency-specific risk and value preferences. "Risk" was derived from the concerns of the HTA bodies pertaining to uncertainty, and "value" from the "other considerations" relating to the disease and treatment characteristics accounted for. Preferences were explored through correspondence analysis, where the associations between the variables (HTA bodies versus uncertainty or "other considerations") were measured and illustrated in correspondence analysis biplots (Bartholomew, Steele, Moustaki, & Galbraith, 2008a; Dickenson, 2010). The horizontal dimension provided a measure of agreement between the HTA bodies in interpreting the same evidence using Cohen's kappa scores (Cohen, 1960), allowing for a more robust evaluation of qualitative findings by comparing observed frequency of agreement with the probability of agreement occurring by chance.

7.3.4. Study limitations

Whereas the objective of this study is to develop and pilot a framework, which would then be applicable to a wider sample of medicines because of its iterative nature, it is not without limitations. One limitation is whether specific aspects of the decision-making process, particularly the context within which a decision was made, were captured; these contextual considerations, however, were not within the scope of this study. A second limitation relates to the purpose and level of detail provided in the HTA reports, which varies by country. This is unlikely to have affected the results given that the key determinants, defined as the main reasons for the final recommendations, were included in all the reports and provide a good overview of the decision criteria; data triangulation ensured sufficient detail was captured in each case, and comprised the HTA reports, other material and case studies, input from HTA experts (e.g. Advance-HTA consortium, conferences), and interviews with HTA bodies where findings were presented and feedback collected. A third limitation is that the framework relies on two case studies. However, these were selected to proxy decision frameworks in orphan oncology and non-oncology treatments because the ways of valuing these may differ. The two cases are very different in terms of both disease and treatment characteristics, and therefore are considered to appropriately cover different dimensions of decision processes. Finally, the transferability of this framework to other countries and therapy areas is limited to those cases where similar decision-making criteria are accounted for, from HTA entities that are arm's length, responsible for issuing coverage recommendations, and have a transparent process where sufficient detail about the appraisal process and reasons for the final decision are recorded in their decision reports.

7.4. Results

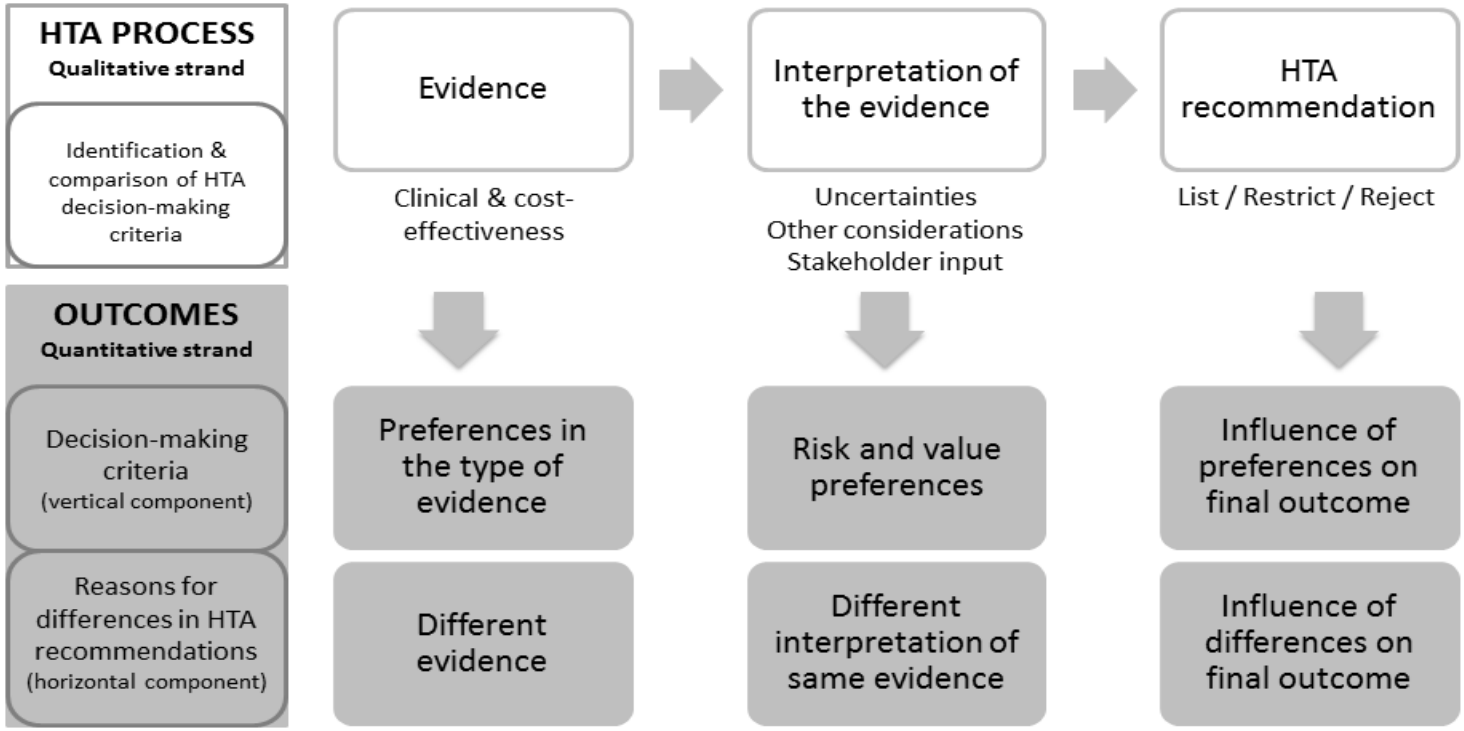
Results are divided into the qualitative (stages I&II) and the quantitative strands (stage III). The first and second sections outline the information collected and coded following the case study template, showcasing how the proposed structure was used to set up and pilot the methodological framework. The third section showcases how the data collected can be quantitatively analysed, where the case studies were used as illustrative examples and results are by no means generalizable due to the small sample size.

7.4.1. Qualitative strand: Developing the methodological framework

Within the HTA process, the manufacturer is responsible to provide evidence about the relative clinical (and cost) -effectiveness of its drug by submitting an HTA application. This evidence is then appraised by the HTA bodies and used as a basis for the recommendation. Three key stages in this process were identified and used as the basis for the methodological framework: a) the evidence, which consists of all the clinical, safety and cost-effectiveness evidence presented (e.g. clinical trials, clinical endpoints, safety, economic model, comparator, comparative effectiveness), b) the interpretation of this evidence, which includes when the evidence was uncertain or incomplete, the "other considerations" accounted for, and the influence from stakeholder input, and c) the final HTA recommendation, focusing on how the previous two stages influenced the recommendation formulation (e.g. HTA recommendations are either to list, restrict or reject the drug for reimbursement for the respective indication) (Figure 7-2).

Within each of these stages, the criteria that were accounted for during the appraisal process were identified and subsequently coded. This was used to establish the case study template and coding manual, both tools forming the methodological framework. The case study gathers all the information accounted for and appraised in a comparable format in one document (Appendix B. Case Study). All the information included in the case studies was then coded in the HTA report, where each new code was included into the coding manual (Figure 7-3). For example, if a phase III RCT was identified, the text providing information about this trial would then be coded once as "phase III" (first-order theme), and would represent the cluster "trial type" (second-order theme), within the cluster "clinical trial" (third-order theme), and so forth. This facilitated the coding and avoided any misinterpretations or missing of information. The next section exemplifies how this was done.

Figure 7-2. Methodological framework for the systematic comparison of HTA processes



Source: (Nicod et al., 2015a).

Figure 7-2 illustrates the three key stages identified and used as a basis for the methodological framework, together with the outcomes from quantitatively analysing the data collected both vertically (e.g. agency-specific preferences) and horizontally (e.g. when differences at each stage of the process explained differences in HTA recommendations across countries).

7.4.2. Qualitative strand: Testing the methodological framework

Eltrombopag (REVOLADE®) and everolimus (AFINITOR®) received diverging recommendations for the treatment of ITP and RCC respectively. Eltrombopag was rejected in England, restricted to patients with severe symptomatic ITP or a high risk of bleeding in Scotland, and listed in Sweden until its re-assessment in 2 years' time. In France, it was valued as having an important medical benefit and providing an important improvement in medical benefit (ASMR II). Everolimus was rejected in England and Scotland, listed as applied for in Sweden, and was considered to provide an important medical benefit in France with a low added benefit (ASMR IV).

Clinical and cost effectiveness evidence

The same phase III primary trials were considered by all agencies for the two study drugs. For eltrombopag, a number of additional clinical trials were considered, one of which was an indirect comparison of eltrombopag with romiplostim, which was appraised by NICE, SMC and TLV, but not by HAS. For everolimus, subgroup analyses by prognostic categories (favourable, intermediate or poor) of the primary trial were also considered by SMC and HAS, whereas NICE additionally conducted a meta-analysis of 28 studies (Table 7-1).

Table 7-1. Clinical trials and their endpoints considered for eltrombopag and everolimus (non-exhaustive list)

ELTROMBOPAG	England NICE	Scotland SMC	Sweden TLV	France HAS	
RAISE					Phase III, 6-month, placebo-controlled (N=197)
Platelet response	✓	✓	✓	✓	
Rescue treatment	✓	✓		✓	
Bleeding (WHO1-4)	✓	✓	✓	✓	
Bleeding (WHO2-4)	✓	✓		✓	Clinically significant bleeding
Bleeding (WHO3-4)	x				Gross (grade 3) and debilitating (grade 4) blood loss
Main reduction in bleeding	x	✓			Seen in grade WHO2
HRQOL (SF36)	✓ 4 x 6	✓ 4			Significant over 4 domains, and not over 6
KUTER					Indirect comparison with RAISE
Platelet response	✓	NR	NR	NR	
Platelet response	x				Splenectomised patients
Platelet response	x				Non-splenectomised patients
EVEROLIMUS					
RECORD-1					Phase III, placebo-controlled (N=416)
Progression-free survival	✓	✓	✓	✓	
Overall survival	x	x	x	x	Blinded phase
HRQOL	R*	x*		x**	*EORTC, FKSI-DRS, **QLQ-C30
Objective response		x		x	
Progression-free survival (subgroup)		✓		✓	Per risk stratification group

Source: (Nicod et al., 2015a).

Legend: √: Statistically significant; x: Non-statistically significant; R: reported; NR: not reported;; EORTC: European Organisation for Research and Treatment of Cancer; FKSI-DRS: Functional Assessment of Cancer Therapy Kidney Symptom Index-Disease Related Symptoms; QLQ-C30: Quality of Life Questionnaire and the Symptoms Associated with the Disease.

A variety of ways to report the clinical endpoints and safety results from the same trials were seen (Table 7-1). For instance, in the case of eltrombopag, WHO 3-4 bleeding events were recorded only by NICE and quality of life relied on the number of domains reported, where it was significant over four (as reported by SMC), but not over six domains (as reported by NICE). Results from indirect comparisons were statistically significant across the whole population, but not significant for the subgroups of patients (as reported only by NICE). A similar scenario was seen for everolimus, where quality of life was not reported by TLV and neither was the objective response rate or progression-free survival in the subgroup analysis by NICE or TLV. Adverse events were reported by all agencies except TLV, but this can be explained by the difference in purpose of the TLV reports, which are of legal nature. Generally, the most common and clinically significant adverse events, and treatment discontinuation rates were reported homogeneously. Additionally, HAS usually provided more detail around the percentage of patients affected and deaths (even if not associated with the treatment).

Three different categories of clinical evidence were identified (as third-order themes): (a) the clinical trials, comprising eight trials (three of which were considered as primary evidence) and five subgroup analyses with their respective comparators coded 51 times across countries; (b) 43 different clinical endpoints (e.g. primary, secondary, health-related quality of life) coded 68 times, and (c) the assessment of safety, recorded in a variety of ways and coded 22 times. This resulted in a total of 141 codes each representing an individual criterion defined as first-order themes, grouped into a number of second-order themes. For example, each trial was coded according to the type of trial (e.g. phase III, phase II and so forth), the type of comparator (e.g. placebo, standard care) and whether it was a primary trial. Trial is a third-order theme, defined by the trial type, comparator and whether it was primary evidence (second-order themes), each of these further defined in detail with the first-order theme (e.g. type of trials are defined by whether they are phase III, phase II, etc.).

Economic models were appraised by all HTA agencies except HAS. For eltrombopag, different economic models were considered. The NICE submission included three models: a) the “watch and rescue model” comparing conventional care to eltrombopag

for splenectomised (£104,100/QALY) and non-splenectomised patients (£116,800/QALY), b) the "long-term continuous model" that included a sequence of treatments with eltrombopag, and c) a cost-effectiveness model comparing eltrombopag to romiplostim. The watch and rescue model was considered the most appropriate (reflecting clinical practice), and the other two were rejected on the basis that they do not represent clinical practice (romiplostim was at the time of this assessment under review at NICE) or were not valid, respectively. In contrast, SMC considered a cost-utility model where eltrombopag was found to dominate romiplostim in splenectomised (savings £12,641 and 0.039 QALY gain) and non-splenectomised patients (savings £2,094 and 0.028 QALY gain), and TLV a cost-minimization analysis with the same comparator based on a non-inferiority claim. For everolimus, cost-utility models were considered by all with best supportive care as comparator. This resulted in a cost per QALY of £49,000 and £51,700 for NICE depending on the approach used, £61,330 for SMC, and was not specified in TLV's HTA report.

Two categories defined the cost-effectiveness evidence (third-order theme): the economic model (second-order theme), coded 7 times (first-order theme), and the comparator included in the model (second-order theme), also coded at 7 occasions (first order theme). This resulted in a total of 14 codes (first-order themes), clustered into two groups (second-order themes). The coding manual summarizes the full list of criteria identified as well as the groups they were clustered in (e.g. second-order and third-order themes) (Figure 7-3). A similar approach was used for the component on the interpretation of the evidence, discussed in the next section.

Interpretation of the evidence

A total cumulative number of 18 (clinical) uncertainties were raised by the agencies for both case studies (10 for eltrombopag and 8 for everolimus) and coded 28 times (e.g. some may have been raised by more than one agency). Some of these were based on evidence commonly appraised by all (56% of 18 uncertainties), while others on evidence included in only some of the appraisal reports (44% of 18 uncertainties). Additionally, some of these uncertainties were also put forward as one of the main reasons for the final recommendation (44% of 18 uncertainties).

Table 7-2 illustrates the different phases in interpreting the clinical evidence, structured in a way that it facilitates the understanding of events in the study countries. The first column reports the evidence (e.g. evidence, considered by) that was interpreted, and whether it came from the primary trial. This enables us to highlight cases when the interpretation of the evidence was based on evidence also considered in the other countries. The second column (e.g. interpretation, raised by) reports whether the assessors highlighted a concern or uncertainty about this evidence, and whether this issue was deemed acceptable or not (e.g. addressed or not). It also highlights whether the HTA bodies raised the same uncertainty and whether it was dealt with in the same manner. The last column (e.g. outcome, main reason for recommendation) reports whether this issue was also one of the main reasons put forward for the final recommendation, and how this compared in the other countries (e.g. was it dealt with in the same manner? was it also put forward in the other countries?).

This allowed to understand how these issues were dealt with across settings. For example, in the case of eltrombopag, the lack of direct comparative data was generally a concern for all and one of the main reasons for the final recommendation for TLV and HAS. The duration of the primary trial was deemed too short to capture the full effects of the treatment for NICE, SMC, and HAS. It was not specifically raised by TLV, but may indirectly be reflected in the conditional nature of the decision with its planned reassessment after 2 years. The primary trial's small sample size was a concern for TLV, but considered acceptable given the treatment's orphan status. Although eltrombopag demonstrated improved outcomes in platelet response and need for rescue medication, NICE was concerned by the fact that this improvement was not significant in the low incidence of the most severe bleeding events (WHO grades 3 and 4); this issue was not raised nor recorded by the other agencies either because this endpoint was not specifically appraised or was not identified as being relevant to the decision. Quality of life data were presented in the submissions to NICE and SMC, but no mention is found in the TLV and HAS reports. This was a concern for HAS, who acknowledged that quality of life is severely affected by the condition, and that there is a need for additional evidence around this.

Table 7-2. Differences and similarities in the interpretation of the clinical evidence and main reasons for recommendation

		Evidence considered by				**Interpretation** uncertainty raised by				**Outcome** main reason			
		✓evidence considered ✓*evidence considered within the pivotal trial				✓positive influence (addressed)		x negative influence (not addressed)		✓positive influence (addressed)		x negative influence (not addressed)	
	Clinical uncertainties	NICE	SMC	TLV	HAS	NICE	SMC	TLV	HAS	NICE	SMC	TLV	HAS
Eltrombopag	Lack of comparator	✓*	✓*	✓*	✓*	x	x	x	x			x	x
	Short duration of trial	✓*	✓*	✓*	✓*	x	x		x	x			✓
	Sample size	✓*	✓*	✓*	✓*			✓				✓	
	Trial population, indication under review	✓*	✓*	✓*	✓*	x				x			
	Trial population, generalizability	✓*	✓*	✓*	✓*	✓	x						
	Trial population, low platelet count patients instead of those with severe risk of bleeding	✓*	✓*	✓*	✓*	✓							
	Significant bleeding events (WHO3-4)	✓*				x				x			
	Quality of life estimate	✓	✓						x				x
	Liver function monitoring		✓				x						
	Uncertain nature of the indirect comparison	✓	✓	✓		x	x						
Everolimus	Bias in overall survival (cross-overs)	✓*	✓*	✓*	✓*	✓	✓	✓	✓	✓	✓	✓	x
	Weak overall and partial response		✓*		✓*		x						
	Lack of comparative safety evidence	✓	✓	✓	✓		x						
	Trial population, co-morbidities excluded	✓*	✓*	✓*	✓*		x						

Trial population, generalizability	✓*	✓*	✓*	✓*	✓	
Risk of pneumonitis, immunosuppression	✓	✓	✓	✓	✓	
Quality of life	✓	✓		✓		x
Risk stratification method (subgroup analysis)		✓	✓	✓	x	

Source: (Nicod et al., 2015a).

In the case of everolimus, the trial was terminated early given the stopping rule around superior efficacy after which patients were given the opportunity to switch from placebo to treatment. Results were biased due to this high number of cross-overs (81% of patients). Nevertheless, for NICE, overall survival was deemed plausible based on clinical expertise and results from a meta-analysis. SMC and TLV also agreed that one specific tool to derive overall survival into more accurate estimates was appropriate. In contrast for HAS, no benefit was demonstrated in overall survival.

A similar analysis about the interpretation of the economic models is possible in order to understand the types of concerns raised by each HTA body, how these are comparable and dealt with. For eltrombopag, different economic models associated with different outcomes were appraised in the three countries considering clinical cost-effectiveness. The comparators differed: NICE considered both conventional care and eltrombopag within different treatment sequences, while SMC and TLV both considered romiplostim. For both NICE and SMC, the trial's secondary endpoint "bleeding events" was included as the main effectiveness endpoint; this was not specified in the report from TLV. The assumption that differences between treatment arms occur because of bleeding events was a concern for NICE and SMC. However, in the latter case, sensitivity analysis showed no differences with romiplostim and a substantially low threshold value for it not to be cost-effective. For everolimus, the interpretation of varying clinical evidence was not associated with differing outcomes, but the different methods used to extrapolate the effects yielded different conclusions. These, together with different willingness to pay thresholds, are likely to have led to different HTA recommendations.

A number of "other considerations" were identified in the HTA reports and coded 62 times. These are summarized into the case study template categorized within their second-order themes (Table 7-3). They were divided between considerations around the treatment and its characteristics, such as the type of benefit provided from the treatment or its innovativeness, and those around the disease and its severity, unmet need, or the nature of the condition. These may have been put forward as part of the reasoning for the final recommendation and/or may have been raised by different stakeholders (e.g. patients, clinicians). For instance, the oral administration benefit of

eltrombopag was one of the main reasons for the final decision by SMC and TLV. Similarly, patients and clinicians stated that adverse events are tolerable and manageable, and that patients are willing to cope with them in order to get treatment. Another example is the life-threatening nature of the condition that was put forward by patients and clinicians in the NICE appraisal for everolimus and was also one of the criteria for recognising the drug as an end-of-life treatment. In total, 41 "other considerations" were identified for the two case studies (first-order themes), and clustered into nine categories according to the information provided (second-order themes) (Table 7-3).

Stakeholder input was seen 46 times in total across the two case studies in the NICE HTA reports and twice in the SMC reports; no stakeholder input was identified in the TLV and HAS reports. In the NICE assessments, patient input was identified in 30% of cases (14 out of 46), providing mainly information on "other considerations" (e.g. impact of symptoms from the disease on daily activities, anxiety from the symptoms affecting quality of life, etc.); and clinical input was identified in 70% of cases (32 out of 46), providing mainly information on "other considerations" (e.g. on limited treatment options when current alternatives fail, issues around the use of treatment alternatives in clinical practice) and commenting on some of the clinical and economic uncertainties raised (e.g. about generalizability of results, issues around clinical practice). For SMC, clinical experts provided input about "other considerations" (e.g. clinical practice) and commented on uncertainty (e.g. generalisability).

Table 7-3. "Other considerations" identified in the HTA reports

			Eltrombopag				Everolimus			
	Subcategory	Illustrative quotations/code	England NICE	Scotland SMC	Sweden TLV	France HAS	England NICE	Scotland SMC	Sweden TLV	France HAS
TREATMENT	Type of benefit	Curative, life-extending				✓	✓ EoL			✓
	Innovativeness	Innovative, new class of drugs	✓ C		✓ Main		✓ C, P			
	Adverse effects	Similar across treatment arms	✓							
		Manageable, tolerated, transient, reversible					✓ P			
	Administration	Oral administration		✓ Main	✓ Main					
DISEASE	Unmet need	Unmet need, no or few treatment alternatives, need for options	✓C	✓ Main	✓ Main	✓	✓ C, P			✓
	Nature of the condition	Disease severity, serious condition			✓ Main	✓			✓ Main	
		Life-threatening	✓ C, P		✓ Main		✓ C, P, EoL		✓	✓
		Short life-expectancy					✓ EoL		✓	
		Impact on quality of life, functional capacity, impact on daily activities	✓ C, P		✓Main	✓				
		Social stigma, limiting of life-style choices, ability to work, travel or undertake leisure activities	✓ C, P							
	Rare disease	Orphan status, small population		✓ Main	✓ Main	✓	✓			

Clinical practice	No routine standard pathway, complex clinical practice, tailored to the patient	✓ C ✓ C		
	Comparator unlicensed for indication, and associated with important adverse events and anxiety	✓ C, P		
	No long term evidence and unknown dosage of comparator	✓ C		
	Late diagnosis (advanced disease)			✓
	Preference for licensed over unlicensed	✓ C		
National priority	Plan Maladies Rares (2004)		✓	

Source: (Nicod et al., 2015a).

Legend: Main: considered one of the main reason for the final recommendation; ✓: considered during the assessment; C: put forward by clinicians; P: put forward by patients; EoL: eligible as “end-of-life treatment” for NICE.

Figure 7-3. Coding Manual

VERTICAL DIMENSION

	THIRD-ORDER THEME	SECOND-ORDER THEME	FIRST-ORDER THEME	DEFINITION	UNIT OF ANALYSIS
EVIDENCE	CLINICAL TRIALS	**Trial type**	Phase III Phase II Extension trial Open-label trial Indirect comparisons Subgroup analysis	Refers to the type of trial included in the assessment, regardless whether or not results (outcomes from the clinical endpoints) were provided.	Paragraph, sentence or part of the sentence with mention of the type of trial and comparator.
		Trial primary	Primary or pivotal trial	Refers to the pivotal trial or primary source of evidenc used as main evidence.	Code pivotal or primary trial together with trial type and comparator.
		Comparator type	Placebo	Refers to cases when the treatment is compared to placebo (no treatment).	Paragraph, sentence or part of the sentence with mention of the type of trial and comparator. Code each comparator only once.
			Treatment	Refers to cases when the treatment is compared to another treatment alternative.	
			Standard care	Refers to cases when the treatment is compared to standard care, which could be, for example, a surgical intervention or palliative care.	
			Standard care-placebo	Refers to cases when comparison is made between standard care (= placebo) and standard care + treatment. Standard care could be, for example, chemotherapy.	
			None	No comparators (e.g. open-label or observational studies)	
	CLINICAL ENDPOINTS	**Effect type**	Primary endpoint Secondary endpoint Health-related quality of life Clinical endpoint (type not specified)	When the type of clinical endpoint is specified, code with specified effect type. If this is not specified, code under "not specified".	Paragraph, sentence or part of sentence with mention of the outcome of the effect. Also code if endpoint is mentioned but
	SAFETY	**Safety type**	Common adverse events	Refers to cases when the most common or frequent events are mentioned.	Paragraph, sentence or part of sentence with mention of the safety assessment. Code each type of safety event reported only once.
			Severe adverse events	Refers to the more severe adverse events or those that are greater or equal to grade 3 events.	
			Discontinuation	Refers to cases when discontinuations are reported.	
			Deaths	Refers to cases when deaths were recoreded, regardless whether they were linked to the treatment under investigation or not.	
			Percent of patients with adverse events	Refers to when the proportion of cases of patients with adverse events was reported.	
	ECONOMIC MODELS	**Economic model**	Cost-utility analysis Cost-effectiveness analysis Cost-minimisation analysis	Refers to the type of economic model included in the assessment.	Paragraph, sentence or part of a sentence with mention of the economic model and comparator. Code each model only once.
		Economic comparator	Placebo	Refers to cases when the economic model compares the treatment to treatment with placebo.	
			Treatment	Refers to cases when the economic model compares the treatment with another treatment alternative.	
			Standard care	Refers to cases when the economic model compares the treatment with standard care (e.g. chemotherapy, palliative care).	
			Sequence	Refers to models that include a sequence of treatments including the treatment under assessment.	

	THIRD-ORDER THEME	SECOND-ORDER THEME	FIRST-ORDER THEME	DEFINITION	UNIT OF ANALYSIS
INTERPRETATION OF THE EVIDENCE	CLINICAL UNCERTAINTIES	**Clinical uncertainties**	Clinical benefit	Uncertain statistical significance or magnitude of the benefit lack of information with regards to a specific endpoint.	Paragraph, sentence or part of a sentence with mention of the clinical uncertainty. Code each uncertainty only once.
			Evidence and study design	Potential bias from the design and conduct of the trials: uncertain nature of indirect comparisons, assumptions around the trial design or randomisation.	
			Trial population	Generalizability of trial results to local clinical practice, representation of the trial population to the indication being appraised.	
			Comparator	Lack of comparative evidence, choice of comparator.	
			Sample size	Trial not sufficiently powered due to sample size.	
			Trial duration	Trial period too short to capture the drug's long term benefit or reflect clinical practice .	
			HRQoL	Absence of any quality of life data in the submission.	
			Administration or provision	Issues around the additional requirements from receiving a treatment (e.g. monitoring).	
	ECONOMIC MODEL UNCERTAINTIE	**Clinical uncertainties**	Clinical benefit	Little or no benefit actually seen.	Paragraph, sentence or part of a sentence with mention of the uncertainty raised around the economic model. Code once each model.
			Clinical practice	Treatment sequences included are not representative of clinical practice.	
			Population	Trial population not representative of the population in clinical practice, differences in trial population between treatment and controls.	
			Trial duration	Trial considered not long enough to capture the full effect of the treatment.	
			Safety	Lack of safety data for comparator treatments.	
		Economic model uncertainties	Clinical assumption	Uncertain assumptions about recurrence rates or the clinical benefit of the control arm.	Paragraph, sentence or part of a sentence with mention of the uncertainty raised around the economic
			Sensitivity analysis	Sensitivity analyses testing how changing the parameters influence the incremental cost-effectiveness ratio.	
	OTHER CONSIDERATIONS - Disease characteristics	**Nature of the disease on the patient and its surrounding**	Negative effects of the disease on quality of life or ability to go to work, disease severity	Refers to the nature of the disease and its impact on the patient and his or her surroundings' quality of lives (e.g. ability to go to work, anxiety from the disease, social stigma).	Paragraph, sentence or part of a sentence with mention of the the other considerations. Code once each other consideration once.
		Unmet need	Unmet need for treatment alternatives, few or no alternatives exist	Refers to cases when the unmet need for new treatments was emphasised, through the fact that little or no alternatives exist or that there is a need for additional treatments.	
		Rarity	Rarity, orphan status	Refers to cases when the rarity of the disease, the orphan status of the drug were put forward.	
		Issues with current treatment alternatives	Complex treatment pathways, no consensus on best practices	Refers to issues set forth around the current treatment pathway, such as the fact that little consensus on best practices, or that treatment pathways are complex.	
		National priority	National priority	Refers to cases when the disease area under assessment is recognised as a national priority.	
	OTHER CONSIDERATIONS - Treatment characteristics	**Type of treatment benefit**	Curative	Refers to the type of treatment benefit (e.g. curative, symptomatic, etc.).	
		Innovativeness	Innovative nature of treatment	Refers to the innovative nature of the treatment, such as the new mechanism of action, or that is is considered a new class of drugs.	
		Indirect benefit from the treatment	Ability to go back to work from the treatment	Refers to any information provided around the indirect benefits from the treatment, such as being able to go back to work, improving functional capacities.	
		Tolerance of adverse events from the treatment	Adverse effects	Refers to any information provided around how adverse events from the treatment are tolerated (e.g. non life-threatening,similar across treatment arms).	
	STAKEHOLDER INPUT	**Stakeholder input**	Clinical expertise	Refers to information (uncertainties or other considerations) provided by on of the stakeholders.	Code section of information provided by stakeholder together with uncertainties or other considerations
			Patient expertise		
			Carer input		
			Consultees		

HORIZONTAL DIMENSION

THIRD-ORDER THEME	SECOND-ORDER THEME	FIRST-ORDER THEME	DEFINITION	UNIT OF ANALYSIS
COMPARISON	**Clinical evidence**	Considered by all	The clinical evidence related to the uncertainty was included in all the	Same unit of analysis as, and to be coded with "uncertainties".
		Not included by NICE	The clinical evidence related to the uncertainty was not included in the NICE submission.	
		Not included by SMC	The clinical evidence related to the uncertainty was not included in the SMC submission.	
		Not included by TLV	The clinical evidence related to the uncertainty was not included in the TLV submission.	
		Not included by HAS	The clinical evidence related to the uncertainty was not included in the HAS submission.	
	Different interpretation	Uncertainty raised by all	The uncertainty was raised by all	
		Uncertainty not raised by NICE	The uncertainty was not raised by NICE.	
		Uncertainty not raised by SMC	The uncertainty was not raised by SMC.	
		Uncertainty not raised by TLV	The uncertainty was not raised by TLV.	
		Uncertainty not raised by HAS	The uncertainty was not raised by HAS.	
INFLUENCE	**Influence on assessment**	Positive impact	The consideration of the evidence, uncertainties, or other considerations had a positive impact on the final assessment	Same unit of analysis as, and to be coded with "uncertainties" and "other considerations".
		Negative impact	The consideration of the evidence, uncertainties, or other considerations had a negative impact on the final assessment	
		No impact	The consideration of the evidence, uncertainties, or other considerations had no impact on the final assessment	
MAIN REASONS	**Main reason for recommendation**	Main positive	Main reason explicitely put forward in the HTA report for the final HTA outcome.	Same unit of analysis as, and to be coded with: "uncertainties" and "other considerations".
		Main restrict		
		Main negative		

Source: (Nicod et al., 2015a).

7.4.3. Quantitative strand (Stage III): outcomes from the methodological framework

Applying the methodological framework showed the variety of ways seen in reporting and interpreting the same clinical evidence. It also highlighted cases when the same clinical evidence was interpreted differently and cases where different evidence was accounted for and interpreted differently (Table 7-2). These, in addition to the main reasons put forward by the agencies for the final recommendation, are defined as the criteria driving these recommendations. This section showcases the outcomes of applying the methodological framework by quantitatively analysing the data collected.

Clinical evidence and its interpretation

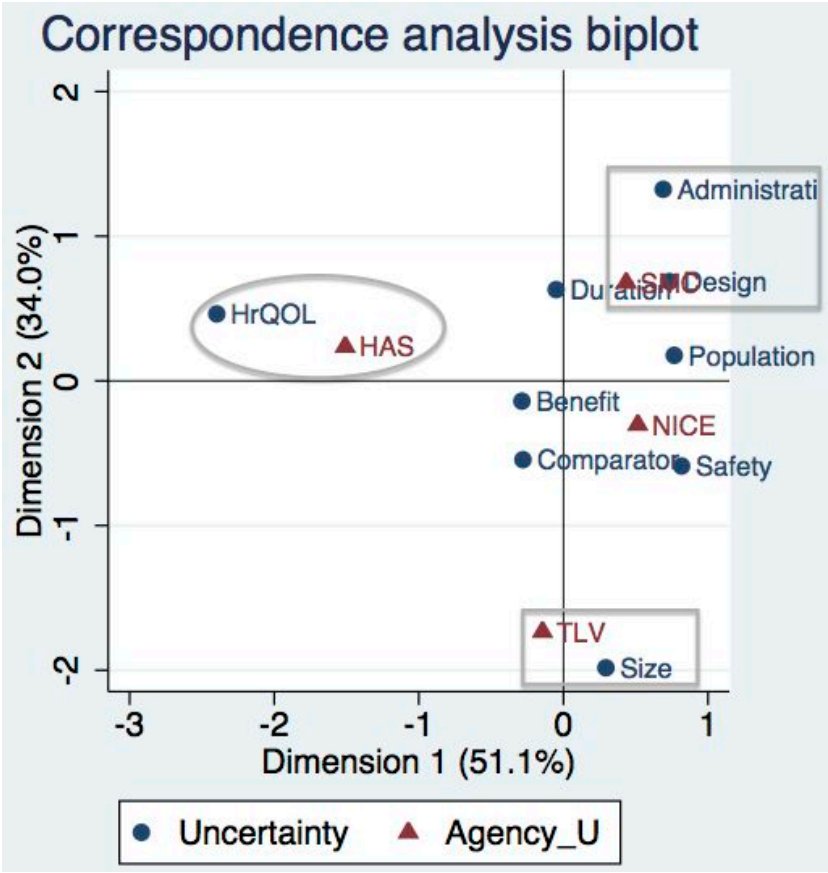
Poor agreement in the interpretation of *the same evidence* was seen ($\kappa = 0.183$, 95% CI [0.015;0.35]) (Altman, 1991). These differences may relate to a subjective (unexplained) component of the decision or to different risk or value preferences. This is illustrated, for example, in the assessment of the short trial duration for eltrombopag.

The correspondence analysis biplot in Figure 7-4 illustrates the relative risk preferences of agencies when appraising *the same evidence*. Although the chi-squared probability of independence is non-significant given the small sample size ($\chi^2=22.49$; $p=0.550$), results nevertheless provide an indication of the relationships among these variables as well as the type of analysis that this framework allows for on a greater sample. On dimension 1 (vertical axis), representing 51.1% of the inertia (or variation), HAS is relatively more likely to raise concerns around quality of life compared to the other agencies. In dimension 2 (horizontal axis), representing 34% of the inertia, TLV is relatively more likely to raise concerns around sample size, and SMC around the administration mode of the treatment and trial design compared to the others. In total, 85% of the inertia is captured across these two dimensions, which provides a good basis for exploring and understanding the associations seen in the data.

Similarly, value preferences were derived from the “other considerations” identified through correspondence analysis and revealed a significant association in terms of the

relative value preferences for these two drugs (χ^2=30.97; p=0.029) (Figure 7-5). In dimension 1 (vertical axis), capturing 57.5% of the inertia, NICE is relatively more likely to account for considerations around clinical practice and adverse effects from the treatment compared to the other agencies, whereas HAS and TLV are more likely to account for considerations around the nature of the disease. In dimension 2 (horizontal axis explaining 24.9% of the inertia), SMC and to a lesser extent TLV, are relatively more likely to account for the treatment's innovativeness compared to the other agencies, and HAS the treatment's clinical benefit. These findings relate to the two case studies and would only be generalizable if conducted on a greater sample of drugs.

Figure 7-4. Correspondence analysis biplots representing associations between the HTA body and the types of clinical uncertainties

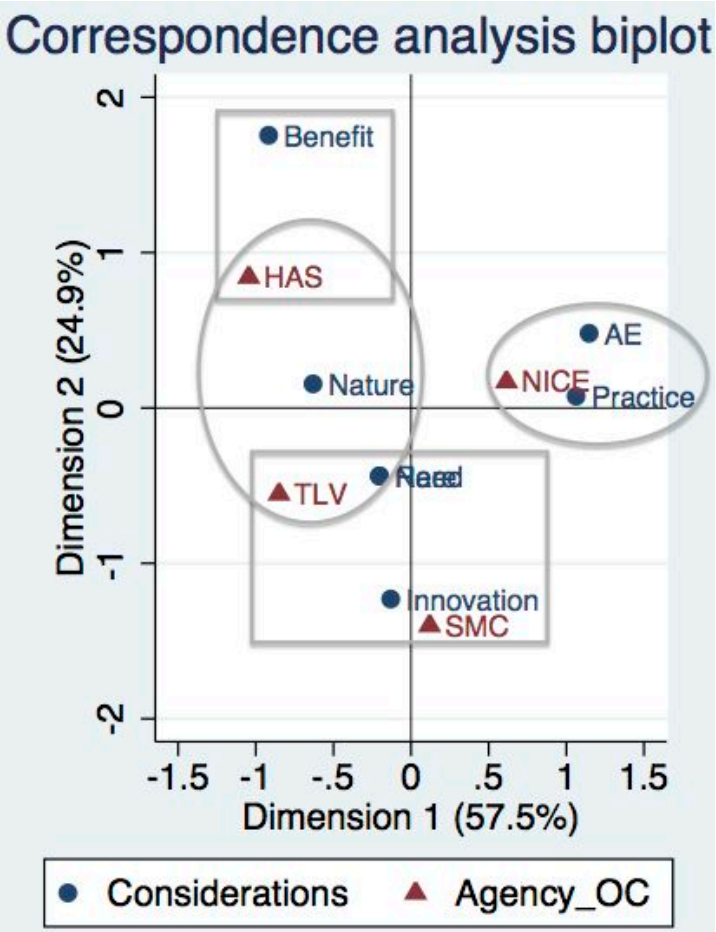


Source: (Nicod et al., 2015a).

Despite the associations being non-significant given the small sample size ($\chi 2=22.49$; $p=0.550$), results provide an indication of the relationships among these variables. Fifty-one percent of the inertia or variation between these variables is captured in dimension 1 (biplot vertical axis). The strongest association is seen with HAS, who is relatively more likely to raise concerns around quality of life compared to the other agencies. In dimension 2, which represents 34% of the inertia (or variation), TLV is relatively more likely to raise concerns around sample size, and SMC around the administration mode of the treatment and trial design compared to the others.

Legend: HRQol: issues with health-related quality of life benefit; Size: issues with sample size; Administration: issues with the administration and provision of the treatment; Design: issues with the trial design and conduct; Duration: trial duration too short; Benefit; uncertain treatment benefit; Population: issues with population generalizability; Comparator: issues with comparator used; Safety: uncertain safety profile of treatment.

Figure 7-5. Correspondence analysis biplots representing associations between the HTA body and the "other considerations"



Source: (Nicod et al., 2015a).

Correspondence analysis revealed a significant association in terms of the relative value preferences for these two drugs ($\chi 2=30.97$; p=0.029). Dimension 1 captures 57.5% of the inertia, where NICE is relatively more likely to consider considerations around clinical practice and adverse effects from the treatment compared to the other agencies, whereas HAS and TLV are more likely to account for considerations around the nature of the disease. In dimension 2 (explaining 24.9% of the inertia), SMC and to a lesser extent TLV, are relatively more likely to account for the treatment's innovativeness compared to the other agencies, and HAS the treatment's clinical benefit.

Legend: Benefit: clinical benefit and type of benefit of the treatment; Nature: disease nature affecting the patient; Innovation: innovative nature of the treatment; AE: adverse events manageable or non-significant; Practice: complex pathway, no best practices; Need: unmet need; Rare: rarity, orphan status.

Criteria driving the decisions

The criteria driving the HTA decisions were defined as: (a) the main reasons for the recommendation identified at each stage of the process (Table 7-2 & Table 7-3), (b) whether and how these were influenced by agency-specific risk and value preferences, and (c) whether they were influenced by decision-makers' judgments about their interpretations of the evidence presented (measured by agreement levels), which resulted into the following decisions.

Eltrombopag was rejected by NICE mainly because of the high uncertainty that increased the ICER to a level greater than what is considered cost-effective. For SMC, although the clinical evidence was weak, eltrombopag was significantly more effective than placebo in platelet response and considered cost-effective, as greater uncertainty in the economic analysis was accepted because it offers additional treatment options, is an orphan drug, and is administered orally. For TLV, eltrombopag was considered cost-effective because of its similar effect at a lower cost compared to romiplostim, which had already been considered as cost-effective by the TLV. The orphan status, severity of the condition, and impact on the patient's quality of life were also put forward. TLV requested that a follow-up on the assessment of effectiveness take place in October 2013. For HAS, while the trial duration was limited and comparative data were lacking, eltrombopag was considered similar to romiplostim until further evidence is provided (risk assessment plan).

Because of the early termination of the trial, the estimated clinical benefit of everolimus was considered biased. In England, overall survival was considered superior to three months and the treatment was eligible as "end-of-life treatment". Nevertheless, sensitivity analysis showed only a very low probability of the drug being cost-effective and was rejected. For SMC, the price was considered too high in comparison with the positive benefits provided. In Sweden, the high cost per QALY was acceptable given the disease's severity. In France, despite being a serious and life-threatening condition, the evidence presented was not sufficient to demonstrate any improvement in survival or quality of life relative to alternatives.

7.5. Discussion and Policy Implications

7.5.1. Summary of key results

This empirical study fitted a mixed methods research design to a research question requiring both an in-depth understanding of the HTA decision-making process and a systematic approach to comparing cases in order to gain a broader understanding of the HTA outcome. The case study analyses highlighted the complexity of these decisions and identified a structure facilitating the understanding and comparability of these processes. This was used to derive and pilot the methodological framework, which divided the decision-making process into three stages within which a set of criteria were identified and coded. This enabled to identify the criteria driving decision processes and explaining cross-country differences.

7.5.2. How do our findings fit with existing evidence?

Comparing our results with existing studies that looked at the criteria influencing HTA decisions in at least one of the study countries (corresponding to the vertical dimension of this study), two studies were identified and their findings are consistent with ours. The literature review of quantitative studies conducted by Fischer aimed to identify the existing empirical evidence on coverage decisions for a range of health technologies (Fischer, 2012). Despite not being directly comparable with our study given it included all types of technologies, it is of interest to ensure that the components of HTA included in our study are comprehensive. Carroll et al. conducted a thematic analysis of the assessments made by the Evidence Review Group at NICE to identify the strengths and weaknesses in the submissions (Carroll, Kaltenthaler, FitzGerald, Boland, & Dickson, 2011). Even though it is again not directly comparable to our study, it remains of interest since it is accounted for by NICE and corresponds to the "interpretation of the evidence" stage in our study; these are summarized in the HTA reports and correspond to the second-order themes within our study. Findings from these two studies validate our classification of the decision-making criteria and confirm that our results are appropriate and comprehensive. Focusing on the horizontal dimension included in our framework, only few comparative studies of

HTA decisions exist, as previously reported in the introduction section. None, however, have compared the decision-making processes in a systematic and comprehensive manner.

7.5.3. The methodological framework

The added-value of this study is that by deconstructing HTA recommendations and developing a taxonomy of criteria that may have contributed to the decision-making process, it enables an enhanced understanding of HTA decision-making (Creswell & Tashakkori, 2007). Without a mixed methods study design, we would not have been able to capture the depth and complexity of these decision-making processes, both within and across countries. The novelty of this methodological framework lies, first, in the systematic approach adopted in analysing the data, and, second, in the inclusion of a horizontal dimension to capture additional aspects of the HTA decisions. The coding and categorization of HTA documents enabled a systematic identification of the decision-making criteria in a homogeneous and comparable manner across countries and drugs. Further, the horizontal dimension also captured, through double-coding, the influence ("positive" or "negative") of each criterion on the final decision ("main reason for recommendation"), and whether it was provided through stakeholder input. Finally, this study exemplifies how this type of design can be implemented to fit a specific research question and disseminated in a clear and transparent manner. It also highlights the inter-disciplinary potential of applying such designs to novel areas, such as HTA.

7.5.4. Policy implications

Based on its application to two cases, results show that a significant number of additional criteria and considerations may be used to inform decisions, which may override pre-existing rules, such as an ICER threshold. This was a consequence of the heterogeneity seen in the evidence and its interpretation or of additional criteria or input, which may have influenced the decision. This may be due, in part, to the orphan nature of these drugs, where accounting for "other considerations" may overcome some of the uncertainty characterising the evidence generated from small patient

populations. It may also be interpreted either as a need to examine in greater depth the available evidence on specific drug-indication pairs rather than stick to a yes-no decision based on otherwise inflexible rules, or as a recognition of the imperfect nature of the HTA process to account for detailed information that may matter when making a decision at the margin, or a combination of both. Results from applying this framework also allow us to raise awareness on the reasons for cross-country differences. Where differences were a consequence of the review, the interpretation of the evidence and dealing with uncertainty, it may contribute to finding solutions to minimising these differences. When applied across a greater sample of drugs, therapy areas and countries, the application of this framework may be beneficial in a variety of ways: to identify decision-making criteria that can feed into other types of models (e.g. multiple criteria decision analysis (MCDA)), to identify agency-specific preferences, to understand the type of stakeholder input meaningful to provide, or to ensure consistency in the "other considerations" accounted for (e.g. accountability for reasonableness).

7.6. Conclusion

Improving the level of access to medicines is a priority at both European and supra-national levels (Bouvy et al., 2013; Pharmaceutical Forum, 2013). This study highlighted the variations seen in the HTA outcomes across countries and the potential reasons for these differences. These differences may be legitimate and reflect context-specific considerations, or may be a consequence of limitations in the application of HTA methodological approaches. Implications would be enormous in seeking to obtain value for money. There is an urgent need to better understand the reasons for these variations and improve the quality of the assessments when they are a consequence of the approaches used. In this study, we have proposed, developed and piloted a methodological framework aiming to account for (part of) the unexplained heterogeneity seen in HTA recommendations across settings. The framework is detailed and provides insights into decision-making practices in the case studies concerned. The framework's external validity is being enhanced and applied to a larger sample of drugs, therapy areas and countries.

8. Why do HTA Coverage Recommendations for Orphan Drugs Differ? Applying a Mixed Methods Framework in Four European Countries[6]

8.1. Abstract

Health technology assessment (HTA) coverage recommendations differ across countries for the same medicine. Unlike previous studies, this study identifies and explains these differences in a systematic manner. HTA recommendations for ten orphan drugs appraised in England (NICE), Scotland (SMC), Sweden (TLV) and France (HAS) (N=35) were compared using an existing methodological framework to identify the criteria driving recommendations and highlight cross-country differences. A sequential mixed methods design was used comprising two stages: (1) qualitative in-depth analysis of the decision-making processes; and (2) quantitative identification of agency-specific risk and value preferences through correspondence analysis, and agreement levels across countries through Cohen's kappa scores. Results showed that six of the ten study drugs received diverging HTA recommendations. This was attributed either to contextual considerations (e.g. NICE end-of-life criteria, SMC modifiers, disease severity for TLV) or to cross-country heterogeneity in: (1) the evidence appraised (50% of six drugs with diverging recommendations), (2) the uncertainties raised when appraising the same evidence (33% of 6 drugs), (3) dealing with the same uncertainty (66% of 6 drugs), or (4) the ability to impose patient access schemes or future re-assessments. Moderate to no agreement across countries was seen in dealing with uncertainty, which was influenced by agency-specific preferences in terms of stakeholder input or considerations relating to treatment characteristics. This research contributes to better understanding how different HTA bodies assess value. As highlighted by the framework used, a more systematic approach is needed in order to look at the predictors of the drivers of coverage decisions in settings using HTA.

---

[6] A version of this chapter is published in the European Journal of Health Economics (Nicod E, 2016)

8.2. Introduction

National competent authorities tasked with making judgments about coverage of new medical technologies are faced with competing needs and goals in containing costs, rewarding innovation while providing safe, effective and quality care to their citizens in a context of a rising prevalence of chronic conditions (European Commission, 2012). Health Technology Assessment (HTA) is commonly used to support such resource-allocation decisions based on the best available evidence of comparative costs and benefits, therefore ensuring that resources are used efficiently (Banta, 2003; Luce et al., 2010).

Providing equal access to affordable medicines across countries is high on the political agenda in many OECD countries including those in the European Union (European Commission, 2013b). In reality, this is far from being achieved even in countries with similar or comparable policies, rules or priorities. In countries using HTA to inform resource allocation decisions, poor to moderate agreement in HTA coverage recommendations across countries is often reported for the same medicine (Clement et al., 2009; Lexchin et al., 2008; Nicod et al., 2015a). These divergences may partly relate to legitimate contextual differences such as the objectives adopted, where it might be a pharmacoepidemiological study in one country and a systematic review of all aspects of using a technology in another (Banta, 2003). Equally, there may be different willingness-to-pay thresholds affecting the extent to which an HTA outcome is acceptable (Pearson et al., 2005; Webb, 2009). Differences may also be due to controversies over the HTA process itself, including questions around the most appropriate methodological approach to undertaking HTA (Brousselle & Lessard, 2011; Drummond, de Pouvourville, Jones, Haig, Saba, & Cawston, 2014a), the application of HTA in each setting, whether the measures used fully capture the effects and costs from taking the treatment (Brazier, 2008; Dolan et al., 2005; Sculpher, 2008), what levels of evidence are acceptable (Gauvin, Abelson, Giacomini, Eyles, & Lavis, 2010; Rawlins, 2008), how to deal with uncertainty (Claxton, 2008), or to what extent “other considerations”, e.g. disease and treatment characteristics, were consistent across decisions (Earnshaw et al., 2008). This, in turn, emphasizes the need to explore these areas of debate in greater depth in order to understand their

importance across settings and provide recommendations for methodological improvements in the conduct or interpretation of HTA and, by extension, evidence-based policymaking in health.

This problem, together with its implications, has been identified and possible explanations examined (Clement et al., 2009; Kanavos et al., 2010b; Lexchin et al., 2008; Morgan et al., 2006; Nicod, 2010; Nicod et al., 2012; Pomedli, 2010; Shah et al., 2013; Van den Aardweg, 2010). These nine studies compared the HTA coverage recommendations made across a sample of drugs and countries and highlighted the extent of these differences. Their research designs were in the form of retrospective descriptive or cohort analyses, and countries compared included Canada, Australia, England, Scotland, France and New Zealand. The reasons for cross-national differences were also explored, but with varying levels of thoroughness. Morgan and colleagues focus on the transparency and rigour of the processes rather than on case-specific reasons for diverging recommendations (Morgan et al., 2006). The three other studies investigate the reasons for these differences (Clement et al., 2010; Lexchin & Mintzes, 2008; Shah et al., 2013), but relied on a few cases or potential reasons. First, they did not outline the key determinants or structure of the decision-making explored, where the reasons set forth may not constitute the full picture. Second, issues relating to the clinical and pharmacoeconomic assessments, also referred to as clinical and economic uncertainty, were identified. However, the level of detail provided in their assessments did not differentiate for the type of uncertainty (e.g. trial duration or magnitude of the benefit), how these were dealt with across countries (e.g. acceptable by some and not by others?), and what the factors influencing these processes were. Third, the methodological approaches used were not sufficiently detailed for these approaches to be transferable. One exception may be the paper from Lexchin and colleagues that describes how the variables were categorised (Lexchin & Mintzes, 2008), which was accounted for when setting up the coding scheme in this thesis. Given that these decision processes are complex and understanding what happened for one same drug in different countries may be challenging, a more systematic, structured, and comprehensive approach to identifying and comparing differences would be required. Additionally, understanding how similar scenarios were dealt with

across settings may also constitute a way forward to identify limitations in applying HTA and learn from how these were dealt with across settings (Nicod et al., 2015a).

Through the application of an existing methodological framework (Nicod et al., 2015a), the purpose of this study is two-fold: (a) to systematically identify and compare the drivers of HTA recommendations for a sample of orphan drugs across four countries; and (b) to identify the reasons for the different HTA recommendations issued across countries at different stages of the HTA process. The subject matter of the analysis was orphan drugs, given they are often not cost-effective due to the small patient numbers, heterogeneous nature of the conditions they treat, and their often high prices (Drummond et al., 2007; Dupont et al., 2011; Kanavos et al., 2012; McCabe, Claxton, & Tsuchiya, 2005). Different studies nevertheless demonstrate that orphan drugs receive the same or a higher level of acceptance compared to other drugs treating more common disease areas (Simoens, 2011; Stolk et al., 2009). Special attention was given to understanding the level of uncertainty in the evidence presented characterising orphan drugs, how it was dealt with, and how disease and drug-specific characteristics were accounted for in different settings.

8.3. Methods

8.3.1. Sampling of study countries and drug-indication pairs

The study countries, England, Scotland, Sweden and France, were selected based on: (1) their well-established HTA agencies and processes, (2) the variety in HTA approaches used, notably clinical (e.g. France) and/or cost-effectiveness (e.g. England, Scotland, Sweden) as decision-making criteria, (3) the different types of HTA body (e.g. advisory in England and Scotland, regulatory in France and Sweden), (4) the different perspectives to HTA adopted (e.g. health service perspective in England, Scotland and France, and societal in Sweden), (5) the public availability of HTA reports, recommendations and other material, and (6) their European location.

In order to arrive at a common sample amongst HTA bodies, NICE was used as a benchmark. All drug-indication pairs with an orphan designation from the European

Medicines Agency (European Medicines Agency, 2012) and appraised by NICE in England through the Single Technology Appraisal process until December 2012 were included and recorded by their indication, generic name, and HTA recommendation (e.g. to list, restrict or reject a drug for coverage). 269 technology appraisal reports were published until December 2012 by NICE, 23 of which related to orphan drugs with an EMA designation. Excluded were: those that underwent the Multiple Technology Appraisal process or were terminated at the time of data collection at NICE (9/23), and those that were appraised by fewer than three study of the four countries (4/23). Additionally, when a compound underwent the abbreviated procedure at the SMC, it was not included given that the rationale for the decision, of interest for this study, was not made available. This resulted in a selection of ten unique orphan drug-indication pairs and a total of 35 country and drug-indication pairs (Table 8-1). Only five were included by TLV in Sweden, which appraised mainly outpatient drugs at the time of the study, while many of the study drugs were inpatient (Faulkner, Matuszewski, & Niziol, 2009).

Special considerations in appraising orphan drugs in each country were identified. These included the recognition of rare diseases as a national priority in France enshrined in legislation (Ministere des Affaires Sociales et de Sante, 2004), and the SMC modifiers (e.g. life-threatening, life-expectancy and health-related quality of life (HRQol) improvement, curative treatment, unmet need) in Scotland (SMC, 2012). By contrast in England and Sweden, orphan drugs follow the same HTA process as drugs for more common conditions.

Table 8-1. List of drug-indication pairs included in the study

Generic name/Brand name	Indication	ICD 10 code[1]	NICE England	SMC Scotland	TLV Sweden	HAS France[2]
Eltrombopag REVOLADE	Chronic idiopathic thrombocytopenic purpura	D2	DNL TA205 Oct 2010	LWC 625/10 July 2010	LWC 3731/2010 May 2011	SMR important ASMR II June 2010
Romiplostim NPLATE	Chronic idiopathic thrombocytopenic purpura	D2	LWC TA221 Apr 2011	LWC 553/09 May 2009	LWC 833/2010 Oct 2010	SMR important ASMR II June 2009
Everolimus AFINITOR	Renal cell carcinoma (2nd line, advanced)	C	DNL TA219 Apr 2011	DNL 595/10 Mar 2010	L 853/2010 Sep 2010	SMR important ASMR IV Jan 2010
Lenalidomide REVLIMID	Multiple myeloma (3rd line)	C	LWC TA171 Jun 2009	LWC 441/08 Apr 2010	L 410/2010 Jul 2010	SMR important ASMR III Oct 2007
Mifamurtide MEPACT	Osteosarcoma	C	LWC TA235 Oct 2011	L 837/13 Jan 2013	LWC 2312/2012 Jan 2013	SMR insufficient DNL Nov 2010
Azacitidine VIDAZA	Myelodysplastic syndrome	D1	LWC TA218 Mar 2011	LWC 589/09 Aug 2011	NA	SMR important ASMR II Apr 2009
Imatinib GLIVEC	Gastro intestinal stromal tumour (adjuvant, after surgery)	C	DNL TA196 Aug 2010	LWC 584/09 Aug 2010	NA	SMR important ASMR III Sep 2009
Mannitol dry BRONCHITOL	Cystic fibrosis	E	LWC TA266 Nov 2012	DNL 837/13 Jan 2013	NA	SMR weak ASMR V Sep 2012

Ofatumumab ARZERRA	Chronic lymphocytic leukemia	C	DNL TA202 Oct 2010	DNL 626/10 Jul 2010	NA	SMR moderate ASMR V Oct 2010
Trabectedin YONDELIS	Soft tissue sarcoma	C	LWC TA185 Feb 2010	DNL 452/08 Jun 2011	NA	SMR important ASMR V Apr 2008

Source: (Nicod, 2016a).

1 WHO ICD10 code classifications: C & D1: Neoplasms, D2: Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism, E: Endocrine, nutritional and metabolic diseases.

2 The ASMR (Amélioration du Service Médical Rendu) ranks drugs according to their relative improvement in clinical benefit in five levels, from a major innovation (level I) to no relative improvement (level V). The SMR (Service Médical Rendu) ranks the drug according to the drug's clinical benefit in five levels, from major to none.

Legend: NICE: National Institute for Health and Care Excellence (NICE); SMC: Scottish Medicines Consortium; TLV: Dental and Pharmaceutical Benefits Board; HAS: Haute Autorité de Santé; L: list; LWC: list with restrictions; DNL: do not list; NA: not applicable.

8.3.2. Study design and methodological framework

The methodological framework applied (Nicod et al., 2015a)allowed for the systematic comparison of HTA decision processes across countries and drugs. The approach used was an exploratory sequential mixed methods design (Creswell & Plano Clark, 2011a), where the qualitative strand took priority and preceded the quantitative strand. The framework consisted in a coding manual and case study template (Chapter 7). This allowed to break down the decision process into sub-components, which can be analysed and compared across countries (Figure 7-2): (a) the evidence appraised (e.g. trial type, clinical and safety endpoints, comparators, economic models), (b) the interpretation of this evidence (e.g. nature of uncertainty, how it was dealt with, and the influence of stakeholder input and "other considerations"), and (c) their influence on the final recommendation (Nicod et al., 2015a). Uncertain evidence was defined as evidence considered not fully capturing the effects of a treatment in the intended population by the assessors (NICE, 2009a). "Other considerations" were defined as the non-quantifiable or non-quantified considerations relating to treatment or disease characteristics not captured by routine methods of HTA (e.g. QALY).

8.3.3. Data analysis

This multi-level research design allowed for an in-depth analysis of the criteria driving these decision-making processes (qualitative strand), and of their role in shaping these decision processes in each country and whether they explained cross-national differences (quantitative strand). This research did not aim to generalise findings, but was interested in exploring and elucidating the reasons behind the HTA decisions, which is mainly qualitative by nature. The quantitative strand aimed to complement and enhance the interpretation of the qualitative findings, and to produce more structured data to be used for subsequent analyses.

<u>Qualitative strand</u>

Data sources comprised publicly available HTA reports, other official documents that include information on the appraisal process and reasons for the final HTA recommendation (e.g. memos in Sweden), and comments from competent authorities.

Although their aims differ (e.g. “advice” in Scotland versus “decision” in Sweden), the HTA reports were assumed to reflect the determinants driving the recommendations given that countries are required to be transparent in their decision processes (European Commission, 1989). Results were also regularly presented to HTA experts (e.g. Advance-HTA consortium) at various occasions, where feedback was collected. This contributed to ensuring that the interpretation of the decisions made by the researcher was accurate.

At each step of the process, all the relevant information driving the decisions was extracted and coded. First, the information from individual HTA reports was compiled into case study forms together with identifiable data (e.g. trial sample size) to ensure that the information collected was comprehensive and comparable across countries. Second, on this basis, thematic analysis was undertaken to code the HTA reports using the software NVivo 10 (QSR International Pty Ltd, 2012). Coding was flexible and iterative with new codes being created for all newly identified criteria and included in the coding manual with their definition and coding rule, ensuring that the multiple dimensions of the decision-making process were captured. The HTA reports already coded were re-examined with these new codes and adjustments were made if necessary. Intra-coding reliability was tested for consistency of coding, as well as content validity for the representativeness and homogeneity of the information coded within codes (Creswell et al., 2011a). Third, the data collected were exported into excel for analysis using different coding matrix queries.

Quantitative strand

The qualitative data collected were transformed into quantitative categorical nominal variables by exporting the data to Stata 13 (StataCorp, 2013). Thematic matrixes and descriptive statistics were used to categorise the variables by types and frequencies of (a) evidence, (b) uncertainty, (c) “other considerations”, and (d) stakeholder input. The data also provided information about their influence on the final recommendation and how all these compared across countries.

Quantitative data analysis consisted in identifying and measuring agency-specific preferences and cross-country agreement levels. Risk preferences were derived from

the types of uncertainty, and value preferences from the "other considerations" identified in the HTA reports. Correspondence analysis was used to measure agency-specific risk and value preferences (Friendly, 1991; Hoffman & Franke, 1986), highlighting when one type of uncertainty or "other consideration" was relatively more commonly raised by one HTA body compared to another. It allowed for the measurement of associations between these variables using the chi-squared statistic test of independence, facilitating the understanding of these complex relationships in a simple bi-dimensional graphical representation (e.g. correspondence analysis biplot) (Bartholomew et al., 2008b). This technique was chosen as it applies to categorical data unlike principal component analysis that applies to continuous data (Bartholomew et al., 2008b). For comparability purposes, TLV was not included in this first part of the analysis given that it relies on only five cases, but in a secondary analysis relying on the five drugs commonly appraised by all.

Descriptive statistics were used to measure the frequency of agreement across countries in their interpretation of the evidence. Cohen's kappa scores of cross-country agreement levels were measured to check the robustness of the results obtained by the primary metric (Cohen, 1960). Kappa coefficients were used to quantify agreement adjusted for chance across HTA agencies about how the same evidence was interpreted, and was done so in a comparable manner given it focused on each individual concern (uncertainty) raised that was common across settings. Two categories of agreement were measured: (a) the type of issues raised by each agency about the same evidence, and (b) how the same issues raised by at least two agencies were dealt with across settings. This allowed to compare agreement observed to agreement expected by chance, ranging from poor ($\kappa = 0$) to perfect agreement ($\kappa = 1$), and where negative values of κ correspond to cases when agreement was less than expected by chance (Altman, 1991).

Finally, the analysis also aimed to identify those issues or considerations that relate to the rarity of these conditions, and understand and compare the different approaches to dealing with them across settings.

8.4. Results

Six of the ten study drugs received diverging recommendations, where they were positively recommended or restricted in some countries and rejected in others (Table 8-1). Out of the four remaining cases with homogeneous recommendations, romiplostim and lenalidomide were restricted in their indications in some countries and not in others, ofatumumab was rejected by NICE and SMC and received the lowest ASMR V rating with a moderate SMR rating (30% reimbursement rate). In only one case (azacitidine) were the recommendations issued really similar. Different trends were also seen, where, for example, mifamurtide received a positive recommendation from NICE and SMC, but was considered insufficient and rejected by HAS. This rarely occurs in France as most drugs considered not to provide any additional benefit would receive an ASMR V rating. Another contrast was seen for eltrombopag, considered important by HAS with a substantially high ASMR rating (II), whereas it was rejected by NICE. These examples illustrate the magnitude and contradictory nature of these cross-country differences, suggesting that they are important and have significant implications for patients in terms of access, and for society in terms of using healthcare resources efficiently. In order to understand why they occur, the decision-making process was scrutinized and compared across countries for these ten drugs on the basis of the structure provided in the methodological framework.

8.4.1. Evidence

The same primary trials were considered, which were predominantly phase III RCTs, one of which was considered only by TLV within an indirect comparison (e.g. lenalidomide with bortezomib), and phase II trials for the two remaining drugs following early marketing authorisation (e.g. trabectedin and ofatumumab). These primary trials had relatively small sample sizes (e.g. less than 300 participants in 60% of trials) and decisions often relied on results from subgroup analyses (e.g. 50% of cases). Comparators were mainly standard care with the exception of two cases comparing different doses of the treatment under investigation (e.g. mannitol dry, trabectedin) and one case with no comparator (e.g. single-arm study for ofatumumab). For 80% of the study drugs, the primary endpoints were surrogate and predominantly validated with the exception of “time-to-progression” for soft tissue sarcoma and “platelet response” for idiopathic thrombocytopenic purpura. In two cases, NICE’s

main outcome of interest was “overall survival” despite it not being the trial’s primary endpoint (e.g. imatinib, ofatumumab).

The inclusion of the remaining non-primary non-phase III trials had very little influence on the assessment. Outcomes from these trials were generally not reported, and when reported, the type of data provided was around safety (e.g. romiplostim, ofatumumab, eltrombopag), dosage research (e.g. eltrombopag) and historical controls (e.g. trabectedin).

Focusing on the economic evidence, similar cost-utility models were considered by NICE, SMC, and TLV except for eltrombopag, for which a cost-minimisation analysis was considered by TLV. Additionally, the comparators used by NICE and SMC for eltrombopag were different: NICE considered conventional care, while SMC and TLV considered romiplostim. No cost-effectiveness models were included in the HAS reviews, as cost-effectiveness was not a requirement for first time approvals at the time of the study.

Different evidence was included by some agencies and not by others. When comparing the trials considered by NICE to those considered by SMC, TLV and HAS, one out of 19 trials, four out of 15, and six out of 23 respectively were not included in the NICE appraisals. These included a database used to estimate HRQol data for trabectedin for SMC; two open-label trials (eltrombopag) and two registries (romiplostim) for TLV; and four phase II open-label trials (azacitidine, eltrombopag), one post-marketing surveillance survey (study extension for eltrombopag), and one indirect comparison (trabectedin) for HAS. HRQol data was not specifically reported in 5 out of 10 cases, and in four other cases, it was not reported homogeneously across the board.

These differences in the evidence appraised were associated with differing HTA outcomes in five cases (Table 8-2): (a) the inclusion of registry data for trabectedin by NICE as historical controls; (b) different primary endpoints for mifamurtide (“overall survival” for NICE and “progression-free survival” for SMC, TLV and HAS); (c) the secondary endpoint “severe bleeding events” for eltrombopag only reported by NICE;

(d) the lack of HRQol data in the assessment of eltrombopag for HAS; and (e) different economic models for eltrombopag.

Table 8-2. Cases when differences at each step of the HTA process explain differences in HTA recommendations (Nicod, 2016a).

Drug and indication pair		Eltrombopag	Imatinib	Mannitol dry	Mifamurtide	Trabectedin
		Thrombocytopenic purpura	*Gastro intestinal stromal tumours (adjuvant, unresectable and/or metastatic)*	*Cystic fibrosis*	*Osteosarcoma*	*Soft tissue sarcoma*
HTA recommendation	**Positively appraised** (list or restricted)	SMC, TLV, HAS (ASMR II)	SMC, HAS (ASMR III)	NICE, HAS (ASMR V)	NICE, SMC, TLV	NICE, HAS (ASMR V)
	Rejected	NICE	NICE	SMC	HAS	SMC
Evidence	**Differences in the level of evidence reported**	**X** Severe bleeding events (WHO grade 3-4) (NICE) **X** Lack of qol data (HAS) Qol data included for NICE, SMC, and TLV **X** CUA-standard care (NICE) ✔ CUA-romiplostim (SMC) ✔ CMA-romiplostim (TLV)			✔ Progression-free survival = primary endpoint (SMC, TLV, HAS) **X** Overal survival = primary endpoint (NICE)	✔ Use of registry data as historical controls (NICE)
Interpretation of the evidence	**Different interpretation of the same evidence appraised**		**Short trial duration** **X** NICE, SMC Not raised by HAS	**No reduction in hospital days and use of antibiotics** **X** HAS Not raised by SMC, NICE **Qol not improved** **X** HAS ✔ NICE Not raised by SMC		
	Different interpretation of a same uncertainty	**Short trial duration** **X** NICE (experts), SMC, TLV ✔ HAS (same as comparator)	**Overall survival not significantly improved** **X** NICE ✔ SMC (orphan) ✔ HAS (on-going trial)	**Risk of bronchospasms** **X** HAS ✔ NICE (expert opinion) Not raised by SMC	**Risk of interaction between treatments** X HAS (other study) ✔ NICE, SMC (expert opinion), TLV (longer term data)	**Lack of comparative evidence (phase II non-comparative pivotal trial)** **X** HAS ✔ NICE (rarity, early marketing autorisation, historical controls) ✔ SMC (rarity, investigational nature of the treatment)

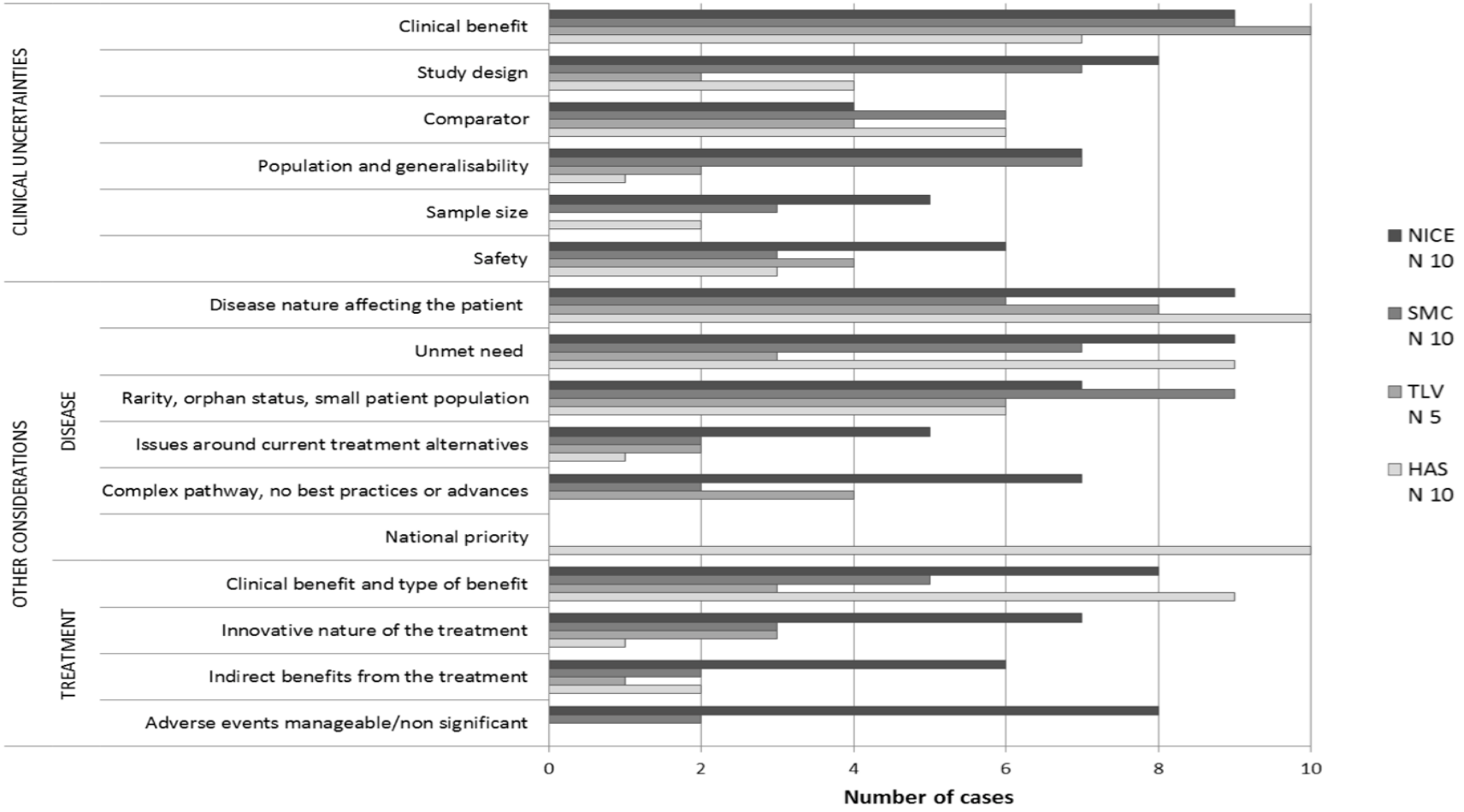
Legend: NICE: National Institute for Health and Care Excellence; SMC: Scottish Medicines Consortium; TLV: Pharmaceutical Benefits Baord; HAS: Haute Autorite de Sante

8.4.2. Interpretation of the evidence

In total, uncertainty was identified 124 times (N_u) in the HTA reports and grouped into ten categories depending on the type of concern raised. Similarly, 125 individual "other considerations" (N_{oc}) were identified and grouped into 16 categories (Figure 8-1).

Although the null hypothesis of independence was rejected (χ^2=18.80; p=0.4040), correspondence analysis gives an indication about the existing relationships amongst the variables analysed, providing insights about the agencies' risk preferences for these ten drugs (Figure 8-2). NICE was relatively more concerned about population generalizability compared to HAS, which was more concerned about issues related to the treatment's administration and provision. In contrast, SMC was relatively more concerned about population generalizability and the treatment's benefit, and HAS about safety and issues the design of the evidence presented. Conducting the same analysis across the five study drugs appraised by all agencies, a non-significant association between variables was seen likely due to the small sample size (χ^2=27.95; p=0.3451). Nevertheless, similar results were seen with additionally NICE being relatively more likely concerned about sample size, HAS with the duration of the study, and TLV about the treatment's administration and provision.

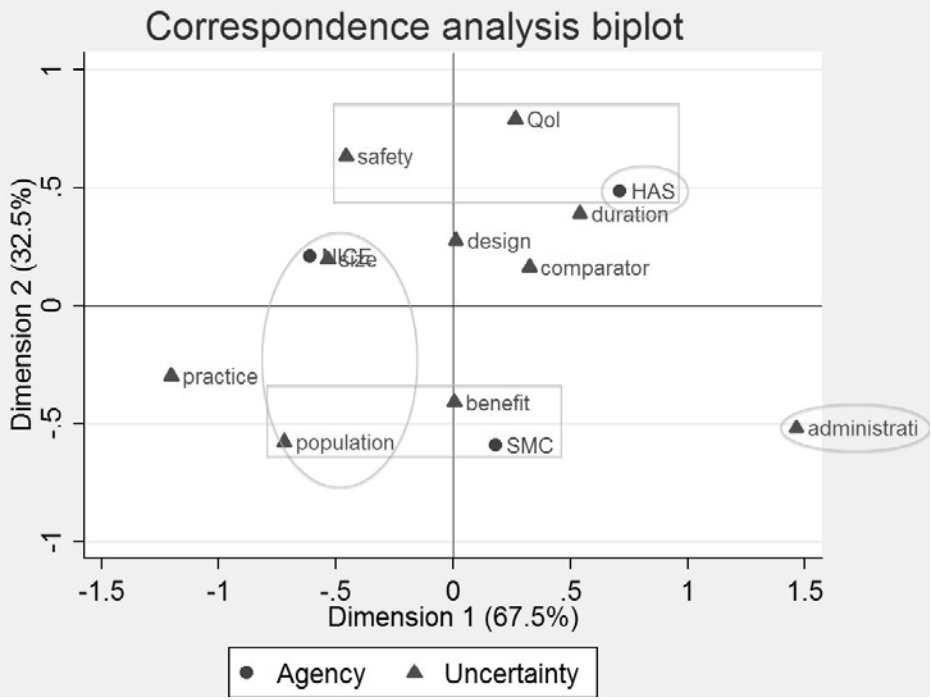
Figure 8-1. Number of drugs where clinical uncertainties and "other considerations" were identified.



Source: (Nicod, 2016a).

This Figure illustrates the number of cases where clinical uncertainties and “other considerations” were identified influencing the decision process in each country. In total 124 clinical uncertainties were identified across the 35 country drug-indication pairs grouped into 10 categories, and 125 “other considerations” grouped into 16 categories. The latter 16 categories were further distinguished between those that relate to living with the disease in question, from those to taking the treatment. The representation of each group was ordered such that the more frequently identified clinical uncertainty, disease-related “other considerations” or treatment-related “other considerations” are represented at the top of the graph.

Figure 8-2. Correspondence analysis biplot illustrating relative associations between the HTA bodies and the issues (clinical uncertainty) raised.

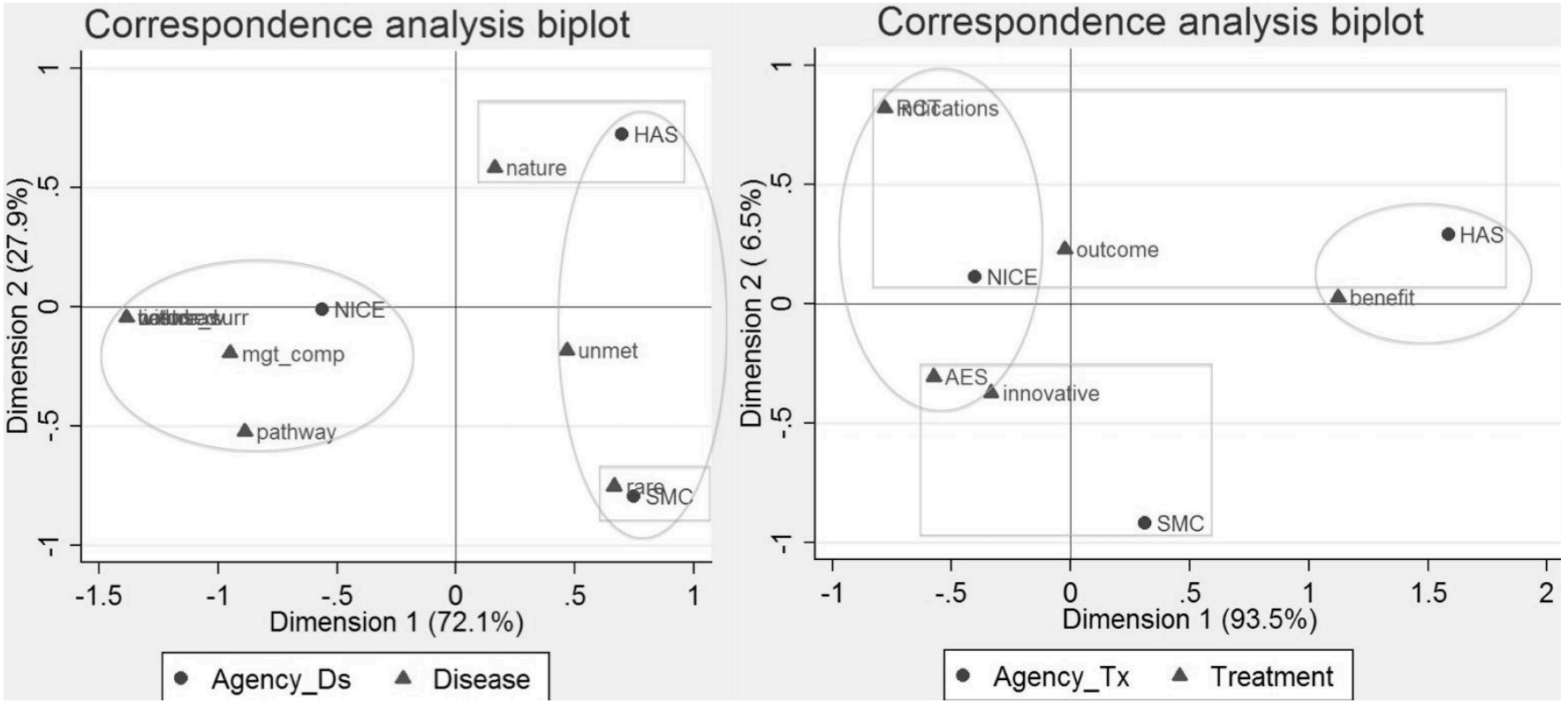


Source: (Nicod, 2016a).

This figure represents the correspondence analysis biplot illustrating the relative associations between the HTA bodies and the clinical uncertainties raised by each HTA body. Although the null hypothesis of independence was rejected (χ^2=18.80; p=0.4040), it provides an indication about specific risk preferences. On the horizontal axis (67.5% of the variation), NICE is more likely concerned about population generalizability and conformity to clinical practice than HAS and SMC, who are more concerned about issues related to the treatment's administration and provision, and the duration of the trial. On the vertical axis (32.5% of the variation), SMC was more likely concerned about population generalizability and the treatment's benefit, and HAS about quality of life improvement and trial design.

Legend: Qol: quality of life; safety: safety assessment; design: trial design; comparator: comparator; duration: duration of the trial; administration: administration and provision of the treatment; benefit: benefit of the treatment; size: sample size; population: population generalizability; practice: clinical practice.

Figure 8-3. Correspondence analysis biplot illustrating the relative associations between the HTA bodies and disease (left) and treatment characteristics (right).



Source: (Nicod, 2016a).

The figure to the left represents the statistically significant relative associations between the HTA bodies and disease characteristics (χ^2=40.05; p=0.0008). On the horizontal axis (72.1% of the variation), NICE was more likely to account for existing treatment alternatives, clinical practice, and the impact of the disease on the patient's surrounding, whereas SMC and HAS for rarity and unmet need. On the vertical axis (27.9% of the variation), HAS was more likely to value the nature of the disease, and SMC the rarity of the condition.

The figure to the right illustrates the significant relative associations between the HTA bodies and treatment characteristics (χ^2=29.46; p=0.0011). On the horizontal axis (93.5% of the variation), NICE was relatively more likely to value the treatment's safety and challenges in conducting RCTs, and HAS the drug's clinical benefit compared to other agencies. On the vertical axis, relationships were relatively less meaningful given that only 6.5% of the variation was captured.

Legend: rare: rarity, small sample size, orphan drug; unmet: unmet need; nature: nature of the condition and its impact on the patient; pathway: complex pathway, no best practice; alternative: issues around current alternatives; cost: cost burden of current treatment alternatives; nature_surr: disease nature affecting the patient's surrounding; mgt_comp: issues around the management of treatment alternatives; benefit: clinical benefit and type of benefit; indirect: indirect benefits from the treatment; innovation: innovative nature of the treatment; AEs: adverse events from the treatment manageable or non-significant; RCT: challenges in conducting RCTs; indications: additional indications of treatment.

Focusing on preferences relating to disease characteristics, the relative associations with NICE, HAS and SMC were significant across the 10 study drugs (χ^2=40.05; p=0.0008) (Figure 8-3). NICE was relatively more likely to account for existing treatment alternatives and clinical practice, as well as the impact of the disease on the patient's surrounding, whereas SMC and HAS were more likely to value rarity and unmet need. HAS was relatively more likely to value the nature of the disease, while SMC the rarity of the condition. Conducting the same analysis across the five drugs appraised by all agencies, associations were statistically significant (χ^2=47.37; p=0.0008). Preferences for NICE, SMC and HAS were similar, where TLV was relatively more likely to value the nature of the condition (e.g. disease severity).

Correspondence analysis examining relative value preferences around treatment characteristics and NICE, SMC and HAS for the 10 study drugs showed a significant association between these variables (χ^2=29.46; p=0.0011) (Figure 8-3). NICE was relatively more likely to value the treatment's safety and challenges in conducting RCTs, and HAS the drug's clinical benefit compared to the other agencies. Conducting the same analysis across the four drugs appraised by all four agencies, similar conclusions were reached (χ^2=21.05; p=0.0496). Additionally, TLV was relatively more likely to value the innovativeness of the treatment compared to the other agencies.

These risk and value preferences identified across the ten study drugs may have influenced these processes and contributed to explaining cross-country differences. Examining each of the concerns more in depth, only 14.5% of the uncertainties identified (18 of the N_u=124) were commonly raised by all agencies, the remainder having been raised by only one or some of the agencies. This was further highlighted in the poor to less than expected by chance agreement measured between agencies in the clinical uncertainties raised about the same evidence (κ range -0.30 to 0.08) (Table 8-3).

In four cases, this poor level of agreement in interpreting the same evidence related to one of the main reasons for the final decision (Table 8-2). For imatinib, the primary trial length was deemed too short by NICE and SMC; this was not highlighted by

HAS. Additionally, the secondary endpoint “overall survival”, considered as main parameter of interest by NICE, was not significantly improved negatively influencing the decision (e.g. imatinib was rejected by NICE). For mannitol dry, the lack of improvement in hospital days and antibiotic use reduction was a concern for HAS, but not for NICE or SMC. Further, the lack of improvement in HRQol negatively influenced HAS’s decision (e.g. ASMR V). This concern was also raised by NICE, who acknowledged that the current measures used do not fully capture the effect of the disease and treatment; this was not highlighted by SMC.

Agreement between two agencies was reached if a concern was considered addressed or not by both, and disagreement if addressed by one and not the other. There was agreement for 13 and disagreement for five of the 18 concerns commonly raised. When comparing agreement across each pair of countries, it varied ranging between moderate to lower than expected by chance depending on the agencies (κ range -0.50 to 1.0) (Table 8-3).

Table 8-3. Agreement between HTA bodies in the uncertainty raised about the same evidence (raised versus not raised); and when the same uncertainty was raised, agreement about how it was dealt with (addressed versus not addressed).

Kappa scores [95% confidence intervals] Standard error (SE) Number of observations (n)	Level of agreement in the uncertainties raised (raised versus not raised)				
Level of agreement in interpreting the same uncertainties (addressed versus not addressed)		**NICE**	**SMC**	**TLV**	**HAS**
	NICE	1	**-0.06** [-0.235-0.124] SE=0.091 n=117	**-0.15** [-0.434-0.143] SE=0.147 n=44	**0.01** [-0.172-0.183] SE=0.090 n=110
	SMC	**0.51** [0.203-0.814] SE=0.156 n=29	1	**-0.30** [-0.588--0.018] SE=0.145 n=43	**0.08** [-0.108-0.261] SE=0.094 n=110
	TLV	**1.00** [1.00-1.00] SE=0.00 n=7	**0.72** [0.232-1.00] SE=0.249 n=7	1	**-0.07** [-0.324-0.180] SE=0.128 n=44
	HAS	**-0.08** [-0.227-0.067] SE=0.075 n=24	**0.18** [-0.272-0.630] SE=0.230 n=22	**-0.50** [-1.00-0.235] SE=0.375 n=4	1

Source: (Nicod, 2016a).

Cohen's kappa scores (κ) rank agreement levels from poor (κ = 0) to perfect agreement (κ =1) and where minus values of κ correspond to cases when agreement was less than expected by chance.

Legend: NICE: National Institute for Health and Care Excellence (NICE); SMC: Scottish Medicines Consortium; TLV: Dental and Pharmaceutical Benefits Board; HAS: Haute Autorité de Santé.

Between 5% and 51% of these clinical uncertainties (N_u=124), depending on the country, were addressed through various means (51% of n_u^{nice}=68 uncertainties for NICE; 12% of n_u^{smc}=60 for SMC; 47% of n_u^{tlv}=21 for TLV; and 5% of n_u^{has}=44 for HAS). First, stakeholder input was used to confirm the plausibility of a (uncertain) clinical claim. Second, the uncertainties were raised but nevertheless considered acceptable by the Appraisal Committee. Third, greater uncertainty was accepted given the rarity of the condition or accounting for non-primary evidence.

In three cases, differences in the interpretation of evidence were also one of the main reasons for the final recommendation (Table 8-2). Two of these were based on expert opinion: the risk of bronchospasms was deemed minimal by NICE clinical experts for mannitol dry, and the risk of interactions with other treatments was deemed minimal by clinical experts from NICE and SMC for mifamurtide. In one case (trabectedin), the lack of comparative data for the primary phase II trial was a concern for all but was addressed differently. It was deemed acceptable given the rarity of the condition and investigational nature of the treatment by NICE and SMC; additionally NICE accounted for registry data as historical controls; in contrast, it was not deemed acceptable by HAS.

A number of other considerations were also put forward by the agencies as one of the reasons for the final recommendation, and associated with differing final outcomes. In a number of cases, greater flexibility was granted to the ICER on the basis of the following considerations: (a) SMC modifiers relating to orphan drugs at SMC (5 out of 10 drugs), (b) national priority for rare diseases by HAS (all 10 study drugs), (c) NICE end-of-life supplementary advice (4 out of 10 drugs) (NICE, 2009c), (d) patient access schemes at NICE (7 out of 10 drugs) and SMC (3 out of 10 drugs), or (e) disease severity at TLV (all five study drugs). In particular, four drugs fulfilled the NICE end-of-life criteria, where three were considered cost-effective with an ICER lower than £50,000/QALY (lenalidomide, azacitidine, trabectedin), and one not cost-effective with an ICER greater than £50,000/QALY (everolimus). Similarly, the high ICERs were accepted by SMC for lenalidomide and azacitidine given the SMC modifiers, and by TLV for lenalidomide given the severity of the disease.

8.4.3. Reasons for different HTA recommendations

Through the application of this methodological framework, differences at each stage of the HTA process were identified partly explaining differences in HTA recommendations (Table 8-2).

First, heterogeneity was seen in the evidence accounted for (e.g. parameter of interest, economic model and comparator, non-primary evidence) partly explaining the different HTA outcomes in 50% of the six cases with diverging recommendations (e.g. eltrombopag, mifamurtide, trabectedin). A further contrast was seen in the main parameters of interest considered for eltrombopag and romiplostim despite treating the same condition, which were bleeding events (SMC, NICE) and platelet response, respectively.

Second, the diverging interpretation of the same evidence (e.g. trial duration, HRQol improvement, reduction in resource use) partly explained differences in HTA outcomes for 33% of the six cases (e.g. imatinib, mannitol dry).

Third, the different ways identified in dealing with the same uncertainty (e.g. trial duration, clinical benefit, safety, comparative data) also contributed to partly explaining differences in HTA outcomes for 66% of the six cases (e.g. imatinib, mannitol dry, mifamurtide, trabectedin).

Some of the differences in interpretation were likely to be a consequence of the agency-specific risk or value preferences identified, as well as the willingness to account for stakeholder input. For example, correspondence analysis identified for HAS a relative risk preference about issues relating the treatment's administration and provision. This was seen for mannitol dry, where HAS was the only one concerned about the lack of reduction in hospital days and antibiotic use from taking the treatment. Similarly, value preferences around the rarity of the condition and the innovativeness of the drug were identified for SMC. This was reflected for trabectedin, where the lack of comparative evidence was deemed acceptable by SMC given the rarity of the condition and investigational nature of the treatment. Finally,

accounting for stakeholder input also explained some of these differences as seen for mannitol dry (NICE) and mifamurtide (NICE, SMC).

Fourth, there were factors that contributed to modulating the ICER or uncertainty to an acceptable level, further explaining differences. These included: disease severity for TLV, end-of-life criteria for NICE, the ability to implement Patient Access Schemes or lower discounting rates, imposing a restriction, or by imposing future re-assessments. For example, a higher ICER was accepted for everolimus by TLV because of its severity, but was considered too high by NICE and SMC and rejected. In the case of lenalidomide, uncertainty was addressed by imposing a restriction to third line treatment (SMC, NICE), or a re-assessment in a near future once more evidence is collected (TLV). Another modulating factor was the ability to implement a lower discounting rate on costs and effects captured in the model, as was seen for mifamurtide by NICE and SMC, whereas the high ICER was acceptable for TLV given the severity of the condition, and was rejected by HAS for the reasons discussed in the next paragraph.

A final contrast was seen when assessing cost-effectiveness versus clinical benefit, resulting in opposite conclusions. Indeed, a number of compounds rejected by NICE or SMC received: (a) an important SMR rating (translating into a 65% to 100% coverage rate): eltrombopag, everolimus or imatinib; and (b) a high ASMR rating (I-III associated with a more favourable pricing scheme): eltrombopag, imatinib. In these cases, the negative recommendations issued by NICE and SMC were mainly because of the high ICER, which also relied on the parameter of interest included in the model (e.g. bleeding events, overall survival). There were also drugs that were positively appraised by NICE or SMC and received very low SMR ratings (moderate (30%) and weak (15% coverage)) and an ASMR V, or were considered insufficient and rejected for coverage: mannitol dry, ofatumumab, mifamurtide. This was because of the lack of comparative data as a result of the early marketing authorisation granted (ofatumumab) and early scientific advice received (mannitol dry), or the highly uncertain evidence presented (mifamurtide). Mannitol dry and mifamurtide also had in common that they were the only two drugs that were not part of the temporary authorisation scheme (ATU) in France, which may be considered a modulating factor

to accepting greater uncertainty when data is being continuously being collected, although results from our limited sample size are insufficient to affirm this.

8.5. Discussion

In summary, this study applied an existing methodological framework to systematically identify and compare HTA decision processes in different settings. Results show that despite its aims in being "transparent, unbiased, robust and systematic" (European Commission, 2015a), there is important heterogeneity in the evidence appraised, its interpretation and the extent to which this influenced the final recommendation during the HTA processes. HTA remains a flexible instrument that is subject to the decision-maker's interpretation about uncertainty and social values, as part of the deliberative process of HTA and in support of existing tools to assess uncertainty (e.g. sensitivity analyses, elicited social values). This study shows that nevertheless, the interpretation may vary based on the decision-maker's own interpretation, observed agency-specific risk and value preferences, or the ability to modulate the ICER or estimate of clinical benefit to an acceptable level, possibly explaining some of the variation in HTA recommendations across countries. These differences may relate to contextual differences or to controversies over the HTA process itself. Identifying and raising awareness around these at each stage of the process is a way forward towards furthering the debate on HTA and its application. Three areas are discussed here, the methodological challenges of HTA, context-specific considerations in these processes, and issues related to the rarity of these conditions and how they are being dealt with. The limitations in this study are then described.

These findings are also relevant for collaborative initiatives such as the EU-level HTA assessments piloted within EUnetHTA, as they contribute to understanding areas where disagreements may arise and clarifying possible solutions to deal with them based on past experiences. An illustrative example was seen for trabectedin and the circumstances under which the use of historical controls was accepted. The retrospective identification of the criteria driving previous decisions, applied in this study, is also recognised as one approach to criteria elicitation for MCDA purposes

when used for priority setting, according to the ISPOR MCDA Emerging Good Practices Task Force. The criteria elicited by the EVIDEM (Evidence and Value: Impact on Decision Making), also for the purpose of MCDA in priority setting, are very similar to those identified in this study, which can contribute to the definition and/or validation of these sub-criteria. For example, unmet needs were categorised as unmet need in efficacy, in safety, in patient-reported outcomes, and patient demand (EVIDEM, 2015). Findings from this study identified the different ways HTA bodies refer to unmet need, such as the importance of having new treatment options, the lack of (satisfactory) treatment alternatives, alternative treatments not being routinely available, the need to improve therapeutic management, and so forth. The distinction about the existence of alternatives, whether they are satisfactory, and the need for improved care despite current practices should be accounted for when referring to unmet need. Results can further contribute to defining the attributes of the criteria being accounted for by EVIDEM or other criteria elicitation processes (e.g. MCDA, Discreet Choice Experiments), such as disease severity, type of benefit, effectiveness, safety and HRQol, cost and cost-effectiveness, and so forth.

8.5.1. HTA methodological challenges

RCT weaknesses are well known and include limitations around safety and generalisability to heterogeneous populations or clinical practice, as well as the cost to conduct these (Rawlins, 2008). These limitations are compatible with our findings where, for example, generalisability to clinical practice was often a concern for NICE (e.g. azacitidine, eltrombopag, everolimus, mifamurtide) and SMC (e.g. eltrombopag, lenalidomide and romiplostim). Other issues relating to the heterogeneity of the trial population comprised the: (a) non-inclusion of certain patient groups (e.g. SMC for azacitidine) or subgroup heterogeneity (e.g. NICE, SMC and HAS for mannitol dry), (b) trial population non-representative of the indication under review (e.g. NICE for eltrombopag), or (c) imbalances in the characteristics or responses across the different subgroups (e.g. SMC for azacitidine and NICE for imatinib).

Given that preference for RCTs was seen in the primary trials appraised, the above concerns emphasise the need to recognise complementary forms of robust and valid

evidence (Rawlins, 2008). Apart from a few cases (e.g. expert opinion to confirm generalizability), this was not seen in practice given the limited role of non-phase III evidence observed in this study. The uptake of such forms of evidence is still modest and likely due to the lack of expertise around dealing with a variety of types of observational evidence including those based on real world data such as electronic patient records, (Berger, Martin, Husereau, Worley, Allen, Yang, Quon, Mullins, Kahler, & Crown, 2014) or patient-reported outcomes (McClimans & Browne, 2011). Their role, however, is to be stressed given their potential use for policymaking in, for example, the value-based system or process for Highly Specialised Medicines at NICE, the Patient and Clinician Engagement (PACE) programme at SMC, the use of managed entry agreements (Ferrario & Kanavos, 2013), and, more recently, the introduction of a pilot study on adaptive licensing at EMA level (Eichler, Oye, Baird, Abadie, Brown, Drum, Ferguson, Garner, Honig, Hukkelhoven, Limn, Lim, Lumpkin, Neil, O'Rourke, Pezalla, Shoda, Seyfert-Margolis, Sigal, Sobotka, Tan, Unger, & Hirsch, 2012; European Medicines Agency, 2014). With these new developments, the environment is shifting towards increasingly relying on expert opinion, observational studies and real world data (Doward, Gnanasakthy, & Baker, 2010), which could provide insights about treatment effectiveness, the burden of illness, the nature of a condition, or the indirect health care costs and benefits from taking the treatment.

There is also a need for a more formalised and consistent recognition of this type of evidence, which could be achieved by generating criteria for their acceptability based on past decisions such as the specific circumstances (e.g. early marketing authorisation) or quality standards (e.g. reliability, validity) required. For example, findings for ofatumumab and trabectedin suggest that comparative evidence is a crucial component for HTA when measuring clinical benefit (particularly for HAS) and that, under certain circumstances, historical control data could be acceptable as a proxy.

Progression-free survival is increasingly being used as the primary endpoint within trials, instead of overall survival (Booth & Eisenhauer, 2012). There is continued disagreement as to whether progression-free survival is the appropriate measure to capture a meaningful improvement for patients, and only very few disease areas have

been able to prove the surrogacy of progression-free survival to overall survival. (Booth & Eisenhauer, 2012). This disagreement was reflected in our results, where progression-free survival was accepted as the main parameter of interest by some and not by others, who only accept overall survival, further explaining differences (e.g. imatinib and mifamurtide). Similar scenarios were seen for other clinical endpoints, where the parameter of interest for two different drugs treating the same condition in the same country differed (romiplostim, eltrombopag).

8.5.2. Context-specific considerations

Context-specific considerations throughout the HTA process are legitimate and reflect different willingness-to-pay thresholds, cost and modelling considerations, perspectives adopted in the assessments (e.g. societal perspective), or agency-specific risk and value preferences. Results highlight those cases when these influenced the appraisal processes explaining differences in recommendations.

The economic models considered were somewhat comparable with one exception (eltrombopag). Nevertheless, they resulted in different outcomes in four cases: everolimus, imatinib, trabectedin, and mannitol dry. This was a consequence of different modulating factors rendering the high ICER acceptable. These included: (a) disease severity for TLV, (b) SMC modifiers, (c) the ability to implement patient access schemes, (c) NICE end-of-life criteria, (d) the restriction to a subgroup population for which the drug is more cost-effective, or (e) the ability to impose a future re-assessment rendering uncertainty more acceptable. The first three reflect different willingness-to-pay thresholds and special considerations for orphan drugs, while the latter cases suggests the ability to go beyond the ICER in order to identify circumstances or subgroups for which the treatment is cost-effective, or accept greater uncertainty for a limited period of time until more evidence is generated.

The ability to implement patient access schemes is another way to improving the cost-effectiveness by improving some of the uncertainty (Towse, 2010), and providing earlier access to these treatments (Russo, Mennini, Siviero, & Rasi, 2010). However, their effects on innovation and expected returns are still unclear (Ferrario et al., 2013),

and a number of issues around their implementation have been already noted (Boggild, Palace, Barton, Ben-Shlomo, Bregenzer, Dobson, & Gray, 2009).

In terms of the societal perspective adopted by TLV, there was no clear differentiation in the results around risk and value preferences, or in the type of evidence appraised that reflected societal preferences. Societal risk and value preferences include, for example, the ability to contribute to society after taking the treatment or the impact of the disease on the patient's family and surrounding. As an example, these were accounted for by NICE for mifamurtide, but not explicitly by TLV.

8.5.3. Issues related to the rarity of the conditions

The nature of the evidence presented, e.g. small trials, phase II primary trials, lack of comparative data, subgroups, surrogate endpoints, is similar to what is commonly seen for orphan drugs, characterised by more uncertainty because of the issues in generating high quality evidence due to their rarity (Bell & Tudur Smith, 2014; Kesselheim et al., 2011). This was clearly reflected in the number of issues highlighted by the HTA bodies that relate to rarity and the orphan status (e.g. small sample size, insufficient statistical power). In some cases, these were considered acceptable because of the condition's rarity or the recognised difficulties in recruiting sufficient patient numbers in trials (e.g. TLV for eltrombopag, NICE for mifamurtide and romiplostim). In contrast, the concerns relating to population subgroups often remained inconclusive because of their lack of statistical power or retrospective nature (e.g. azacitidine or mannitol dry). When comparing the prevalence rates used by SMC in their budget impact analysis and the HTA recommendations issued, two observations arise. The three drugs treating less than 20 patients per year (ofatumumab, mifamurtide, trabectedin) had generally poorer outcomes: they all received the poorest ASMR (V) rating, and were more likely to be rejected by the other agencies (ofatumumab by all, trabectedin by SMC). This was a consequence of the lower quality of the evidence from small sample sizes or the lack of comparative data. In the "more prevalent" rare conditions analysed (between 200-300 patients per year in Scotland), similar issues were encountered but to a lesser extent were these linked to the small sample size (eltrombopag, mannitol dry). These experiences could be a good starting point to

generate the circumstances under which small sample sizes or other issues specific to rare diseases may be acceptable due to the rarity of the condition, also ensuring these are accounted for consistently across cases.

Rare diseases are often characterised by greater unmet need, and therefore few or no treatment alternatives are often available. It also entails that little is known about the natural progression of these conditions and what the appropriate comparators should be, as illustrated in our sample where 30% of cases did not include any comparative data and the remainder mainly relied on comparisons with standard care. This had a negative influence in France because of the lack of comparative data, given that the ASMR assessments rely on the treatment's additional clinical benefit.
The high number of validated or non-validated surrogate endpoints identified in this study (80% of cases) is again representative of orphan drug characteristics and were dealt with discordantly raising questions about the selection of the appropriate endpoints and whether they reflect what brings value to the patient. This was illustrated when dealing with progression-free-survival and overall survival, where NICE always seems to prefer the latter and the others tend to account for the trial's primary endpoint. Another subtlety was seen for eltrombopag and romiplostim, which treat the same condition and where the main parameter of interest differed (e.g. bleeding events and platelet count).

The limited number of cases that included evidence about HRQol improvement also raises the question as to whether the measures used were sufficient to capture the full effects of the treatment to the intended population. These are important considerations for rare diseases, given that the issues around validated surrogate endpoints (Miyamoto & Kakkis, 2011) and availability of HRQol data (Price, Klaassen, Bolton-Maggs, Grainger, Curtis, Wakefield, Dufort, Riedlinger, Soltner, Blanchette, & Young, 2009) are more likely to be a concern in these cases, further emphasising the need to account for other levels of evidence.

Finally, given that orphan drugs are characterised by greater uncertainty, there were instances (TLV and HAS) where greater uncertainty was accepted because of continuous generation of evidence and planned re-assessments. Given that orphan

drugs are characterised by greater uncertainty due to the small patient populations they treat, it is all the more important to ensure that coverage decisions do not only rely on one assessment at one time point, but that the proper incentives are implemented for continuous data collection and assessment.

The above experiences could be a good starting point to generate the circumstances under which small sample sizes or other issues specific to rare diseases may be acceptable due to the rarity of the condition, also ensuring these are accounted for consistently across cases.

8.5.4. Limitations and need for further research

This research is not without its limitations.

First, the data was mainly collected from secondary sources. It would have been preferable to have full information about the submissions (e.g. manufacturer submission), but this was not possible in the current scheme. The information obtained by applying the methodological framework was unavoidably limited by the level of detail provided in the HTA reports and whether the framework captures all aspects of the decision-making process (Nicod et al., 2015a). The information published was assumed to be transparent and reflect the main determinants driving the decisions (Transparency Directive). The analysis of these published documents were considered to provide sufficient detail and explain how decisions were reached. Validation of the findings was enhanced through a triangulation of the data collected with semi-structured interviews of HTA body representatives. The objective of the interview was to validate the findings that arose from the interpretation of the researcher and obtain additional insights about the different approaches to valuing orphan drugs, which may have not been captured by even the most complete documentation.

Third, sampling issues arising from differences among the four agencies in the way they select topics for their assessments. Despite these differences, a suitable sample was identified.

Fourth, this research focused specifically on orphan drugs, which undergo the same HTA process as drugs for more common conditions. Some of the findings may also be applicable to these more common conditions. One component of the analysis did focus on identifying those challenges that are specific, but not necessarily always unique to, dealing with these rarer conditions and draw key lessons from these.

A final limitation is the relatively small sample size, which does not allow for multivariate regression analysis. However, this research resulted in meaningful outputs derived from a more in-depth and qualitative component showing that differences across countries do matter. A more structured understanding of the possible explanations for differences were derived from the findings, allowing for subsequent more quantitative analyses to focus on certain aspects of the decision-making process across a greater sample. Further research could look at the drivers of these differences across a larger sample of drugs and therapy areas using multivariate regression analysis for a greater generalisation of the results, by extending it to other types of drugs to assess how different agencies assess different drug and disease characteristics. In order to maintain the depth and breadth of the analysis building on the methodological framework used in this study, it is highly recommended to begin by prioritising the qualitative strand to ensure that the depth of the processes are captured and comparable across settings.

8.6. Conclusions and policy implications

This research contributes to better understanding what matters in decision-making beyond the assessment of cost-effectiveness and clinical benefit, and explains the reasons for differences across countries. The added value is the approach used, in the application of an existing methodological framework enabling to identify and compare in a systematic and comprehensive manner these decision processes across settings.

The results of this study have implications for policy. First, they confirm the high level of uncertainty characterising orphan drugs from the types of concerns raised by the HTA bodies (e.g. small trials, short trial duration, issues with the trial design and

conduct, lack of comparative data, phase II primary trials, surrogate endpoints, subgroup data). Substantial differences in dealing with these were also observed, as emphasised by the poor level of agreement measured across countries when dealing with uncertainty, which in turn contributed to explaining differences in the HTA recommendations issued across countries. These consisted in the different levels of evidence accounted for and the different interpretations of this evidence, including how these were influenced by agency-specific risk and value preferences. There were also modulating factors that enabled to address uncertainty or modulate the ICER to an acceptable level, as well as differences depending on whether the assessment relied on cost-effectiveness or clinical benefit. Policymakers should be aware of this more comprehensive range of factors accounted for the decisions, as well as the different expressions of rarity in practice and the different ways to deal with the issues specific to -but not limited to- orphan drugs.

Second, given the limitations identified from dealing with –predominantly-experimental evidence (RCTs) and evidence from these rarer conditions, this study suggests that HTA processes may be insufficient. This was illustrated by the lack of clear pattern observed around how countries dealt with the issues identified. The reasons explaining differences were identified at each stage of the decision-making process, with contrasting trends within and across countries. Therefore, not only is more consistency and clarity needed within countries in the means used to address uncertainty or accept a greater ICER (e.g. “other considerations”, stakeholder input), but there is a need for new mechanisms to deal with uncertain evidence and high ICERs, often characterising these orphan drugs.

An important limitation of HTA, particularly with orphan drugs, is that the process often relies on the assessment of evidence at one particular time point (often shortly after marketing authorisation). This was quite clearly illustrated in this study, and given the difficulties to generate robust evidence and obtain significant point estimates with small patient populations, it is all the more important to introduce novel approaches to assessing these. This could be through a system of continuous data generation and assessment to reduce uncertainty over time, which requires proper incentives at European level at early stages to implement high quality registries and

processes that include re-assessments over time. This is already in place in some countries such as Sweden or France (under the ATU scheme), which has contributed to dealing with uncertainty in some of the cases evaluated without imposing additional conditions or restrictions.

A last implication is the misalignment observed between marketing authorisation and HTA, particularly in those three cases that received either early marketing authorisation or early scientific advice (about the trial design), which received generally poor acceptance rates due to the lack of comparative data or highly uncertain nature of the evidence produced. This further emphasises the need for a better alignment in the incentives in place as well as innovative approaches to continuously assessing these treatments.

The implications of this research are all the more important given the shift towards more niche markets and personalised medicine, where more and more of the treatments undergoing regulatory and coverage processes are characterised by some of the important issues discussed in this paper. This research is also topical and in line with the recognised need to better understand pricing and reimbursement systems through cross-country learning and sharing of experiences (Kaplan, Wirtz, Mantel-Teeuwisse, Stolk, Duthey, & Laing, 2013). It is also useful for European-level initiatives, such as the pilot for a common European HTA (EUnetHTA) as it helps better understand how different countries are dealing with value assessments and what the causes for variation are, including areas where HTA methods may be improved. It is a valid tool for policymakers and contributes to ensuring that resources are used efficiently, ultimately improving access to medicines.

9. How do Scientific and Social Value Judgments Influence HTA recommendations in England, France, Scotland, and Sweden?[7]

9.1. Abstract

This chapter explores how broader aspects of a treatment's value and the impact of the condition on patients not captured by routine HTA methods using clinical and economic evidence, defined as "other considerations", influence these HTA processes in different settings. Countries included were England, Scotland, Sweden, and France, and data sources were the publicly available reports on HTA recommendations. Ten drugs with EMA orphan designation for a specified indication and all appraised in England were selected. Qualitative thematic analysis was used to systematically identify and code all "other considerations" considered based on an existing methodological framework, which also coded how it influenced the decision and how it was provided. These pertained to the scientific assessment or to societal preferences, which were quantified or elicited, or non-quantified or non-elicited, respectively. On this basis, a classification framework was developed and used throughout the study. In total, 125 different "other considerations" were identified and grouped into 16 subcategories based on the information provided. Between 18% and 100% of these, depending on the agency, were put forward as one of the main reasons for the final decision potentially contributing to accepting a higher ICER or uncertain clinical benefit. Some of these were non-quantified or non-elicited and pertained to the assessor's judgment of the treatment's value (e.g. oral administration benefit, unmet need, innovativeness). Results were used to create a taxonomy of criteria accounted for as scientific or social value judgments that can be used in future cases to ensure their consistent use. Results also contributed to better defining the determinants of social value and to improving the lack of accountability for reasonableness. Identifying and understanding the scientific and social value judgments made provides a way forward to improving their transparency and consistency across decisions.

---

[7] A version of this chapter is under peer review with *International Journal for Technology Assessment in Health Care* (submitted June 2015)

9.2. Introduction

Health care decision-makers are responsible for resource allocation decisions with the primary objective to maximise health or social welfare in the whole population (Brouwer, Culyer, van Exel, & Rutten, 2008; Hurley, 2000), alongside other ambitions such as rewarding innovation. HTA helps make such decisions about whether to reimburse a new treatment by providing guidance on the efficient use of resources, ultimately, optimising access to patients. It relies on systematic approaches to appraising evidence about the value of using this treatment in terms of benefits (and costs) in real world settings, while including considerations of social, ethical and legal aspects to inform coverage for this technology (Luce et al., 2010). When both clinical and economic evidence are considered, the HTA outcome is most often the ICER that provides information about the cost per unit of effect from taking this new treatment compared to existing standards of care in real world settings. It provides "value for money" if the ICER is worth the investment and ranges within the payer's maximum WTP.

Routine HTA methods that rely on clinical (and economic) evidence may not adequately capture all the important considerations of a treatment's value and the impact of the condition on patients in real world settings (O'Donnell, Pham, Pashos, Miller, & Smith, 2009). This is partly because HTA is undertaken at the time of the treatment's launch onto the market when evidence is often incomplete or uncertain since real world evidence is generally not available. HTA bodies also tend to prefer experimental evidence collected within controlled environments (e.g. RCTs) (Rawlins, 2008), in spite of their limitations in capturing real world settings (Black, 1996). In such cases, scientific judgments about the reliability, generalisability and meaningfulness of this evidence in the clinical context are made (Rawlins, 2014; Rawlins & Culyer, 2004). Elicited societal preferences, referring to cases when society agrees to forego health in order to treat specific populations (e.g. preference to treating children), are also captured by routine HTA approaches (e.g. disease severity in Sweden, EQ-5D preference weights). Preference granted during the appraisal process may nevertheless have been granted despite it not having been previously elicited by the general population; these relate to the social value judgments made by

the assessors (Rawlins, 2014; Rawlins et al., 2004). These judgments are usually made as part of the deliberative process of HTA, during which experts and key stakeholders are consulted and the evidence is discussed until a decision is taken (Garau, Shah, Towse, Wang, Drummond, & Mason, 2009b). The main criticisms of this process is the lack of accountability for reasonableness given that there is not always a clear process to account for the inclusion of these other forms of evidence in the assessment process, as well as the lack of consistency in accounting for these "other considerations" (Daniels & Sabin, 2008; Earnshaw et al., 2008; Garau et al., 2009b; Schwappach, 2002).

Drugs used to treat rare conditions with an orphan designation are often characterised by high and uncertain ICERs, and likely not cost-effective according to standard WTP thresholds (Denis et al., 2010a; Drummond et al., 2007; McCabe et al., 2005). This is a consequence of the difficulties in producing robust evidence due to the small patient populations and the heterogeneity of these conditions, as well as their high prices (Clarke, 2006; Drummond & Towse, 2014b; Dupont et al., 2011; Simoens, 2011). In such cases, these reimbursement decisions rely on whether society is willing to forego health to the whole population in order to treat fewer patients with a rare condition (Drummond et al., 2014b).

Little evidence in support of a societal preference for rare conditions exists, and the few studies that attempted to elucidate this suggest the contrary when patients with more common diseases are deprived of treatment in order to treat fewer patients with a rare condition (Desser, Gyrd-Hansen, Olsen, Grepperud, & Kristiansen, 2010; McCabe et al., 2010; NICE Citizens Council Report, 2004; SMC, 2011a). In such cases, these decisions partly rely on the decision-makers' willingness to accept high and uncertain ICERs based on additional factors that influence their judgment of (scientific and social) value, such as, for example, disease severity, the treatment's orphan status, or to what extent uncertain evidence is acceptable (Dupont et al., 2011; Kanavos et al., 2012; Nicod, 2015; Simoens et al., 2011). They also rely on the flexibility of these processes in, for example, their ability to implement managed entry agreements or the availability of separate funding programs for certain conditions (e.g. Scottish fund for ultra-orphan drugs, Cancer Drug Fund in England). It is somewhat different in France,

where a procedure has been set up to expedite access to drugs for rare diseases, as a means to support development and dissemination of treatment for populations suffering from rare conditions.

The purpose of this study goes beyond the assessment of clinical and economic evidence into other areas that help explain value. We explore how broader aspects of a treatment's value and the impact of the condition on patients, not captured by routine HTA methods, influence these HTA processes in different settings. Particular focus was given to those cases with high ICERs or poor SMRs, in view of understanding which "other considerations" contributed to accepting higher ICERs or greater uncertainty. The subject of analysis was a sample of orphan drugs in four countries (England, Scotland, Sweden, and France), due to the likelihood of greater uncertainty in the evidence generated. We then examined whether the social value judgments revealed pertain to orphan drugs furthering the debate as to whether they have a preferential status.

9.3. Methods

9.3.1. Study sample

The methodological approach to HTA is built around pre-specified criteria, such as whether clinical and/or cost-effectiveness are considered and the context within they operate insofar as what is being captured (e.g. health service or societal perspective). On this basis, purposive sampling was used to select the study countries with the aim of having a good representation of different types of decision-making characteristics, in terms of: (a) the criteria used in the appraisal process; (b) the perspective adopted; and (c) any existing elicited preferences (Table 9-1). The HTA agencies included were: NICE in England, SMC in Scotland, TLV in Sweden, and the Transparency Committee of HAS in France. The three former focus their assessment on the treatment's clinical cost-effectiveness (NICE, 2008; Rawlins, Barnett, & Stevens, 2010; SMC, 2013b; TLV, 2013). HAS assesses the drug's medical benefit to inform whether it should be covered and at what rate (Service Medical Rendu (SMR)), and the relative improvement in medical benefit to inform the pricing negotiation

(Amelioration du Service Medical Rendu (ASMR)), ranking treatments in five levels. Both the drug's medical effectiveness (risk-benefit ratio) and its interest in terms of public health (ISP) are accounted for in the SMR assessment. No economic modelling was done by HAS at the time of the sampled drugs' appraisals. For cost considerations, NICE and SMC agencies adopt a health service perspective and TLV a societal perspective.

Ten drug and indication pairs, given that the HTA appraisal process focuses on one drug for one specific indication, were selected based on (a) whether they received an orphan drug designation by the European Medicines Agency, (b) were appraised by NICE through the Single Technology Appraisal process until December 2012, and (c) by at least two other of the study countries (Table 9-2).

9.3.2. Data collection and analysis

This empirical study applied an existing methodological framework enabling the systematic identification and comparison of the criteria driving HTA decisions for the same drugs in different countries through a mixed methods research design comprising three key elements: the evidence appraised, its interpretation and its influence on the final recommendation (Nicod et al., 2015a). The focus of this chapter is on the results from one of the components about the interpretation of the evidence. Specifically, the analysis focused on exploring the elements beyond cost-effectiveness, cost, effectiveness and safety that were raised by the HTA agencies and whether these played a role on the decision-making.

Thematic analysis was conducted to identify and code all the "other considerations" accounted for during the appraisal process and recorded in the study drugs' appraisal report(s). Bottom-up coding was performed, where codes were inductively created while examining the data to summarise what was put forward and categorise this data depending on the type of information provided (Onwuegbuzie et al., 2010). The section of text coded included all the text referring to the "other consideration". For example, if the assessors pointed out that very few treatment alternatives exist, this whole section of text would be coded as "few treatment alternatives". Codes were

then categorised into subcategories depending on the type of information provided, and recorded in a coding manual. These were considered as part of two clusters, those relating to living with the disease (disease characteristics) and those to taking the treatment (treatment characteristics). For example, “few treatment alternatives” were clustered under “unmet need” and considered as a disease characteristic. Coding was iterative and flexible to ensure transferability of codes to other drugs and countries, and additional codes were created with newly identified “other considerations”. Coding was conducted by the PhD candidate. Reliability and validity of the coding was tested by an academic colleague, who re-categorised each individual code into one of these. Where differences were observed, adjustments were made. Feedback from external experts were also received, this included the HTA bodies concerned, who have been presented most of this work, from HTA experts that are part of the Advance-HTA consortium, and from colleagues at different conferences.

Coding was performed vertically and horizontally. In the former, all “other considerations” were coded in a systematic manner as prescribed in the coding manual. The latter was implemented by double coding all “other considerations” to capture whether it was put forward as one of the main reasons for the decision, where the information came from (e.g. experts), and whether it was accounted for in the other countries. The data collected qualitatively was then quantitatively analysed to determine: (a) the type and frequency of “other considerations” accounted for; (b) cases when these were one of the main reasons for the decisions; (c) how they were provided (e.g. expert opinion); and (d) how they compared across agencies. The qualitative statistical software NVivo 10 was used for the data collection and analysis,(19) and Excel for further data analysis. Data sources consisted in the HTA reports publicly available from each HTA body, complemented with a review of the literature and input from key stakeholders, e.g. HTA bodies and HTA experts from the Advance-HTA consortium.

Each sub-category of “other considerations” were further explored to determine whether they are more likely to pertain to orphan drug and rare disease characteristics. Orphan drug and rare disease characteristics were identified from reviewing the literature.

9.4. Results

9.4.1. Value judgment classification framework

When the evidence appraised is uncertain or incomplete, scientific value judgments are made about the acceptability of this evidence. This includes about the uncertainty (e.g. reliability, generalisability, meaningfulness), the assumptions made (e.g. economic modelling) or about certain non-quantified considerations around disease and treatment characteristics. Societal preferences are also accounted for by HTA approaches. These pertain to giving preference to certain (non-quantifiable) aspects of living with a disease or taking a treatment, which are translated into prioritising certain groups of patients over others (Table 9-1) (Rawlins, 2014). These can be either elicited or not (Rawlins et al., 2004). The former are typically elicited by a group of representative citizens. In England, this formal process is conducted through NICE's Citizens Council (NICE Citizens Council, 2014). In Scotland, a societal preference for orphan drugs exists by means of the "SMC modifiers", which were defined by the SMC, input from clinical experts and Patient Interest Groups (SMC, 2011a, 2011c). In France, rare diseases are recognised as a national priority under the 2004 Law (Ministere des Affaires Sociales et de Sante, 2004). In Sweden, preference is given to the more severe conditions, for which a higher ICER is accepted. They define disease severity "on the basis of the relevant, initial condition and risk of permanent injury, ultimately death without treatment…. All the positive effects the medicine has on people's health and quality of life are accounted for" (TLV, 2012). Non-elicited preferences originate from the individual appraisal committee member's value judgment based on their experience or on what they believe society would prefer, and are usually made as part of the deliberative processes of HTA. These are referred to as social value judgments, judgments made about societal preferences. Within the scope of this study, these scientific and social value judgments are defined as "other considerations" (Table 9-1).

Table 9-1. Classification framework of scientific and social value judgments

HTA Body	Scientific assessment		Social or societal preferences		
	HTA criteria & perspective *-quantified-*	**Scientific value judgments** *-non-quantified-*	**Preferential status** *-elicited-*	**Orphan drug preferential status** *-elicited-*	**Social value judgments** *-non-elicited-*
England National Institute for Health and Care Excellence - NICE	ICER NHS & PSS perspective	As part of the deliberative process, judgment about the ***acceptability of uncertain or incomplete evidence***, including about the ***assumptions*** made (e.g. economic modelling), or about certain ***non-quantified*** considerations around treatment and disease characteristics. Examples: health-related quality of life, administration benefits, uncertain resource use, clinical pathways, discount rate, disease severity	End-of-life supplementary advice: life-threatening, small patient numbers, life-extending		As part of the deliberative process, giving preference to certain ***non-quantifiable*** considerations around treatment and disease characteristics when these have not been elicited from a representative population of citizens. Preference originates from the individual judgments of the appraisal committee based on their ***experience*** or on ***what they believe society would prefer*** or on ***conclusions of citizen's councils / juries.*** Examples: orphan status, unmet need, treatment innovativeness, disease severity
Scotland Scottish Medicines Consortium - SMC	ICER NHS & PSS perspective			SMC modifiers: life-threatening, life-extending, quality of life improvement, curative intent, unmet need	
Sweden Dental and Pharmaceutical Benefits Board - TLV	Human value, need and solidarity, ICER Societal perspective		Disease severity & unmet need		
France Haute Autorité de Santé - HAS	SMR & ASMR			Public Health Act 2004, recognising rare diseases as a national priority	

Source: (Nicod & Kanavos, 2016b)

Legend: ICER: clinical cost-effectiveness; NHS: National Health Service; PSS: Personal Social Services; SMR: Clinical Benefit; ASMR: Relative Improvement in Clinical Benefit

9.4.2. Study drugs and HTA recommendations

In countries that use the ICER, the WTP threshold is the amount above which a drug is not cost-effective unless certain pre-specified characteristics are fulfilled or the assessors are willing to accept this high ICER. No precise WTP threshold exists in England, but the ICER is considered within a maximum allowable range of £20,000 to £30,000 per QALY depending on the certainty of the evidence and whether quality of life and the treatment's innovativeness were appropriately captured, and £30,000 per QALY if a strong case is made (e.g. end-of-life treatment (NICE, 2009c), disadvantaged populations and children (Rawlins et al., 2010)) (NICE, 2008). Although no WTP threshold exists, SMC guidance notes that NICE's threshold may influence the assessment (SMC, 2011b). TLV does not have a fixed threshold but based on previous decisions, the average ICER accepted between 2002 and 2007 was EUR 36,000 per QALY and the highest granted was EUR 90,000 per QALY (Persson, 2012). In France, no threshold exists though a two-stage process is used where coverage relies on the clinical benefit (SMR) and the price negotiation uses the (relative) improvement in clinical benefit (ASMR).

The study included ten drugs for specific indications (Table 9-2). Five were not appraised by TLV because they were inpatient drugs and at the time of the study, TLV only appraised outpatient drugs. Based on the indicative cost-effectiveness thresholds, a number of drugs with an ICER greater than the acceptable range received a positive recommendation: mannitol dry, azacitidine, lenalidomide, mifamurtide, and trabectedin for NICE; azacitidine, lenalidomide, mifamurtide, and imatinib for SMC; everolimus, mifamurtide and romiplostim for TLV. In some instances, the ICERs for these products were improved by application of a Patient Access Scheme that provided a confidential discounted drug price. In France, where coverage is disconnected from the ICER and no threshold exists, only one case was rejected for reimbursement (mifamurtide), three drugs received an ASMR V where no additional benefit was recognised (ofatumumab, mannitol dry, trabectedin), and the remainder were considered to provide additional benefits.

Table 9-2. ICER & coverage decision

	NICE England		SMC Scotland		TLV Sweden		HAS France	
Drug Indication	**ICER**	**Decision**	**ICER**	**Decision**	**ICER***	**Decision**	**SMR**	**ASMR**
Eltrombopag Chronic thrombocytopenic purpura	✗ £104,000-£116,000/QALY (standard care)	**Reject**	✔CUA dominant compared to romiplostim (SMC modifiers)	**Restrict** (Subgroup severe ITP and high risk of bleeding)	✔CMA dominant compared to romiplostim (severity)	**Restrict** (Re-assessment, and for hospital	Important	**II**
Ofatumumab Chronic lymphocytic leukemia	✗ £50,300 - £81,500/QALY, depending on subgroup (PAS)	**Reject**	✗ £108,815/QALY	**Reject**			Moderate	**V**
Mannitol dry Cystic fibrosis	✗ £50-£80,000/QALY rhDNase ★< £30,000/QALY no rhDNase	**Restrict** (Subgroup with no rhDNase, rapid decline of lung function, intolerant to osmotic agents)	★£20,736/QALY no rhDNase	**Reject**			Weak	**V**
Everolimus Renal cell carcinoma (2nd line, advanced)	✗ £51,700/QALY (EoL, PAS)	**Reject**	✗ £61,330/QALY	**Reject**	★Cost/QALY high but justified given the severity of the disease (severity)	**List**	Important	**IV**
Azacitidine Myelodysplastic syndrome	✗ £47,200/QALY (best case scenario) (EoL, PAS)	**List**	✗ £51,275/QALY (SMC modifiers, PAS)	**List**			Important	**II**
Lenalidomide Multiple myeloma (2nd, 3rd line)	✗ two or more prior therapies: £41,300-43,800/QALY (chemo alone) (EoL, PAS)	**Restrict** (Subgroup 3rd line)	✗ £34,286-£41,381/QALY (chemo alone) (SMC modifiers)	**Restrict** (Subgroup 3rd line)	✔SEK290,000/QALY (bortezomib) = EUR 32,000/QALY (severity)	**List**	Important	**III**
Mifamurtide Osteosarcoma	✗ £36,000/QALY (1.5% discount, PAS)	**List**	✗ £48,579/QALY (1.5% discount, PAS)	**List**	★-✗ SEK 700,000-900,000/QALY = EUR 77-99,000/QALY (severity, 3% discount)	**List**	Insufficient	**DNL**
Trabectedin Soft tissue sarcoma	✗ £34,500/QALY (EoL, PAS)	**List**	★£36,841/QALY (PAS)	**Reject**			Important	**V**
Imatinib Gastro-intestinal stromal tumours (GIST) (adj. unresectable and/or metastatic)	★£21-£23,000/QALY (significant and moderate risk of recurrence)	**Reject**	★£20,655/QALY (SMC modifiers)	**Restrict** (Subgroup of patients with high risk of recurrence following complete resection)			Important	**III**
Romiplostim Chronic thrombocytopenic purpura	✔High risk of bleeding < £20,000/QALY slenectomised = £30,000/QALY non-splenectomised (PAS)	**Restrict** (Subgroup with high risk of bleeding, risk management plan)	✔High risk of bleeding: £15,220/QALY splenectomised £16,673/QALY non-splenectomised (standard care) (SMC modifiers)	**Restrict** (Subgroup with high risk of bleeding, 2nd line or when surgery is contraindicated)	★SEK 400-600,000/QALY = EUR 44-66,000/QALY	**Restrict** (Re-assessment & risk management plan)	Important	**II**
✔	Acceptable ICER: - within 20,000/QALY for NICE. - SMC: no threshold, but accounts for NICE threshold - TLV: no threshold, but based on previous decisions average of drugs approved is Eur 36,000/QALY							
★	Acceptable ICER accounting for other factors: - NICE: £20-£30,000/QALY - SMC: no threshold, but accounts for NICE threshold - TLV: no threshold, but based on previous decisions average of drugs approved is Eur 36,000/QALY, up to Eur 90,000/QALY							
✗	High ICER, likely not acceptable except if exceptional circumstances: - NICE: > £30,000/QALY (e.g. end-of-life treatment) - SMC: no maximum threshold, but accounts for NICE threshold - TLV: no maximum threshold, but based on previous decisions ICER greater than EUR 90,000/QALY							
	*1 SEK = 0.110202 EUR Legend: PAS: Patient Access Scheme; EoL: End-of-Life treatment; severity: disease severity considered high; QALY: quality-adjusted life years; ICER: incremental cost-effectiveness ratio; NICE: National Institute for Health and Care Excellence; SMC: Scottish Medicines Consortium; TLV: Dental and Pharmaceutical Benefits Board; HAS: Haute Autorite de Sante.							

Source: (Nicod & Kanavos, 2016b)

Legend: The grey-shaded cells are for when the HTA recommendations are in line with the known willingness to pay for an ICER, or in France, have been granted an ASMR V or a rejection. The non-shaded cells are for those HTA recommendations that were positive despite an ICER greater than the expected willingness to pay threshold.

9.4.3. "Other considerations": an overview

In total, 125 individual "other considerations" were coded and grouped into 10 categories (Figure 8-1). 94 of these 125 codes were included by NICE and used 173 times across all 10 cases (e.g. one "other consideration" may have been coded for more than one drug), followed by 24 codes used 67 times by HAS, 23 codes used 50 times by SMC, and 33 codes used 56 times included by TLV. The most common disease characteristics raised by all agencies were about the nature of the disease and its rarity, and the treatment's unmet need. The most common treatment characteristics included the type of treatment benefit (e.g. curative), the treatment's innovative nature (e.g. new class of drugs), its indirect benefit (e.g. such as the ability to go back to work from the treatment) and the non-significance of adverse events.

9.4.4. "Other considerations" as pivotal factors in the decision processes

A proportion of these 125 "other considerations" were also put forward by the HTA bodies as one of the main reasons for their decisions and identified through the double coding process. These represent 18% of the "other considerations" put forward by NICE (32 of 173), 24% by SMC (12 of 50), 34% by TLV (19 of 56), and 100% by HAS (67 of 67) (Table 9-3). For the purpose of HAS, these "other considerations" were mainly discussed in the conclusions of the Transparency Committee when assessing the ISP and consequently have all been considered as having been put forward together with the main reasons for the final recommendation.

A proportion of these (e.g. cases with a superscript in Table 9-3) pertained to those preferences elicited by each HTA body (Table 9-1), where higher ICERs or uncertain evidence may be accepted. Four drugs were eligible under the NICE end-of-life supplementary advice, three of which were considered cost-effective with an ICER ranging between £34,000-£47,000/QALY (lenalidomide, azacitidine, trabectedin), and the fourth (everolimus) not cost-effective with £51,700/QALY. Weaknesses in the economic model were deemed acceptable because of the SMC modifiers in four cases (eltrombopag, imatinib, azacitidine, lenalidomide). For HAS, all study drugs were recognised as targeting patients with rare diseases, and assessed within the framework

of one or more ministerial plans. In Sweden, the severity of the condition was put forward by TLV for all five cases and resulted in accepting higher ICERs.

Cases without a superscript in Table 9-3 represent the additional (non-quantified or non-elicited) "other considerations" put forward as one of the main reasons for the decision, which relate to the scientific and social value judgments made. For NICE, these included the treatment's unmet need for lenalidomide, mifamurtide and mannitol dry, its innovativeness for azacitidine and mifamurtide, and the severity of the disease for mannitol dry. Additionally, the impact on families' and friends' quality of lives, the rarity of the disease, and the ability to contribute to society, and live an active and fulfilling life were also put forward for mifamurtide. For SMC, these included the benefit from oral administration, the orphan status and unmet need for eltrombopag; the potential reduction in resource use for romiplostim; and the life-extending nature of the treatment for mannitol dry and azacitidine. Similarly, TLV also valued certain treatment characteristics, such as the benefit from oral administration (e.g. eltrombopag), the treatment's novel mechanism of action (e.g. eltrombopag, romiplostim), the impact of the disease on quality of life and daily activities (e.g. eltrombopag, romiplostim, lenalidomide), or the treatment's orphan status (e.g. eltrombopag). Unmet need was also recognised (e.g. eltrombopag, romiplostim), and in one case, TLV acknowledged the changing environment in clinical practice (e.g. lenalidomide). For HAS, both disease and treatment characteristics were put forward during the assessments for all drugs, namely around the nature of the disease and its effect on the patient, the need for treatment alternatives, as well as the type of direct or indirect benefit from taking the treatment. In France, orphan drugs are presumed to be innovative and thus subject to fast-track HTA consideration. In the assessment, the innovativeness of a drug is recognized for those drugs with ASMR I-III.

Table 9-3. “Other considerations” as pivotal factors in the decision (Nicod & Kanavos, 2016b)

	Eltrombopag	Romiplostim	Everolimus	Lenalidomide	Azacitidine	Imatinib	Mannitol dry	Mifamurtide	Ofatumumab	Trabectedin	Scientific value judgment *-non-quantified-*	Social value judgment *-non-elicited-*
National priority												
- Rare disease plan	HAS[rare]	HAS[rare]			HAS[rare]		HAS[rare]			HAS[rare]		
- Cancer plan			HAS[rare]		HAS[rare]	HAS[rare]		HAS[rare]				
- Plan for improving qol in patients with chronic diseases			HAS[rare]									
- Public Health Law 2004			HAS[rare]	HAS[rare]		HAS[rare]		HAS[rare]				
- Falls in the scope of the fight against cancer	HAS[rare]								HAS[rare]			
Issues around current treatment alternatives												
- changing treatment pathways				TLV							★	
Disease nature affecting the patient												
- Short life expectancy			NICE[end]	NICE[end], HAS	NICE[end], HAS					NICE[end]	★	
- Disease severity	HAS, TLV[severity]	TLV[severity]	TLV[severity]	TLV[severity]			NICE				★	
- Disease with a poor prognosis					HAS						★	
- Serious condition		HAS	HAS				HAS	HAS, TLV[severity]		HAS	★	
- Life threatening	TLV	TLV	HAS	HAS		HAS	HAS	HAS	HAS	HAS	★	
- Incurable				TLV			HAS				★	
- Requires life long treatment	HAS										★	
- Affects quality of life	TLV, HAS	HAS, TLV					HAS				★	
- Affects daily activities and functional capacity	TLV	HAS, TLV									★	
Disease nature affecting the patient's surrounding												
- Impact on quality of life of family and friends								NICE			★	
Rarity, orphan status, small patient population												
- Small patient population	TLV							NICE				★
- Minor public health burden because of rarity			NICE[end]	NICE[end]	NICE[end]	HAS	HAS			NICE[end]		
- Orphan status	HAS	HAS						HAS		HAS		
	SMC			SMC[modifiers]	SMC[modifiers]	SMC[modifiers]						
Unmet need												
- Importance of new treatment options							NICE					★
- Few developments in last years							NICE	NICE				
- No (satisfactory) alternatives exist	TLV	HAS	HAS			HAS			HAS	HAS		
- Alternatives exist				HAS			HAS	HAS				
- Need to improve therapeutic management									HAS			
- Few therapeutic options	HAS	TLV										
- New treatment would offer new options	SMC											
- Alternative treatments not routinely available				NICE								
Type of treatment benefit												
- Curative	HAS	HAS, SMC[modifiers]	HAS					NICE[discount], HAS	HAS	HAS		★
- Palliative				HAS	HAS							
- Preventive						HAS						
- Symptomatic							HAS					
- Salvage treatment									HAS			
- Life-extending			NICE[end]	NICE[end]	NICE[end], SMC[modifiers]	SMC[modifiers]	SMC	SMC[modifiers]		NICE[end]		
- Benefit extended over a long period								NICE				
Innovative nature of the treatment												
- Important advance					NICE							★
- Novel mechanism of action	HAS	TLV						NICE				
- Significant innovation for a rare disease								NICE				
- New class of drugs	TLV											
- Potential valuable new therapy								NICE				
- Oral administration advantage	TLV, SMC, HAS											
Indirect benefits from taking the treatment												
- Ability to lead an active and fulfulling life								NICE			★	
- Ability to contribute to society								NICE			★	
- Significant impact on morbidity					HAS						★	
- Significant impact on mortality					HAS						★	
- Significant impact on quality of life					HAS						★	
- Resource use reduction		HAS, SMC		HAS	HAS						★	

Legend: end: NICE End-of-life supplementary advice; severity:severe disease; modifiers: SMC modifiers; NICE: National Institute for Health and Care Excellence; SMC: Scottish Medicines Consortium; TLV: Dental and Pharmaceutical Benefits Board; HAS: Haute Autorite de Sante

9.4.5. Stakeholder input

No mention of stakeholder input was found for TLV given that this is done informally and generally not documented, which is reflected in the results. In contrast, formal channels exist to collect stakeholder input during the technology appraisal processes through the Public Involvement Programme (PIP) at NICE, the Patient and Public Involvement Group (PAPIG) at SMC, and the procedures for rapporteurs at HAS (HAS, 2015; NICE, 2004; SMC, 2013a). The CT meeting minutes at HAS note how many outside experts provided input but not the content of their advice.

"Other considerations" were provided by stakeholders in 116 out of 173 "other considerations" identified in the NICE appraisals. 41% of these (n = 116) were provided by clinical experts, 21% by patient experts, and 35% by both clinical experts and patient representatives. Clinical experts provided information about the nature of the disease affecting the patient (27%), issues around current treatment alternatives (13%), the treatment's unmet need (11%) and innovativeness (10%), and the non-significance of adverse events (10%). Patient representatives provided information about the nature of the disease affecting the patient (33%), the non-significance of adverse events (14%), the indirect benefits from taking the treatment (12%) and the patient's unmet need (11%). In Scotland, all drugs received a Patient Interest Group submission except for trabectedin and imatinib. The detail of these submissions was not accessible. Additionally in two cases, clinical input was recorded in the HTA reports, where they commented that treatment pathways depend on symptoms (e.g.eltrombopag) and existing treatments are unlicensed (e.g romiplostim).

9.4.6. Orphan drugs and special status

Table 9-4 represents the subcategories of "other considerations" identified in the sampled drugs (rows) and whether they pertain to certain characteristics specific, but not limited to, rare disease and orphan drugs (columns). Unmet need is more likely to characterise, but is not limited to, rare diseases given the scarcity of relevant knowledge and expertise and the fact that often no effective cure exist. This is due to issues around the diagnosis of some of these rare diseases, the complex and unknown

nature of these conditions, together with the lack of coordination amongst centres of expertise at EU- and international-levels, and the lack of knowledge around best practices (Commission of the European Communities, 2008; EPIRARE, 2013; EUCERD, 2013; Rare Best-Practices, 2013). Further, given that orphan drugs often do not have any effective cure, hence the reason for implementing incentives at marketing authorisation level, treatments for rare diseases are more likely to be innovative. On this basis, the "other considerations" that were put forward as one of the main reasons for the final decision identified previously, therefore influencing the final decision, may favour orphan drugs compared to drugs to treat normal conditions. This was seen, for example, with "unmet need" for lenalidomide, mifamurtide and mannitol dry by NICE, and for eltrombopag, romiplostim by SMC.

Table 9-4. Special status of orphan drugs

	Special status?	Disease or treatment characteristic specific to rare diseases and orphan drugs								
Subcategories of "other considerations" (non-quantifiable or non-quantified)	✔ specific to orphans ★ likely more characteristic of orphans	small patient numbers	scarcity of relevant knowledge and expertise	genetic origin	chronic, progressive, often life-threatening	disabling	no effective cure	75% of rare diseases affect children	30% of rare disease patients die before age of 5	High level of suffering for patient and family
Nature of the disease affecting the patient eg disease severity, impact on quality of life and daily activities	★				✔	✔	✔	✔	✔	✔
Nature of the disease affecting the patients' surrounding eg impact of the disease on the families' quality of lives, anxiety, limiting life choices	★									✔
Rarity, orphan status, small patient numbers	✔	✔								
Unmet need eg no or few treatment alternatives exist, treatment pathway unclear	★		✔				✔			
Type of treatment benefit eg curative, life extending	★						✔	✔	✔	
Innovative nature of the treatment eg new mechanisms of action	★		✔				✔			
Indirect benefit from the treatment eg quality of life improvement, ability to live normal lives, improved symptoms, administration benefit	★				✔	✔		✔	✔	✔
✔ Characteristic specific to rare diseases and orphan drugs										
★ Characteristics likely specific to rare diseases and orphan drugs										

Source: (Nicod & Kanavos, 2016b)

9.5. Discussion

This study identified the value judgments made for a sample of ten orphan drugs in four countries in order to understand how they influenced the assessment process, particularly in those cases with a high ICER or, in the case of France, a high ASMR (I-III), which allows the manufacturer to set the price (consistent with other European markets). The study also identified those cases when these "other considerations" were provided by different stakeholders, by type of information provided, as well as those cases when the "other considerations" pertain more to orphan drug compared to more normal conditions. Implications from these findings are discussed in this section, and focus on five topical areas: (a) the added-value of a classification framework, (b) how results compare with existing literature, (c) the determinants of social values, (d) accountability for reasonableness, and (e) orphan drugs and special status.

Based on what we know about value judgments (Rawlins, 2014; Rawlins et al., 2004), one of the significant contributions of this study is the proposed classification framework of these (Table 1). Its application enabled to identify and differentiate the scientific and social value judgments made (Table 3), where implications are two-fold. First, it constitutes a way forward to highlighting needs for further research (when evidence is incomplete or preferences are non-elicited). Second, if they continue not to be elicited or quantified, retrospectively identifying these to prospectively create a taxonomy of criteria may facilitate their being used more consistently when similar scenarios are encountered in the future. For example, NICE emphasised the impact of osteosarcoma on families' and friends' lives when assessing mifamurtide, or SMC and TLV recognised the "oral administration benefit" when assessing eltrombopag. These are non-quantified or non-elicited criteria for which preference could be given in future cases by their inclusion in the taxonomy of criteria to be accounted for. This is all the more important when considering the extent to which these considerations are different across countries and likely also across decision-making bodies within one HTA agency. These differences are either a consequence of agency-specific value preferences (Nicod, 2016a), or of committee-specific preferences reflecting the

composition of the decision panel and their individual judgments driven by their experiences, and it is therefore important to improve the consistency in their use.

The different "other considerations" identified and their classification into sub-categories and clusters are in line with findings from the literature on (social) value judgments. Schwappach (2002) divides the determinants of social value into those factors relating to patient characteristics and those to the treatment (Schwappach, 2002). Our study clustered these determinants in a similar manner into two clusters (e.g. treatment or disease characteristics), and takes one step further by applying this same classification to both social and scientific value judgments. Second, a number of individual social values were identified in the literature. One study in England used qualitative techniques to define these, where respondents agreed to favour need, preventive care, quality of life, health improvement and life expectancy, in addition to not favouring certain populations according to age or socio-economic status (Baker, Bateman, Donaldson, Jones-Lee, Lancsar, Loomes, Mason, Odejar, Prades, Robinson, Ryan, Shackley, Smith, Sugden, Wildman, & Team, 2010). Generally there is agreement about what these social values are, but the determinants of social value remain broadly defined and no exhaustive list of these exists. When comparing these results to the topics defined within the ethical, organisation and social domains of the EUnetHTA Core Model (EUnetHTA, 2013), commonalities and differences are seen. The topics included in the ethical domain relate to the societal preferences (elicited social values), those in the social domain relate to treatment characteristics, and those in the organisational domain to the financial impact or organisational impacts of using the treatment. The Core Model domains do not necessarily capture those aspects put forward about the patient experience in living with the disease, but rather focuses on those aspects that change when taking a treatment. In contrast, results presented in this chapter do not capture all of the domains highlighted in the Core Model, likely because these had not been put forward in these particular cases analysed. Nevertheless, this study contributes to understanding how those topics included in the Core Model may be expressed in practice.

Given that the determinants of social value are only broadly outlined, this study contributes to better defining these. For example, "need" or "unmet need" is a

determinant of social value. It is accounted for in the weighing of disease severity by TLV, as one of the SMC modifiers, and discussed by a NICE Citizen Council meeting. Nevertheless, no clear definition of unmet need exists. Our results captured the variety of ways of expressing "unmet need" (Table 8-3), which can be used to define it. Another example is disease severity, for which no single definition exists. It is characterised by a number of determinants, which include the impact on quality of life and mobility, or considerations of life expectancy (Dolan & Shaw, 2004; Garau et al., 2009b). Severity is included into HTA either through a weighing of the QALY (or of other measures of HTA) or as part of the deliberative process (Garau et al., 2009b). The latter would apply to our study countries since no specific weighing for severity was seen, including in Sweden where it is explicitly accounted for despite the definition of disease severity being broad (as noted earlier). Our results identified the various forms of expressing severity, which can be used to better define severity for future cases. For TLV, these included: the life-threatening nature of the disease, the negative impact on daily activities including functional capacity and on quality of life, and the short life expectancy from having the disease. In France, where no ICER or threshold exist, informal methods are used to incorporate societal and political values into the assessments. This is explicit in the evaluation of the public health value (intérêt de santé publique) of drugs as part of the coverage evaluation (SMR), however, whether these determinants of (social) value are accounted consistently across cases is another question, which could be partly addressed by applying the taxonomy of criteria.

For a resource allocation decision to be accountable for reasonableness, the process should be transparent and public, based on reasons that are relevant, decisions should be revisable when new evidence is available, and the process should allow for these conditions to be enforced (Daniels, 2000; Daniels et al., 2008). This usually takes place during the deliberative process of HTA, during which the Committee discusses the evidence and accounts for stakeholder opinion until a decision is made. The decision and reasons for the decision should then be documented in the HTA report, most often publicly available, as is the case with our study countries. In terms of stakeholder input, a clear process exists at NICE and SMC where they are given the opportunity to voice their concerns or opinions. Our analysis confirmed that this is

well-reported for NICE (given the high number of "other considerations" provided by different experts), but is not as detailed in SMC's summary of advice, probably because it is a less detailed report. HAS has specific procedures governing outside experts (rapporteurs) who provide advice and input in the evaluation process. For TLV, no official procedures exist, although some of the key stakeholders are represented within the Appraisal Committees (e.g. clinical experts). Generally, the type of input from these stakeholders could be better documented or transparent. Some argue that it is not sufficient to have a formal procedure to account for stakeholder input and value judgments, but that it should also be clear how these have influenced the decision, which is often lacking (Garau et al., 2009b). Our results further confirm this in the number of "other considerations" (from stakeholders or not) identified, where it is not entirely clear how these factors contributed to the decisions particularly in those cases where these were (non-elicited or non-quantified) value judgments. The taxonomy of criteria developed together with the type of input from different stakeholders may help understand the criteria that are relevant to decision-making and their sources that go beyond routine methods of assessing clinical benefit and ICERs.

Little agreement exists on whether patients with rare diseases requiring orphan drug treatments deserve a preferential status (Desser et al., 2010; McCabe et al., 2010; NICE Citizens Council Report, 2004). Nevertheless, governments recognise the difficulties in appraising these treatments and the fact that they should be treated differently. In France, patients with orphan diseases have a preferential status, but their needs go much beyond drugs. Only recently, NICE and SMC have implemented new procedures for end-of-life and ultra-orphan drugs. The treatment's additional benefit and other elements not captured by the ICER (e.g. unmet need, disease severity, added value the patient and surrounding) are now accounted for by SMC, together with patient and clinical engagement. These other elements correspond to the "other considerations" identified in this study. Similar questions are arising in Sweden, where a consultation on how to appraise orphan drugs has recently been issued. Further, in NICE's recent consultation on value-based pricing, they attempted to find novel approaches to capturing burden of illness and other issues. They concluded that approaches to adjusting the QALY were insufficient, and therefore it is

essential to identify the criteria that are important in decision-making and that go beyond the ICER. This study provides an alternative to the issue of preferential status by accounting for the non-elicited or non-quantified "other considerations" that influenced previous decisions, and query whether it would be worth eliciting preferences for these. This could then feed into novel approaches in assessing orphan drugs.

9.6. Conclusions

This study systematically identified the scientific and social value judgments made in four countries for a sample of orphan drugs, and explored how they influenced the deliberative process of HTA. The proposed classification framework of these value judgments was used to identify needs for further research and to improve consistency in their use across drugs within one agency. This was then used to address different issues around identifying and better defining the determinants of social value or how to improve the lack of accountability for reasonableness particularly in cases when it was not clear how the "other considerations" identified influenced the decisions. It also provided a way forward to eliciting whether these orphan drugs deserve a special status by eliciting preferences around some of the social value judgments made which are more likely to pertain to orphan drugs compared to normal condition, rather than focusing on the opportunity cost of these. Given the challenges in producing robust evidence for orphan drugs due to the small patient numbers and heterogeneity of the diseases, scientific and social value judgments are unavoidably part of the decision processes for these drugs. Identifying the scientific and social value judgments through the application of this framework enables us to create a taxonomy of criteria that were relevant in these decision-making processes and which go beyond route methods for HTA.

10.Dealing with Uncertainty and Accounting for Social Value Judgments in Value Assessments for Orphan Drugs: Qualitative Evidence from Four European Countries[8]

10.1. Abstract

We compared the value assessment of orphan drugs in four European countries and explored differences in reimbursement decisions. Semi-structured interviews with HTA body representatives in England, Scotland, Sweden, and France were conducted. An interview topic guide was developed based on findings from a systematic comparison of HTA decisions for ten orphan drugs. Qualitative thematic data analysis was applied to the interview transcripts using the Framework Approach. Eight HTA body representatives were interviewed between March and June 2015. Evidentiary requirements and approaches to dealing with uncertainty were discussed around: trial design, population and duration, comparators, relevant endpoints and economic modelling. HTA bodies agreed that decisions regarding orphan drugs are made in a context of greater uncertainty. The threshold of acceptable uncertainty varied by country and was generally not related to the risk of not marketing the drugs. The acceptability of surrogate endpoints was not consistent across countries nor were the validation requirements. Different mechanisms were used to modulate the ICER in cases of uncertainty (e.g. Patient Access Schemes). Some countries require higher evidentiary standards for greater clinical claims, which may be more challenging for orphan diseases. The most common social value judgments identified related to innovation, disease severity and unmet need. Trivial differences were seen in the way these concepts were defined and accounted for across countries. Although agreement was seen in evidentiary requirements or preferences, there were subtle differences in the circumstances where uncertain evidence may be considered acceptable, possibly explaining differences in HTA recommendations.

---

[8] A version of this chapter is soon to be submitted for publication to *Value in Health* (Nicod E, Berg Brigham K, Durand-Zaleski I, Kanavos P, 2015)

10.2. Introduction

HTA aims to ensure that technologies offered are safe, efficacious and provide value for money (Hurley, 2000). Value may be perceived within the context of efficiency by reimbursing only the most efficient technologies within an allowable budget; however, this does not necessarily account for what truly matters to society and to those being treated (Caro, 2009). Value may also reflect specific attributes, such as innovation, with direct implications for patients in terms of improved prognosis or quality of life, and indirect ones for society through increased productivity, other societal contributions, possibly also benefiting patients in other disease areas through knowledge spill-overs. HTA is best viewed as an umbrella concept in a multi-disciplinary field that aims to capture the value of a new technology through a range of facets relating to different stakeholder perspectives and attributes of interest.

For a given drug, two bodies may reach opposite conclusions despite appraising the same evidence and using similar measurement outcomes (Nicod et al., 2012). These processes may rely on different attributes of value or on different approaches to dealing with often imperfect evidence. The acceptability of uncertain clinical benefit and cost-effectiveness therefore depends on the tools used to address uncertainty and on the judgment of the decision-makers with consideration of additional qualitative criteria (e.g. stakeholder input, disease or treatment characteristics) (Nicod et al., 2015a). The internal regulations of HTA bodies explain the frameworks under which they operate, and the opinions or recommendations regarding specific health technologies generally provide documentation of the evidence considered and the bases for the decision. However, subtleties may not be captured even in the most complete documentation.

This is particularly the case with respect to orphan drugs, because it may be impossible to apply the generally-applicable rules in defining the appropriate evidence for small populations facing very serious chronic or life-limiting diseases. Producing high quality evidence has proven to be challenging because of the small patient numbers, heterogeneous nature of these conditions, and lack of scientific expertise (Vickers, 2013). This has implications for clinical trial design (e.g. appropriate endpoints, trial

duration or clinical pathways) and conduct (e.g. recruitment from scarce patient numbers and specialists) (Vickers, 2013). Two studies compared trials for orphan and nonorphan conditions and found that orphan drugs were generally characterised by lower quality evidence (Bell et al., 2014; Kesselheim et al., 2011). These challenges, in addition to their high prices, result in orphan drugs generally not being cost-effective (Drummond et al., 2007). Orphan drugs often undergo the same HTA processes as for more prevalent conditions. Ongoing debate is whether we (society) are willing to pay more for these rarer conditions. This was not supported by a number of surveys that aimed to elicit this question and found the contrary when resources are taken from more prevalent conditions (Drummond et al., 2014b). A better understanding is therefore needed about how HTA bodies value orphan drugs and deal with issues related to rarity, and further the debate on whether they deserve a special status or their processes should be differentiated.

In a previous study, the decisions of four HTA bodies for 10 orphan drug-indication pairs were analysed based on the opinions and in light of each entity's internal regulations (Nicod, 2016a). While the same clinical trial evidence was considered by each HTA body, their analyses and conclusions were not uniform, where a substantial number of drugs (60%) received diverging recommendations (Table 10-1). On this basis, this study aimed to develop a broader perspective about how value is assessed for orphan drugs in four European countries and how differences affect reimbursement decisions based upon interviews of representatives of the HTA bodies.

Table 10-1. Drugs included in previous study and overview of key characteristics seen in the trial submission

HTA recommendations ICER or ASMR (pricing scheme) and SMR (coverage rate) in France	**NICE England, N 10**	**SMC Scotland, N 10**	**TLV Sweden, N 5**	**HAS France, N 10**
Eltrombopag Chronic idiopathic thrombocytopenic purpura	**Reject** >£100k	**Restrict (high risk of bleeding)** Dominant	**List** Dominant	**II (EU)** important (65%)
Romiplostim Chronic idiopathic thrombocytopenic purpura	**Restrict (high risk of bleeding)** £30k splenectomised, <£20k non-splenectomised	**Restrict (high risk of bleeding)** £15k splenectomised, £17k non-splenectomised	**List** Sek4-600k	**II (EU)** important (65%)
Everolimus Renal cell carcinoma (2nd line, advanced)	**Reject** £52k	**Reject** £61	**List** High	**IV** important (100%)
Lenalidomide Multiple myeloma (2nd, 3rd line)	**Restrict (3rd line)** >WTP 2nd line <£44k 3rd line	**Restrict (3rd line)** £34-41k 3rd line	**List** Sek290k	**III (EU)** important (65%)*
Mifamurtide Osteosarcoma	**List** £36k	**List** £48k	**List** Sek7-900k	**Reject** insufficient (0%)
Azacitidine Myelodysplastic syndrome	**List** £47k	**List** £51	**NA**	**II (EU)** important (65%)*
Imatinib GIST (adjuvant, after surgery)	**Reject** £21-23k	**Restrict (high risk of recurrence after resection)** £21k	**NA**	**III (EU)** important (100%)
Mannitol dry Cystic fibrosis	**Restrict (no rhDNase)** £50-80k rhDNase <£30k no rhDNase	**Reject** £20k no rhDNase	**NA**	**V** weak (15%)
Ofatumumab Chronic lymphocytic leukemia	**Reject** £50-80k	**Reject** £108k	**NA**	**V** moderate (30%)*
Trabectedin Soft tissue sarcoma	**List** £34k	**Reject** £36k	**NA**	**V** important (65%)*
Evidence appraised				
Primary trials (#)	**13**	**13**	**6**	**13**
Design				
Phase III (% trials)	85%	85%	100%	85%

Phase II (% trials)	15%	15%	0%	15%
Study population				
Less than 300 patients (% trials)	62%	62%	50%	62%
Subgroup data (% drugs)	40%	40%	20%	40%
Comparators				
Alternative treatment (% trials)	0%	0%	17%	0%
Standard care (% trials)	92%	92%	83%	92%
None (% trials)	8%	8%	0%	8%
Relevant endpoint				
Clinical endpoint (% drugs)	20%	10%	0%	10%
Surrogate endpoint (% drugs)	80%	90%	100%	90%
Health-related quality of life				
Included (% drugs)	50%	50%	40%	20%
Not explicitly reported (% drugs)	50%	50%	60%	80%
Economic models				
Cost-utility models (% drugs)	100%	100%	80%	NA
Cost-minimisation models (% drugs)	0	0	40%	NA

Source: The author, based on (Nicod, 2016a).

Legend: N: # of drugs; * Coverage rate not specified in report; EU: European price levels and price negotiation.

NOTES: Two economic models recorded in TLV's report for lenalidomide.

10.3. Methods

10.3.1. Study sampling and data collection

Purposeful sampling was used to select the study countries, each of which undertake assessments using well-established processes and criteria, have publicly available recommendation reports and represent a cross-selection in terms of HTA approach and perspective. The selected study countries were England (National Institute for Health and Care Excellence, NICE), Scotland (Scottish Medicines Consortium, SMC), Sweden (Dental and Pharmaceutical Benefits Board, TLV) and France (Haute Autorité de Santé, HAS). Each HTA body also accounts for elicited societal values, which allows greater flexibility in the face of high and uncertain ICERs or outcomes.

We conducted semi-interviews with HTA body representatives identified by partners of a European research consortium (Advance-HTA) in each of the study countries. An interview topic guide was developed by the lead author and reviewed by all co-authors (Appendix C). It included open-ended questions derived from actual questions that arose in the context of our cross-national comparison of 10 orphan drug-indication pairs, including how certain identified criteria were considered and influenced the opinions. The interview questions were divided into four general themes: (a) General evidentiary requirements for orphan drugs, to better understand what scientific evidence is required and to what extent it is different for orphan drugs compared to other drugs; (b) Dealing with uncertainty, to understand whether more flexibility is given to accepting certain types of uncertainty when other types of evidence are presented or because of the rarity of the treatments under review; (c) Social value judgments, to understand the type of value judgments made by the assessors and how these are accounted for during the processes; (d) Stakeholder involvement, to understand who else other than the manufacturer and the HTA body members have input in the assessments. The structure and responses to the questions in the interview topic guide are shown in Table 10-2. The interview guide was developed in such a way as to ensure consistency of focus while providing flexibility in the discussions, so

that interviewees were free to offer additional insights and interviewers could ask spontaneous questions.

An invitation to participate in a face-to-face or telephone interview was sent to each of the identified interviewees by email along with the interview topic guide. Interviewees were ensured anonymity, and their responses remained confidential pending their confirmation and approval of the content. The study protocol underwent the LSE Research Ethics procedure and received exemption. Interviews were recorded and transcribed by the lead author and sent to the interviewees for comment and validation. Following the interviews, the topics or issues that emerged as relevant or different across countries were compiled and analysed by the authors together with a summary of the views of interviewer(s), circulated amongst co-authors and accounted for during the analysis.

10.3.2. Data Analysis

Qualitative thematic data analysis was undertaken using the Framework Approach (Gale, Heath, Cameron, Rashid, & Redwood, 2013). After familiarisation with the topics discussed, a number of subthemes within each general theme were identified and inductively coded. These included, for example, “preferred type of trial” or “comparator”, “appropriate endpoint”, etc. An interview matrix was created in Excel to facilitate comparison of each subtheme across the four HTA bodies. The key findings from each of these subthemes were summarised in tables and incorporated illustrative quotes from the interviewees. The findings were discussed by all co-authors, and a list of follow-up questions for the four HTA bodies was developed to complement the interviews where information was unclear or incomplete. These questions were sent to each of the interviewees along with the summary findings for their particular HTA body for response and confirmation.

Results focus on the contrasts across countries identified within each theme. Themes were reorganised as follows: (a) clinical evidence and uncertainty, (b) comparators, (c) treatment outcomes and safety, and (d) additional qualitative criteria. Results about stakeholder input were excluded, as no additional information was provided compared

to the previous chapters. Each theme protrays the agencies' perspectives about the clinical evidence appraised and whether this evidence for orphan drugs is characterised by even greater uncertainty compared to more prevalent conditions. This evidence-based used for HTA is imperfect or incomplete, therefore uncertain, as it relies on estimated values from experimental or observational studies (Rawlins, 2014; Rawlins 2014, Claxton, 2008). Decision-makers make scientific value judgments about the extent to which this uncertain evidence is acceptable. This includes judgments about whether the evidence presented fully and accurately captures the effect of the intervention, whether it is generalizable to the local context of the decision, whether quality of life changes are accurately captured, or whether it is appropriate to impose restrictions to population subgroups (Rawlins 2014). I aimed to shed light on the different perspectives adopted when making these judgments about uncertainty regarding the themes discussed, where cross-country differences had already been identified.

10.4. Results

Eight representatives from each of the four HTA bodies agreed to participate in the interviews between March and June 2015. Interviewees occupied senior positions in their agencies (e.g. Head of the Technology Appraisal Programme, Head Economist or Pharmacist, Chair of the Appraisal Committee). Interviews were conducted face-to-face and, in one case, by telephone, lasting between one to three and a half hours. This section summarises the most relevant and contrasting findings (Table 10-2).

10.4.1. Clinical evidence and uncertainty

Trial design

No formal requirements around minimum levels of evidence are imposed, though phase III comparative trials are often preferred. This was illustrated in the primary trials considered, which were predominantly phase III trials (with relatively small patient numbers), with the exception of two phase II trials that received early marketing authorisation (e.g. trabectedin and ofatumumab) (Nicod, 2016a). HAS also

requires all existing and available data at the time of the HTA submission. An important distinction was seen in their expectations about the quality of the evidence submitted when examined within the context of the clinical claim. TLV has higher scientific and methodological demands for superior efficacy with a price premium, and greater uncertainty is accepted for non-inferior efficacy (and low price) or for treating untreatable diseases as long as the treatment is safe. Similarly, the highest ASMR rating should demonstrate in a good way the treatment's benefit on survival. HAS also judges whether the evidence presented is of sufficient quality by accounting for the situation of the disease in terms of prevalence and number of recruitable patients.

The type of information generated from non-primary non-phase III trials was about safety, dosage research and historical controls (Nicod, 2016a). A contrast was seen in the acceptance of historical controls for trabectedin only by NICE due the non-comparative nature of the phase II trial presented. Although rarely used, the agencies agreed that historical controls are acceptable when no other data is available (NICE), when it is the best data available (SMC, TLV), to collect data on disease progression when no alternative treatments exist (HAS), or when the disease is rare or other special circumstances are seen (SMC). Historical control data is considered generally of poor quality, which would explain its limited use and in our example and the fact that it was only considered by NICE for trabectedin. All agencies recognised that no established treatment alternatives exist and were presented historical control data. It was not accepted by SMC and HAS because of the statistical methods used in analysing the data, whereas it was accepted by NICE because of the investigational nature of the treatment. The interviewees recognised the usefulness of registry data to obtain historical controls, and information about disease progression for economic modelling purposes (NICE), or to obtain longer term data (particularly for rare conditions, which are often chronic and rely on short-term data) about efficacy and safety (HAS). Their limited use is explained by the unknown nature of the type of data useful in the future (NICE) or the missing data where comparisons become inappropriate (HAS), as illustrated in our example.

Trial length

In our previous study, the HTA bodies were often concerned about the trial duration being too short (eltrombopag, imatinib, mannitol dry, romiplostim) or the uncertain optimal treatment duration (imatinib). When questioning the agencies about how the appropriate trial length is assessed, different approaches were described accounting for the summary of product characteristics (SMC), the durability of response (TLV, SMC), the assessment by the EMA (TLV), or the natural prognosis of the disease (TLV, HAS) (Table 10-2). Given the often chronic nature of these rarer conditions, issues around the short trial length may be more common, as recognised by SMC and TLV who are willing to accept greater uncertainty for orphan drugs, but not for NICE who does not make any differentiation for orphan drugs. This was illustrated for mannitol dry to treat cystic fibrosis, a chronic condition, where the length of the two trials assessed were considered too short by HAS and SMC given that the treatment is to be taken over a prolonged period of time, but not by NICE nor TLV, possibly because they were the pivotal trials considered for marketing authorisation. In the previous section, HAS recognised the usefulness of registry data particularly for the rarer and often chronic conditions. This may explain the negative decision for mannitol dry, which was not made available under the temporary authorisation scheme (ATU) and for which no longer-term data was being collected.

Study population

Four out of the ten drugs analysed relied on subgroup data (Nicod, 2016a) (Table 10-1). When asking the interviewees about how they deal with subgroup data, different perspectives were given. Pre-specified subgroup data is generally preferred, but acceptable under certain circumstances (Table 10-2). This was illustrated for mannitol dry, where HAS was concerned that the trial population (including children and adults) was not representative of the indication for marketing authorisation (adults). This was due to the data presented in the HTA assessments, which included both adults and children and were presented either together (SMC, HAS) or only for adults (NICE). Additionally, extrapolating treatment effects from subgroup data to a wider population would also not be accepted by HAS. For the others, preference would be given to the subgroup driving the results (most cost-effective subgroup). This was seen for eltrombopag restricted by NICE and SMC to patients with a severe

risk of bleeding, given that the need for rescue medication was the main driver of costs and cost-effectiveness, and for which the ICER across the whole population was not cost-effective. This was also highlighted by TLV, who rather than impose a restriction, requested a re-assessment in the future to assess its cost-effectiveness in practice.

Comparators

The comparators included in the primary trials for the ten orphan drugs analysed were placebo or standard care, except for three cases without comparators due to early marketing authorisation (ofatumumab) or comparing different doses of the treatment under investigation (trabectedin, mannitol dry) (Nicod, 2016a). The scoping process to identify the appropriate comparator differs across countries, where it takes place before the appraisal process at NICE and during the appraisal process for the others. Subtle differences were seen across countries in the selection criteria of the appropriate comparator (Table 10-2), where the judgment about its appropriateness usually relies on clinical expertise and local clinical guidance. For NICE and TLV, one of the most frequent issues encountered relate to the comparator, whereas this is more seldom seen for SMC and HAS. This contrast was not reflected in the systematic comparison, where the lack of comparative data was an issue in five cases with varying levels of concern and approaches in their dealings. For example, the lack of comparative data with another treatment for romiplostim, which relied on standard care, was either acceptable due to the rare and heterogeneous nature of the disease (NICE) and the numerous (unlicensed) treatments used in practice (SMC), or not, due to missing transposability into clinical practice (HAS). A similar scenario was encountered for trabectedin. These illustrate common scenarios encountered for rare diseases (e.g. no comparator, unlicensed options, unknown clinical pathway, no comparative data, unknown optimal dosage), where the validation of the appropriate comparator may be more challenging due to fewer experts with sufficient knowledge in these disease areas, as highlighted by SMC.

Treatment outcomes

The relevant clinical endpoint used to assess a treatment's clinical benefit varies depending on the agencies, circumstances, and economic models used (Table 10-2). NICE only accepts the QALY, whereas it is required by SMC and TLV when cost-utility models are appropriate. If non-inferiority between two treatments is demonstrated, then cost-minimisation models are preferred (SMC, TLV). For HAS, the choice depends on the situation of the disease (e.g. short term consequences) and aim of the drug (e.g. symptoms for a symptomatic treatment).

Orphan drugs often rely on surrogate endpoints (Joppi, Bertele', & Garattini, 2013), as illustrated in 8 of the 10 study drugs analysed that were predominantly validated except for two cases (Nicod, 2016a). Surrogate endpoints are defined as "biomarkers intended to substitute a clinical endpoint"(Biomarkers Definitions Working Group, 2001), which is the definitive or clinically meaningful endpoint to the patient, such as overall survival. The acceptability of surrogate endpoints by the interviewees depended on their validation, against a hard or soft endpoint, with the exception of HAS for the latter. A non-validated endpoint would probably not be accounted for by NICE, whereas it may be acceptable under certain circumstances for the others (Table 10-2). Greater acceptance of surrogate endpoints for orphan drugs would not be accepted NICE or TLV, and accepted if no other option is available (HAS) or if the drug fulfils the SMC modifiers (e.g. life-extending treatment, life-threatening condition), where greater uncertainty including around surrogate endpoints is accepted.

Progression-free survival is one of the most commonly encountered surrogate endpoint. Different levels of acceptability were seen: NICE always prefers overall survival to progression-free survival, even if it is the trial's secondary endpoint (e.g. mifamurtide and imatinib). Progression-free survival is accepted by SMC when there is an established link with life extension or HRQol, or when the main benefits of the treatment are improved HRQol rather than life extension. TLV also prefers overall survival, but understands it is often not available and relies on progression-free survival considered potentially closer to patients' needs: at the point of conversion between two survival curves, the area between the curves are likely to reflect a benefit

to patients. In HAS's approach focusing on short-term risks, progression-free survival would not replace overall survival in a situation where the patient would die shortly unless it were a validated surrogate of overall survival.

The main contrast seen for HRQol data was in its considerations as a hard (NICE, TLV, SMC) or soft endpoint (HAS). Despite the weight of HRQol not being explicit, HAS highlighted issues when this data was not provided or no improvement was demonstrated (e.g. eltrombopag, everolimus, mannitol dry, trabectedin). Subtle differences were seen in the preferred types and sources of HRQol data (Table 10-2). In our study, HRQol data was not commonly reported (50% of cases), nor was it reported homogeneously when reported. Issues relating to HRQol were raised for nine out of the ten study drugs and related to the lack of HRQol data (all), the lack of HRQol data collected from the pivotal trial (NICE), the lack of improvement in HRQol (HAS), or the limitations in the HRQol data presented (SMC, NICE) (Nicod, 2016a). These issues influenced the decision either negatively or were dealt with in different manners, via other types of HRQol data or through clinical expertise. Challenges in collecting HRQol data may be seen, particularly for the rarer conditions (TLV), but are not considered necessarily specific to orphan drugs (SMC).

Safety

Safety is not, per se, part of the assessment for NICE, SMC and TLV given it has already been assessed for marketing authorisation. It is considered if it impacts on QALY gains or whether it is adequately captured by utility, survival and cost estimates (SMC, NICE). It is somewhat different for HAS, which assesses safety in the same way as efficacy. There were also situations where agencies agreed that safety can modulate the assessment of efficacy (e.g. if efficacy is the same and safety is worse), including in the context of the consequence of not giving a treatment (TLV). In practice, this would translate into ensuring that comparative safety data is available between the treatments being compared or whether one alternative shows greater events than the other, and whether this produces extra costs. As seen in our examples, the safety issues highlighted related to specific risks (e.g. hearing loss) in six cases

(NICE, HAS) or to the uncertain nature of the evidence (e.g. lack of long-term or comparative safety data) in five cases (NICE, SMC and HAS) (Nicod, 2016a).

Table 10-2. Summary of interview findings about clinical evidence and uncertainty, including illustrative quotations

			NICE	SMC	TLV	HAS	Illustrative quotations/comments
Trial design	Requirements	None	☑	☑	☑		No formal requirements
		All available evidence				☑	HAS: "HAS requires all the clinical trial data available at the time of the assessment"
	Preferences	Best or highest available evidence	☑			☑	HAS: "HAS has a preference for demonstrative data, which means data that are the highest level of evidence (e.g. phase III comparative well-designed and conducted trial)"
		Phase III comparative trials	☑	☑			NICE: "The Committee feels more comfortable about making decisions on clinical effectiveness based on phase III trials, but it is very rare to actually have phase III trials with the correct comparator"
		Requirements similar for all drugs	☑	☑			SMC: "each case is viewed upon its own merits"
		Higher methodological requirements for superior/higher efficacy claim			☑	☑	TLV: "If their price is really low (and clinical claim is non-inferiority), than any uncertainty is ok as long as patients don't die (which has already been checked by the EMA)" HAS: a higher claim should demonstrate in a good way the effect of the treatment - "ASMR I is granted for drugs that have a demonstrated in a good way a substantial effect on survival"… The ASMR IV is for a demonstration that is not so perfect and with a quantity of effect which exists but is not very important"
		Lower methodological requirements when the consequence of the decision is severe			☑		TLV: "Greater uncertainty accepted if the consequence of the decision is severe"
		Quality of the evidence is assessed according to the situation of the disease (prevalence and number of recruitable patients)				☑	HAS: "accounts for the real situation of the disease, considering the prevalence and number of patients that are recruitable in trials, as often seen for orphan drugs"
	co ntr ols, acc ept abi	When no other data available	☑				NICE: "Historical controls are rarely seen mainly in cases when no other data

							is available"
		When it is the best evidence available		☑	☑		
		When no other treatments are available and to obtain data on disease progression				☑	HAS: "HAS is very much in favour for prospective appropriate data collection on natural history of a disease that can serve as a comparison when another comparison is not possible"
		When the disease is rare or other special circumstances are seen		☑			SMC: "the acceptability of registry data by the Committee would depend on many factors already discussed (e.g. rarity, etc.)…"
	Registry data, use	Historical controls	☑	☑	☑	☑	
		Natural progression of the disease (e.g. to obtain long-term data)	☑			☑	NICE: "lifelong modelling of the disease and therefore need long term data about disease progression, which will never come from any trial"
		Treatment efficacy and safety				☑	HAS: "at the first assessment for reimbursement, in general only short term data is available and orphan disease are in majority chronic diseases so they also rely on registries to have longer term data on efficacy first, and safety second"
Trial length	Acceptability criteria	Natural progression of the disease	☑			☑	NICE: "the Committee always welcomes data on natural history of the disease to validate any extrapolation curves." HAS: " If the duration is too short compared to the natural course of the disease then it will be criticised"
		Likely durability of treatment response	☑	☑	☑	☑	NICE: "The Committee always welcomes data on the parameter, or seeks sensitivity analyses with different assumptions if no data is available."SMC: "… the likely durability of treatment response (may be informed by other sources)"HAS: "…sufficiently long to generate solid evidence about the type of benefit to the patient"TLV: "The trial length needs to cover the time spam up to the point where we can see that both treatments converge"
		Corresponds to the EMA assessment			☑		TLV: "This is done by the EMA and TLV trusts that the right recommendations were given."

		SPC advice for treatment duration	☑	☑			NICE: "NICE would be bound by the treatment duration specified in the SPC, unless a stopping rule is proposed by the company and supported by the clinical community." SMC: " It will relate to consideration of factors such as the duration of the trial relative to what the SPC advises in terms of treatment duration"
		Greater flexibility accepted for orphan drugs	☒	☑	☑		NICE: "we do not differentiate orphan drugs" SMC: "Trial duration isn't something that is specifically teased out as an issue but may be something that is noted as a general weakness of the evidence base (particularly if very short in relation to a very long term economic model). To the extent that we offer greater flexibility in dealing with the general limitations with orphan drugs, issues with limitations in trial duration would be afforded similar flexibility." TLV: "greater uncertainty regarding the clinical effect is accepted"
Study population	Subgroup data, acceptability criteria	Pre-specified	☑	☑	☑	☑	HAS: "The Transparency Committee will not be confident in the results if the subgroup was posthoc, and have a clear preference for pre-planned or pre-specified subgroups"…"has to be pre-specified in the protocol of the trial"
		Posthoc	☑	☑	☒		NICE: "sometimes the population in the licence is from a posthoc group, in which case NICE needs to consider it. Otherwise, these are very rarely included, only if there is a strong biological plausibility of a strong cost-effectiveness argument for including it." TLV: "Subgroups must have been pre specified before using them, it is an absolute rule" SMC: "consideration would be given to whether the subgroup was pre-specified or post-hoc and also the relative size of the subgroup and the potential significance of any results."
		Relative size of the subgroup	☒	☑			NICE: "ideally, small subgroups are not considered, but that depends on what population the licence covers."
		If it is the only available evidence for a very severe rare condition, non-specified data could be acceptable			☑		

		Significance of results (credibility, relevance, and practicalities)	☒	☑			NICE: "often not possible with subgroups" SMC: "From a clinical point of view, important considerations are whether the subgroup has clinical credibility, relevance, and practicality (where it can be easily identified as a group of patients in Scottish practice)"
		If the subgroup is driving the clinical trial results, the indication should be restricted to this group	☑	☑	☑		NICE: ".... the drug needs to be cost-effective in the subgroup and not in the overall group. Only then is recommending the drug for a subgroup only appropriate."SMC: "This may particularly be the case for a medicine that looks to have poorer cost-effectiveness for the whole group as we may try to find ways that can maximise the chance of the medicine being accepted at least for some patients. "TLV: "if the whole study is driven by a subgroup, then very important to treat this subgroup and to exclude the study as a whole because of evidence demands could be very counter-productive."
		Limited to the marketing authorisation and trial indication				☑	
Comparator	Scoping process	By HTA agencies *before* the HTA process (literature review, expert opinion)	☑				
		Based on MAH's submission *during* the HTA process (clinical experts queried about choice of comparator)		☑	☑	☑	SMC: "Within the critical appraisal process, SMC will go to a bank of clinical experts with a set of generic questions about the medicine, which tend to elicit responses about comparators, treatments used in current practice, what would be displaced with the new treatment, etc." TLV: "Experts are the most important source of information. As well as guidelines from the Swedish Medical Products Agency about the treatment recommendations for different conditions. The choice of comparator needs to be very specific to Swedish circumstances. Therefore literature reviews does not play."
	Selection	Existing treatment/standard of care to be replaced/routine practice	☑	☑	☑		NICE Guide to Methods of Technology Appraisal TLV: "Criteria for relevant comparator: most cost-effective, treatment most likely to be replaced (e.g. if the patient doesn't get this new drug, what it he/she getting instead)" SMC: SMC Guidance to Manufacturers for Completion of New Product Assessment Form

		Therapeutic technology used at the same stage of the therapeutic strategy				☑	HAS: "therapeutic technologies that you can use at the same stage of the therapeutic strategy"
		Most cost-effective			☑		
Treatment outcomes	Relevant clinical endpoint, criteria	Overall survival, utility, QALY	☑		☑		NICE only accepts QALYs, and TLV has a clear preference for QALYs (except when the clinical claim is non-inferior efficacy) TLV: "Take home message: TLV is big on QALYs"
		The endpoint in the MAH submission is critically appraised (not identified by Committee)		☑			SMC: "SMC does not identify the endpoint. The company presents the endpoint and SMC judges whether it is appropriate or not. This would take place as part of the deliberative process"…
		Unmet need based on expert opinion		☑			SMC: "It depends on a range of things such as what was presented in the dossier by the manufacturer, need and unmet need fed from experts."
		Endpoint used for the economic model (e.g. survival + quality of life for cost-utility analysis)		☑			SMC: "A long-term model would require important information on overall / long term survival, which is not always possible other than with extrapolation from short term trials."
		Should reflect the aim of the treatment				☑	HAS: "if the treatment is symptomatic, they will consider the symptoms…"
		Should reflect the short term consequence of the disease				☑	HAS: "If the disease is leading patients to die shortly, survival should be chosen."
	Surrogate endpoints, acceptability criteria	Validated for life expectancy (=hard endpoint)	☑	☑	☑	☑	NICE: "If they are not validated against the outcome of interest (qol or life expectancy), they are probably not going to be taken into account" HAS: "If surrogate is validated as predictive for the change of a more hard endpoint, then it will be accepted"
		Validated for HRQol (=soft or subjective endpoint)	☑	☑	☑		SMC: "acceptability is greater where the committee can see that there is an established link between the surrogate outcome measure and the final outcome of interest"
		Clinically relevant			☑		TLV: "Surrogate endpoints must be clinically relevant. How do they relate to qol and life expectancy"

		Certainty of the validation	☑	☑			NICE: "they will look at the certainty or uncertainty of that validation"
		Non-validated	☒		☑	☑	NICE: " They have to be validated" TLV: "If not validated, a surrogate may have to be accepted if it is an important new treatment (and depending on the consequences)" HAS: "If it is not validated, they would not accept surrogates"
		Surrogates for orphan drugs more acceptable	☒	☑	☒	☑	NICE: "we don't differentiate orphan drugs" SMC: "the committee does have more latitude to accept greater uncertainty (through the modifier) and this can lead to a greater acceptance of a surrogate outcome" TLV: "Surrogates are not necessarily more accepted for orphan drugs" HAS: "if there is no other possibility, intermediate endpoints are accepted."
		Situation of the disease		☑	☒	☑	HAS: "HAS adapt their assessment to the situation. If a disease that has 25-30 patients and the trial has included the same amount of patients in a world-wide situation, they will accept a non-comparative study, with a surrogate endpoint, etc. They will consider whether they have tried to reach the highest level of evidence they could."
	Overall and progression-free survival, requirements and preferences	Preference for OS	☑	☑	☑	☑	TLV: "TLV has a preference for OS, but it is hardly the case that that information is available"
		Even if OS is a secondary trial endpoint	☑				NICE: "It doesn't matter if it is a primary or secondary endpoint (like utility), NICE will always prefer OS"
		When QALYs depend on life extension		☑			SMC: "OS is preferred where QALY gained depends on life extension"
		When patients may die shortly				☑	HAS: "PFS cannot replace survival in a situation where the patient would die shortly"
	Progression-free survival, acceptability criteria	PFS validated for OS		☑		☑	HAS: "There is some literature showing that in some kinds of cancer, PFS has shown to be a surrogate of OS and in those cases they would be accepted"
		PFS validated for HRQol		☑			
		PFS better predictor of (validated) HRQol than OS		☑			SMC: "For some analyses, PFS is a reasonable outcome to use because it is likely that the main benefits of treatment will be in terms of quality of life rather than in any degree of life extension"

		PFS may be a better predictor of patients' needs (If OS same for two alternatives, area between the curbs may be a value to patients)			☑		TLV: " there might be cases when PFS is at least as interesting and as relevant to patients and clinicians as OS"
	Health-related quality of life, requirements and preferences	Required in submission	☑				NICE: utility measures are needed for the cost-utility model (NICE requirement)
		Preferred if claim is superior efficacy (with a price premium), or if a cost-utility model used (hard endpoint)		☑	☑		TLV: "We need to have some knowledge of qol. Or need to make an assumption. Rare are the cases when it not accounted for (e.g. CEA), apart from CMA"
		Preference for generic utility measures (e.g. EQ-5D)	☑	☑			SMC: "SMC has a preference (rather than a requirement) for utility estimates from a validated generic utility instrument such as the EQ 5D"
		Collected within the clinical trial	☑	☑	☑		SMC: "Where utility assessment has taken part within the key clinical studies, we would have a preference for the company using this data in their economic analysis, unless there was a good reason to expect that the data were not appropriate"
		Secondary to assessment (soft endpoint)				☑	HAS: HAS will first look at results on the hard endpoint, and second will look at HRQol to see how the life is for the patient"
	Non-generic utility measures, acceptability criteria	Validated mapping techniques	☑	☑	☑	☑	SMC: "SMC can accept other sources of utility values, for example, via use of validated mapping techniques or use of values from the literature or registries" TLV: "It is very important to have validated mapping techniques"
		Values from the literature or registries	☑	☑		☑	
		Values from other diseases areas	☒*	☑	☑	☑	*NICE: "only under exceptional circumstances"

		Values from expert opinion	☒*	☑*	☑	☑	*NICE: "only under exceptional circumstances" SMC: " The use of expert opinion as a source of utility values would likely be perceived as the most uncertain source of utility values, but has been used in some submissions for some health state valuations" TLV: "Expert opinion can be done for the QALY though the delfi panel (but not when clinical claim is superior efficacy)...It can be used to estimate QALY gains in terms of simplified administration, or parameters that are softer."
	Safety assessment, requirements and preferences	Not assessed per se given it has already been done for marketing authorisation	☑	☑	☑		
		Accounted for if it impact on QALY/utility gains	☑	☑			
		Considered if adequately not captured in the utility values	☑	☑			
		Assessed in the same way as efficacy data				☑	
		Life-threatening diseases, more likely to accept uncertain efficacy if the risk of adverse events is low	☑		☑	☑	NICE: "probably" TLV: "If the risk of severe adverse events from the treatment is low, and that patients can only get better even from taking the treatment, even if we don't know for sure, TLV would allow them to take the chance by paying for this drug. This is also considering that EMA had already looked into safety."

10.4.2. Additional qualitative criteria considered

In addition to appraising the evidence, additional qualitative criteria were considered that relate to treatment innovativeness, disease severity and unmet need (Nicod et al., 2015b). This section aimed to further the understanding of these determinants, summarised in Table 10-3.

Treatment innovativeness

NICE explicitly accounts for the innovative nature of the treatment and defines innovation if it renders a step change for patients rather than belongs to a new class of drugs or contains a new mechanism of action. TLV and SMC do not have specific criteria for innovation as it is considered to be captured by gains to patients. In France, the innovative nature would be captured by a higher ASMR rating (ASMR I, II, III) and a price set within European levels without price negotiations (CEPS). If the innovativeness is recognised in the opinion issued and the drug is for hospital use, it would be covered by a special list on top of the hospital DRGs tariff. Additionally and before the assessment by HAS, if a drug is considered innovative based on three pre-defined criteria (e.g. new mode of action, good efficacy and correct tolerance, covers an unmet need), the MAH can submit a pre-file before the marketing authorisation in order to undergo the fast track procedure and submit an application at the same time as the EMA application.

Disease severity

A higher ICER (up to £30,000/QALY) is accepted by NICE for the more severe conditions, decided during the deliberative process. Severity is defined by how a person's quality of life is affected without the treatment, rather than how the treatment improved survival, which is considered to be captured in the model together with its baseline severity. Generally, most of the cases appraised are terrible or severe. Severity is also explicitly accounted for by TLV, where higher severity is considered to correspond to a greater unmet medical need and higher ICER levels are accepted. No explicit weighing or definition of severity exists (work in progress). In contrast,

severity is not explicitly accounted for, though it may be intrinsically, by the Committee members at SMC. For HAS, severity is captured within the SMR ratings and comprises five pre-defined categories: severe, not so severe, affecting quality of life, life-threatening, etc. These categories are not given explicit weights and their influence on the assessments has shown to be minimal according to unpublished evidence (e.g. severe disease in 50% of sufficient SMR, and severe disease in 50% of insufficient SMR).

Unmet need

For SMC, unmet need is assessed for each case by drawing on clinical expertise to understand current treatment options and how the new treatment option might fit in clinical practice. 'No treatment' would likely have some priority over a situation where 'few treatments' were available, and unmet need would also be recognised in cases where few treatments with intolerable side effects are available. Unmet need would be accounted for as part of the deliberative process through the application of a SMC decision modifier ("lack of available treatments of proven efficacy") and are strictly applied when there is no treatment available of proven efficacy in that particular indication. For TLV, unmet need is captured in the severity measure by focusing on the consequence of the decision without treatment. For NICE, unmet need would be also captured in a similar manner as severity and the consequence of the decision. It is considered in the context of the drug's place in the therapeutic strategy and the medical/healthcare needs. A real unmet need would be recognised when no treatment options are available. For HAS, unmet need is considered within the context of assessment the place of the treatment in the therapeutic strategy, as part of the analysis on identifying the comparators. This includes a description of how the disease is treated, where the new drug would fit, and whether other options at the same stage of treatment are available. If no other options were available, it would be considered a great unmet need.

Consistency across decisions

For NICE, consistency in accounting these other considerations is not a concern if the ICER is below £20,000, which is the case in many therapeutic areas with relatively cheap treatments that produce enormous benefits (e.g. cardiovascular). For cancer however, it is more difficult and every case is different, where different Committees may apply the end-of-life criteria more or less generously. SMC agreed that precedent can be an important factor for consideration and that manufacturers sometimes argue that similar circumstances apply to their medicine in question. TLV also agree that accounting for severity consistently across cases is very complex and should be more distinct. In France, evidence around the use of severity also suggests inconsistencies across cases.

10.4.3. Economic analysis and pricing considerations

Our sample of orphan drugs were characterised with relatively high ICERs, most being above £20,000/QALY in England and Scotland, and similarly in Sweden. This was mainly due not only to the high prices (considering all models were cost-utility except for the one appraised by TLV for eltrombopag), but also the uncertain evidence presented (Nicod, 2016a). Each country has its own mechanisms allowing to modulate these ICERs to a more acceptable level, such as: Patient Access Schemes, lower discounting rates when treatment effects prolonged over a long period of time, additional considerations, stakeholder input (confirming the plausibility of an uncertain assumption). These countries also elicited certain circumstances where greater uncertainty or higher ICERs are accepted. These include the SMC disease modifiers for orphan drugs, disease severity in Sweden, or end-of-life considerations at NICE. Their application has helped to improve ICERs but nevertheless, important differences remain in the HTA recommendations issued across countries, explained by some of the trivial differences discussed in this paper.

The situation is somewhat different in France, where the ASMR rating drives the pricing and the SMR the coverage decision. Prices of drugs with an ASMR I-III are set according to European price levels and negotiated with the Pricing Committee (CEPS), whereas those with a lower ASMR should be lower than their comparators. The SMR also provides information about the coverage rate (15%, 35%, 65%), where

the difference (depending on the rate) is often covered by private health insurances. Additional consideration should be made as to whether the drug is under temporary authorisation (ATU), as seen for all of the study drugs with the exception of mifamurtide and mannitol dry, or if it is a hospital drug, in such case it may also be covered on top of hospital DRGs. The lack of comparative data and uncertain nature of the evidence presented drove the low ASMR ratings (V) in three cases (mannitol dry, ofatumumab, trabectedin), and a rejection for mifamurtide. This contrast compared to the other agencies is very important to highlight, as the lack of comparative data was acceptable in certain cases, also thanks to the different mechanisms in place allowing to modulate and interpret (e.g. sensitivity analysis) the ICER. The ATU may also be an important consideration as the two drugs (mannitol dry and mifamurtide) that did not have an ATU received the lowest SMR ratings, which has a significant impact on the level of coverage.

Table 10-3. Information provided about innovation, unmet need and severity

	Innovation	Unmet need	Severity
NICE	Elicited = defined by whether the treatment benefits patients, determined during the deliberative process E.g. delaying chemotherapy, first oral treatment replacing intravenous administration. Counter-examples: new class of drugs, new mechanism of action (without visible benefits to patients)	Non-elicited = defined by the consequence of the decision, determined during the deliberative process where NICE is willing to accept a higher ICER (up to £30K/QALY) for the conditions with a high unmet need E.g. effect on quality of life of patients without treatment	Non-elicited = defined by the consequence of the decision, determined as part of the deliberative process where NICE is willing to accept a higher ICER (up to £30K/QALY) for the more severe diseases E.g. effect on quality of life of patients without treatment
SMC	Non-elicited = intrinsic to the decision, likely captured differently. Anything providing benefits to patient, captured by the ICER or accounted for during the deliberative process E.g. a first in class could fulfil an unmet need, new mode of action or administration benefits, advantages in terms of service delivery, reduced severe adverse events, step-change in patient management	Elicited (for orphan drugs through the modifiers) = "lack of available treatments of proven efficacy", determined as part of the deliberative process and from clinical experts E.g. "No treatment" would be prioritised over "few treatments". If there were "few treatments" with intolerable side effects, it would be considered an unmet need	Non-elicited = no definition, may be accounted for intrinsically during the deliberative process
TLV	Non-elicited = benefits to patients, captured by the ICER or as part of the deliberative process E.g. improved administration form benefits patients and reduced costs	Elicited = defined by the consequence of the decision, determined as part of the deliberative process Disease severity and unmet need are considered to be related: the greater the severity, the greater the unmet need	

HAS	Elicited (captured within the ASMR) = a drug with an ASMR I, II or III would be considered as innovative. Prices would be set at European levels and would not be negotiated with the economic committee (CEPS).	Elicited (captured within the SMR) = place in the therapeutic strategy: if no other options at the same stage of the disease, based on the analysis of comparators and the description of therapeutic strategy (how the disease is treated, where the drug would fit and what are the current existing alternatives). E.g. a real unmet medical need would be recognised when there are no other treatment options.	Elicited (captured within the SMR) = Different categories of severity defined: severe, life-threatening, short life expectancy, affects quality of life, not so severe.

10.5. Discussion

10.5.1. Differences in the HTA process and application of HTA, implications for orphan drugs

The HTA approach adopted, in terms of clinical benefit and cost-effectiveness, the subtle differences in the acceptability criteria of evidence and uncertainty outlined in the results, and the willingness to accept greater uncertainty in specific circumstances relating to rare diseases may have implications on the assessments. These are discussed here, together with their implications when valuing orphan drugs.

Despite the known limitations in generating robust evidence for orphan drugs (Vickers, 2013), formal evidentiary requirements are similar for orphan and non-orphan drugs, with the exception of HAS that accounts for prevalence when examining the evidence. Higher evidentiary requirements are also seen for superior efficacy by TLV and high ASMR ratings by HAS. This has implications for orphan drugs often characterised by a lack of treatment options (Kesselheim et al., 2011), whereby a new treatment would likely be a superior one. Demonstrating survival benefits or the more clinically relevant benefits to patients usually require well-designed phase II trials or phase III trials (Korn, Freidlin, Abrams, & Halabi, 2012; Wieand, 2005), and treatment effects should be greater in smaller trials to attain statistical significance (Boudes, 2013). Innovative trial designs exist to deal with small patient trial populations (Gagne et al., 2014), and could be complemented with historical data from registries (Boudes, 2013). As shown in our study, the use of innovative trial designs and registry data is still limited, likely because their quality is often poor. In line with the feedback received during the interviews, registries are often difficult to analyse and time consuming as it should capture what is historically known about the effectiveness of a product (Haffner, 1998). Their usefulness, particularly for rarer conditions, is recognised in collecting information about the patient experience and natural history of the disease. Collecting this evidence over time has shown to improve its quality and reliability (when collected from patients, particularly if they understand the added-value of collecting this data) (Howie, Hirsch, Locklear, & Abernethy, 2014).

Evidentiary requirements for cost-effectiveness (NICE, SMC, TLV) differ from clinical benefit assessments (HAS). In the former, the estimate of clinical benefit is tailored to fit the economic model to represent clinical practice in terms of the magnitude of effect (e.g. life extension and/or HRQol) over a period of time (e.g. life-long), resulting in a single quantified measure (ICER). Assumptions are required about these determinants. In the latter scenario, clinical benefit is considered the hard endpoint, and quality of life and other qualitative criteria are accounted for as soft endpoints during the deliberative process. The approach used had a number of implications, some of which are discussed in this section. One important consideration is in the tools made available to deal with uncertainty when conducting an economic evaluation, where sensitivity analysis may help assess whether uncertain evidence is acceptable or not. This is true for any parameter tested within the economic model, and may be more relevant for orphan drugs given their greater uncertainty.

A distinction was seen in assessing the appropriate trial length, which is particularly relevant in the context of orphan drugs given they are often characterised by shorter clinical testing phases compared to nonorphan drugs (Kesselheim et al., 2011). This may relate to challenges in defining the appropriate trial length particularly when the natural history of the disease is unknown (Vickers, 2013) or when the disease is chronic or has an early age of onset; as reflected in our findings. The criteria in assessing the appropriate trial length were similar across countries and related to the natural course of the disease and likely duration of the treatment. Further, TLV relies on the judgment made by marketing authorisation authorities. Nevertheless, pivotal trials may not be sufficiently long to capture the benefits of a drug in clinical practice, particularly for life-long conditions (as shown for mannitol dry). A further consideration relates to the temporary authorisation (ATU) applicable in France for severe or rare conditions, which may also contribute to a greater acceptance of uncertainty. This could be illustrated by mannitol dry, which was not under an ATU at the time of the assessment, and for which one of the main issues highlighted by HAS was the uncertainty around its long term benefits (in addition to other issues).

Another contrast was seen in the consideration of subgroup data, whether within or outside of an economic model. There was agreement that if a particular subgroup

drives the cost-effectiveness results, the indication should be restricted to this subgroup. Nevertheless, differences were seen in practice, where some impose restrictions and others future re-assessments, as was the case for romiplostim with NICE and SMC versus TLV respectively. In contrast, this treatment received top ratings by HAS (important SMR, and ASMR II). In addition to the preferences seen for pre-specified subgroups, it should also be the same as for marketing authorization or included in the trial for HAS. Subgroup analyses in trials are included to identify whether certain patient groups are more likely to benefit from treatment, and are usually defined by their characteristics, such as age, sex, stage of disease, genomics characterized by biomarkers, etc. Despite the increasing body of research around predictive biomarkers, very few of these are included in the licensing indication and likely do not reflect their use in practice (Malottki, Biswas, Deeks, Riley, Craddock, Johnson, & Billingham, 2014). Therefore, the marketing authorization indication may not necessarily reflect clinical practice. A review of 894 RCTs showed that half of these reported subgroup analyses, of which 46% were planned in the trial protocols and 10% of those matched those reported in the publication (Kasenda, Schandelmaier, Sun, von Elm, You, Blumle, Tomonaga, Saccilotto, Amstutz, Bengough, Meerpohl, Stegert, Olu, Tikkinen, Neumann, Carrasco-Labra, Faulhaber, Mulla, Mertz, Akl, Bassler, Busse, Ferreira-Gonzalez, Lamontagne, Nordmann, Gloy, Raatz, Moja, Rosenthal, Ebrahim, Vandvik, Johnston, Walter, Burnand, Schwenkglenks, Hemkens, Bucher, Guyatt, & Briel, 2014). Another study showed that when the primary endpoint was not statistically significant across the whole patient population, subgroup analyses were most likely to be reported, particularly for industry-sponsored trials (Sun, Briel, Busse, You, Akl, Mejza, Bala, Bassler, Mertz, Diaz-Granados, Vandvik, Malaga, Srinathan, Dahm, Johnston, Alonso-Coello, Hassouneh, Truong, Dattani, Walter, Heels-Ansdell, Bhatnagar, Altman, & Guyatt, 2011). Subgroup data is therefore to be assessed with caution, particularly given that this evidence is likely to be even more uncertain (with greater confidence intervals) when the trial population is small, as is commonly the case for orphan indications. In such cases, considerations should be given as to whether the trial subgroup was pre-specified and matched what was planned in the trial protocol, and if discrepancies were to be seen, to impose a post-marketing follow-up to collect additional longer-term evidence and assess the effect of the treatment on a hard endpoint within clinical practice.

Issues relating to the quality or availability of comparative evidence is more common for orphan drugs, as they are more likely to rely on single arm and non-randomised studies (Kesselheim et al., 2011). This was illustrated for three cases: mannitol dry, trabectedin, which compared different doses of treatment, and ofatumumab (phase II non-comparative trial). This was translated into an ASMR V rating in France, given that it was not possible to assess the effect of the treatment compared to usual care. In contrast, more flexibility was granted by NICE and SMC because of the investigational nature and rarity of the treatment, and existing unlicensed alternatives (trabectedin). Such contrasts also illustrate consequences from misalignments between marketing authorisation and HTA processes. In our sample, the comparators were a consequence of the early marketing authorisation granted for ofatumumab and trabectedin, and the early scientific advice for mannitol dry (Bilton, Robinson, Cooper, Gallagher, Kolbe, Fox, Jaques, Charlton, & Investigators, 2011).

Surrogate endpoints are more common for orphan drugs compared to nonorphan drugs (Joppi et al., 2013; Kesselheim et al., 2011), further confirmed in our study (80% of drugs). Additionally, different levels of acceptability of progression-free survival was seen, which may explain differences in the assessments in two cases (e.g. mifamurtide, imatinib). There were subtle differences in the endpoints to use for the validation (hard versus soft endpoint), which may have implications for orphan drugs given also their often questionable clinical relevance (e.g. 6-minute walk or platelet response) (Joppi et al., 2009) or difficulties in establishing their validation (Boudes, 2013). Evidence suggests that surrogate endpoints tend to overestimate treatment effects, which can be minimized by quantifying their magnitude and certainty through their validation with the relevant patient outcomes (Ciani, Buyse, Garside, Pavey, Stein, Sterne, & Taylor, 2013). As with subgroup data, ongoing data collection through registries or other sources may render these uncertain outcomes more acceptable pending longer-term data about the clinical endpoint (e.g. overall survival).

A last significant contrast in the acceptability criteria for evidence was seen in the consideration of HRQol data as hard (captured in the economic models) or soft endpoint (HAS). In the last case, it may be the case that a lack of improvement in

quality of life data may have greater implications than if captured in an economic model, as seen for everolimus and mannitol dry. Further, for cost-utility modelling, generic utility data is generally preferred, despite not always being the most appropriate way to capture quality of life (Tordrup, Mossman, & Kanavos, 2014). This may have implications for rare diseases given they are often chronic, severe and disabling diseases, affecting quality of life and beyond, such as aspects of hopelessness linked to illness chronicity, or the search for normalcy in being part of a community and gaining social recognition (Caputo, 2014).

In addition to the different preferences and levels of acceptability of uncertain evidence, qualitative criteria are also accounted for in these processes. The most frequently identified being innovation, unmet need and severity, which have played an important role in modulating the decisions in accepting greater uncertainty or ICERs. Trivial differences were seen in the way these concepts are captured or defined. Despite common agreement about the definition of innovation (treatment benefits to patients), differences were seen in the way it is being accounted for: explicitly by NICE and HAS (through the ASMR), and captured within the ICER (or implicitly as a value judgment of the Committee members) by SMC and TLV. Examples of cases when innovation was highlighted in the decisions include the oral administration benefit of eltrombopag highlighted by SMC and TLV, or the new principle of treatment of romiplostim highlighted by NICE and TLV. Contrast was also seen in the definition of unmet need, defined in two different manners: by the consequence of the decision (NICE and TLV) or by the lack of available treatment options (SMC and HAS). The first closely relates to the severity of the disease and how patients would be affected without the treatment, whereas the second accounts for alternative treatments (e.g. a new treatment with better adverse effects would cover an unmet need) without differentiating the degree of severity. While severity is captured with unmet need for NICE and TLV, it has little or no weight for SMC and HAS given it is not explicitly accounted for by SMC and although categorized by HAS, unpublished evidence suggests that it does not have any influence on the assessment. However, it may be reflected through the temporary authorisation (ATU) scheme, where greater uncertainty may be accepted for these drugs due to the ongoing collection of data

about efficacy or safety. This contrast is suggested in our sample but no evidence exists to support this statement.

This furthers the discussion about defining the social values which need preference elicitation, and how accountability for reasonableness and consistency in their use can be improved. For innovation, this would pertain to those aspects of innovation not captured by the ICER or ASMR, which could include aspects around managing and living with a disease, and their importance for patients. The main question around unmet need relates to whether it should capture severity and prioritise the most severe conditions or whether it is a way to ensure that patients have treatment options at each stage of the disease. Prioritising the most severe conditions would put less weight on certain "less severe" or non-life-threatening problems from living with the disease or taking the treatment (e.g. pain, adverse effects, reduced mobility).

10.5.2. Study limitations

The structure of the qualitative research review guidelines (RATS) were followed in order to ensure the quality and clear dissemination of the research (Clarke, 2003). Despite this, the study is not without its limitations. First, the interview questions were derived from the analysis of ten orphan drugs. While this sample may be considered limited and not representative of all issues surrounding orphan drugs, a number of scenarios nevertheless repeated themselves suggesting that many common issues have been covered. Additionally, the topics covered dealt with all different levels of evidence (primary, non-primary, outcomes, etc.) and a number of additional considerations, and on this basis, we assumed that the analysis was sufficiently comprehensive. However, the sample did not include ultra-orphan drugs and the issues highlighted may in reality be even more uncertain than those analysed. The main advantage of focusing the interview questions around scenarios that were encountered is that we were then able to compare the responses with how this was enforced in practice. Second, different levels of detail may have been captured during the interviews due to their varying durations. A second round of questions together with the tables summarising the interpretation of their interviews, were sent to the interviewees to ensure the comparability and reliability of the research. Third, the

differentiation between how these findings apply to orphan and nonorphan drugs may at times appear unclear. This was because orphan and nonorphan drugs undergo the same process. We were nevertheless able to highlight some of the issues specific to orphan drugs and focus on these, and explore whether some aspects of the process could be differentiated for orphan drugs in order to overcome some of the specific limitations encountered in appraising these. Finally, the interviewees in each agency were interviewed together. This allowed to foster discussion amongst the respondents and capture richer discussions. However, this approach did not allow to capture potential contrasting opinions across interviewees in one same agency, particularly considering that they were given the opportunity to prepare their answers in advance.

10.6. Conclusions and Policy Implications

HTA bodies agreed that decisions regarding orphan drugs are made in a context of greater uncertainty, as illustrated throughout the paper. Despite the broadly similar agreement seen in the evidentiary requirements or preferences, subtle differences were identified with respect to the circumstances under which uncertain evidence may be considered acceptable. These relate to differences in the expectations around the quality of the evidence dependent on the clinical claim, in the recognition of usefulness and acceptability of registry data (e.g. historical controls), in their criteria for acceptability of uncertain trial duration, subgroup data, comparative data, surrogate outcomes, HRQol data, or safety data, and the extent to which more flexibility is granted to uncertainty because of the rare nature of these conditions. These, together with the varying approaches used for HTA (e.g. clinical benefit versus cost-effectiveness) and the special considerations given to orphan drugs, may modulate the interpretation of the outcomes of HTA and explain differences in the HTA recommendations made across countries. The former relate to the ability to implement patient access schemes, the NICE end-of-life criteria, disease severity at TLV, and implementing lower discounting rates when the benefits are sustained in the long run, or the way qualitative criteria are accounted for and weigh on the decisions. The latter ones specific to orphan drugs relate to the SMC modifiers, and possibly also those orphan drugs that are made available through the temporary authorization scheme (ATU) in France, where the continuous collection of data may contribute towards a

greater acceptance of uncertainty. These three categories of modulating factors may contribute towards accepting greater uncertainty or higher ICERs that what would normally be permitted, and explain the significant differences in the HTA recommendations made across the four countries.

Better understanding these modulating factors is essential to improve HTA and decide whether these processes should be differentiated in particular circumstances and how. For example, continuous data collection and re-assessment is even more relevant for rare conditions given the substantial uncertainty characterizing their assessments at the time of HTA – as illustrated in this study (e.g. trial length, surrogate outcomes, subgroups). Policy-makers should ensure that the correct incentives are implemented to collect this data at early stages of the drug development process. The limitations around RCTs, in collecting comparative data and implications on the decisions were also highlighted in this study, which only reinforces the need to ensure that what is being measured for HTA is responsive to patient needs, preferences and values, through continuous involvement of patients throughout the whole drug development process. Their input could help determine, for example, whether overall survival or progression-free survival respond better to patient needs in a particular disease setting. The fact that rare diseases often affect children may highlight the need to consider in a more explicit manner the circumstances when lower discount rates should be implemented to reflect the long term effects of treatments, based on existing evidence around their appropriate use (Severens, 2004). This study contributed to showing the reasons why current systems, and the HTA methodological approaches being used, are not sufficiently suitable to tackle the issues that relate to rarity, as highlighted not only be the magnitude of and conflicting differences in the HTA recommendations made, but also be the contrasts seen in the various ways of dealing with these issues emerging from the rare nature of the diseases they treat. This is all the more important in a changing pharmaceutical environment that is shifting towards more niche and targeted therapies ("the right patient with the right drug at the right dose at the right time"), where such issues will soon become a daily reality.

11. Conclusions

11.1. The Contribution of this Thesis

This section outlines the main contributions of this thesis grouped into three categories: methodological and empirical contributions, and how findings fit with existing research, which are discussed here.

11.1.1. Methodological contribution

This thesis used a mixed methods research design to address a specific empirical research question. This type of study has never been done before in the area of HTA. It shows the inter-disciplinary potential of applying mixed methods research designs to novel areas in social sciences. Specifically, an instrument-based model in the form of a methodological framework was developed and piloted (Chapter 7), which allowed for an enhanced understanding of complex decision processes in different HTA settings. This was done using a sequential exploratory mixed methods design, which captured both the depth and breadth of these decisions. It provided a structured understanding of the decision-making process, by breaking it down into three distinct steps, each comprising a specific type of criteria (evidence, its interpretation, and influence on the final decision). This structured categorisation is also what allowed to conduct cross-country comparisons of the criteria that were accounted for in each setting (commonalities and differences across countries). Each step undertaken to set up and test this framework was outlined in detail to ensure its transparency and also transferability to third parties, by means of a coding manual and case study template (Figure 7-3 and Appendix B).

The novelty of this research is twofold: (1) its innovative design through the integration of both a vertical and horizontal component; and (2) the applicability of the methodological framework to other HTA settings and disease areas given its iterative and flexible nature. The vertical component comprises coding each individual drug in a systematic, homogeneous and comprehensive manner (based on the existing and flexible coding manual) in order to quantitatively devise (through correspondence

analysis) agency-specific preferences in the type of evidence, concerns (e.g. risk preferences), and "other considerations" (e.g. value preferences) identified. The horizontal component of applying the framework enables to capture how HTA processes compare across countries, whether the criteria had a positive or negative influence on the decision, where the criteria put forward came from (e.g. stakeholder input), and whether the criteria were one of the main reasons for the final recommendation.

Therefore, the main methodological contribution of this research is the methodological framework developed, piloted and tested, which allows to systematically compare HTA decision processes across settings and drugs. Through its application, it is possible to identify a more comprehensive range of criteria accounted for during the decision processes in a structured manner, including those that go beyond standard measures used for HTA (e.g. ICER). It also allows for a comparison across countries of these criteria identified as influencing these processes.

11.1.2. Empirical contributions

Four main empirical contributions resulted from this thesis in addressing its overarching objective to understand why there are differences in HTA recommendations across countries for a same drug and indication, and whether there is consistency with countries, focusing on orphan drugs. The empirical contributions follow the same structure as the sub-research questions addressed in this thesis and repeated again in the boxes below.

Findings from this thesis first contributed to understanding the commonalities and differences in the HTA recommendations made across countries, and also across therapy areas.

Question 1

What are the commonalities and differences in HTA recommendations made across five countries and three therapy areas? On this basis, are these differences meaningful?

Variations in HTA recommendations across countries were already recognised at the time this thesis commenced. These may be a consequence of context-specific considerations, such as WTP thresholds, available budgets, context-specific costs or national priorities. A number of studies explored this matter, however, the extent of and reasons for differences were not thoroughly scrutinized. This thesis further demonstrated the magnitude of these differences across a larger sample of 287 drug-indication pairs and five countries, where as much as 46% of the 122 compounds appraised by at least two agencies received diverging recommendations across countries. The extent of these differences suggests that other reasons may cause this divergence, such as differences in the methods adopted or in the interpretation of the same evidence.

When comparing the acceptance rates across therapy areas, the assumption was made that the same trends would be observed across countries for the same drugs. For example, if the cancer drugs included in the sample had a greater acceptance rate in one country compared to the other therapy areas, the same trend would be assumed in the other countries. This was not seen in practice, where different countries were more likely to accept certain therapy areas compared to others, suggesting that expectations in terms of relative effectiveness differ depending on the drug and disease characteristics despite agency-specific guidelines being generally homogeneous for all drugs (Chapter 6). Therefore, the first empirical contribution of this thesis was to demonstrate that these differences are meaningful because of their magnitude and the contradictory trends observed across countries, and that there is a need to query why they occur.

This is all the more important in the current European and international context, where various initiatives aim to foster greater cooperation and incentivise cross-country learning and sharing of expertise. This includes the EUnetHTA at European-level that aims to enhance HTA cooperation across countries towards a system of knowledge sharing and promotion of good practice and methods (European Commission, 2015a), or the willingness to minimise different levels of access to medicines across European Member States with the EU Cross Border Directive (Directive 2011/24/EU). The

final conclusions and recommendations of the Pharmaceutical Forum also recognise the need to improve access to orphan medicines for all EU citizens, in line with their objectives to "offer equal access to medicines at an affordable overall cost" (Pharmaceutical Forum, 2013). Consequently, understanding the reasons for differences can contribute towards these goals, and ultimately, improve and harmonise patient access to treatment across EU Member States.

Given that these differences in the HTA recommendations issued across countries matter, the second research question in this thesis aimed to understand why countries issued different HTA recommendations. The analysis focused on ten orphan drugs appraised in four countries (N=35).

Question 2

Why are there differences in HTA recommendations for orphan drugs in four countries? What can we learn from these differences?

Through the application of the methodological framework, this sample of drugs and countries were systematically compared. One of the first main observations from the results is that no clear and systematic reasons for differences across countries were identified. Instead, differences were identified at each step of the decision-making process (as per the structure of the framework developed, e.g. evidence, its interpretation, influence on final recommendation) and were non-homogeneous within and across countries. No clear pattern emerged. Differences were seen in the main parameters considered of interest by the HTA bodies despite coming from the same primary trials. When assessing the same evidence, heterogeneity in the concerns raised was observed (e.g. in what was considered uncertain by the agencies). In certain cases, but not for all, these concerns were in line with those types of concerns more commonly raised by some agencies compared to others (e.g. agency-specific risk preferences, identified through correspondence analysis). There were also instances where the same concerns were raised by more than one HTA body, but were dealt with in different ways. Different means were used to address uncertainty in addition to existing standard tools (e.g. sensitivity analysis). This was achieved, for example, through expert opinion or the assessor's own judgments during the deliberative process

of HTA. These value judgments are about the scientific evidence or about social values, and whether they are willing to pay more for drugs that fulfil certain characteristics. Some of the social values identified were more likely to be made by some agencies compared to others (e.g. agency-specific value preferences, identified through correspondence analysis). All these contributed to explaining some of the differences in the HTA recommendations made across countries.

In summary, this thesis contributed to generating the ability to identify the various reasons explaining differences in the HTA recommendations, which includes also the identification of the softer endpoints (e.g. value judgments) and their influence on the final decision. In addition to the ICER or magnitude of the clinical benefit, these decisions are also influenced by the type of evidence preferred, the different attitudes (agency-specific preferences) or means (e.g. sensitivity analysis, expert opinion, additional evidence) in accepting uncertainty, the availability of certain modulating factors that may help improve the ICER (e.g. PAS, SMC modifiers) or measure of clinical benefit (ATU), and finally, a subjective component that depends on the Committee's judgment (e.g. value judgments). All these reasons relate to how agencies are dealing with uncertainty, which always rely on the scientific and social value judgments of the assessor's and depend on the level of ambiguity around what constitutes acceptable evidence or acceptable ways to deal with uncertainty, the processes or means available that may allow to modulate the ICER to an acceptable level or allow for a greater acceptance of uncertainty (future re-assessment).

Shedding light on these cross-national differences also allows to query why they occur and what we can learn from them, particularly in those cases where they may reflect weaknesses in the HTA methodological approaches used. For example, issues relating to the limitations of RCTs (generalisability, safety) are well-known, nonetheless, they are still considered the most robust and preferred type of evidence. Given that the majority of primary trials were RCTs, these issues were unavoidably likely to be raised, as seen in the results. This contributed to understanding how these issues were dealt with in the different settings. For instance, generalizability to clinical practice was a frequent concern for NICE and SMC, and was most often addressed through clinical expertise in the former case, and had a negative influence on the assessments

in the latter. Other issues included that (a) patients with certain characteristics were not included in the trial or around the heterogeneity of the subgroups appraised, (b) the trial population did not represent the indication under review, or (c) imbalances were seen in terms of the characteristics or responses across the different subgroups. This highlights the potential value of accounting for other forms of evidence, such as expert opinion or observational data to complement RCT evidence.

Raising awareness about these differences (and similarities) facilitates cross-country learning including sharing practices about how value is assessed in different settings. Better understanding these processes and how value is assessed in different settings can be useful for all stakeholders involved in the process: HTA bodies, manufacturers, patients, clinicians, payers, etc.

One of the components contributing to addressing uncertainty and explaining differences across countries were the value judgments made and identified through the application of the framework developed. This thesis contributed to furthering the understanding around what these value judgments look like in practice, how they can be categorised, and whether orphan drugs have special status (Chapter 9).

Question 3

How do scientific and social value judgments influence HTA decisions? And on this basis, do orphan drugs have a "special status"?

An important contribution is the classification framework proposed about the value judgments made that distinguishes those judgments about the scientific evidence, usually quantified, from those about social values, usually elicited. This is an important distinction as the former relates to the acceptability about uncertain or incomplete evidence and the latter relate to non-quantifiable aspects of living with a condition and taking a treatment. Both of these occur during the deliberative process and rely on the decision-maker's experience, on what they believe society would prefer, or on conclusions of citizen's councils or juries. Based on this classification, it is possible to categorise these value judgments made and identified by applying the

framework, and differentiating those that were quantified or elicited from those that were not quantified or elicited, respectively.

Their identification contributes towards assessing whether these value judgments are consistent across cases. It also allows to better define the determinants of these social values based on how they were accounted for retrospectively (e.g. unmet need). For those value judgments that were not elicited, it may provide a good basis to seek elicitation about societal preferences, particularly for the more commonly identified ones (e.g. unmet need, innovation, severity). Their identification improves the lack of accountability for reasonableness particularly in cases when it is not clear how the "other considerations" identified influenced the decisions. It also contributes to eliciting whether these orphan drugs deserve a special status by eliciting preferences around some of the social value judgments made which are more likely to pertain to orphan drugs compared to normal condition, rather than focusing on the opportunity cost of these.

In the last empirical chapter of this thesis, insights from HTA body representatives were collected about the differences seen across countries or the challenges that HTA bodies were facing when appraising orphan drugs, with a particular focus on those relating to the rarity of the conditions. Results focus on some of the contrasts and subtleties identified across countries and how these may have contributed towards explaining some of the cross-country differences observed.

Question 4

How is the value of orphan drugs assessed across different settings and how do differences affect coverage decisions?

HTA bodies agreed that decisions regarding orphan drugs are made in a context of greater uncertainty. The main contribution of this last part of the thesis was to understand the contrasting approaches in the application of HTA when appraising orphan drugs. Despite the broadly similar agreement about evidentiary requirements or preferences, subtle differences were identified with respect to the circumstances under which uncertain evidence may be considered acceptable. These relate to

differences in the expectations around the quality of the evidence dependent on the clinical claim, in the recognition of usefulness and acceptability of registry data (e.g. historical controls), in their criteria for acceptability of uncertain trial duration, subgroup data, comparative data, surrogate outcomes, HRQol data, or safety data, and the extent to which more flexibility is granted to uncertainty because of the rare nature of these conditions. These, together with the varying approaches used for HTA (e.g. clinical benefit versus cost-effectiveness) and the special considerations given to orphan drugs, may modulate the interpretation of the outcomes of HTA and explain differences in the HTA recommendations made across countries. The former relate to the ability to implement patient access schemes, the NICE end-of-life criteria, disease severity at TLV, and implementing lower discounting rates when the benefits are sustained in the long run, or the way qualitative criteria are accounted for and weigh on the decisions. The latter specific to orphan drugs relate to the SMC modifiers, and possibly also those orphan drugs that are made available through the temporary authorization scheme (ATU) in France, where the continuous collection of data may contribute towards a greater acceptance of uncertainty. Shedding light on the three main categories of modulating factors contributes to understanding means to accepting greater uncertainty or higher ICERs that what would normally be permitted, which in turn explain the significant differences in the HTA recommendations made across the four countries.

11.1.3. How do findings fit with existing research?

Since the beginning of this thesis and the review of the literature conducted to identify the gaps in the literature, a number comparative studies that include the study countries have been published (Drummond, de Pouvourville, Jones, Haig, Saba, & Cawston, 2014a; Cerri et al., 2014; Maynou Pujolras & Cairns 2015).

Drummond and colleagues (2014a) contrast the advantages and disadvantages of the two approaches to HTA used in England and France (cost-effectiveness and cost/QALY versus added benefit (ASMR)). By comparing the assessments of QALY increases and ASMR ratings, they show that the two countries arrive at similar assessments of added clinical benefit, but arrive at different conclusions when NICE

accounts for costs and cost-effectiveness. Findings from this thesis are complementary and provide more detail about what happened to explain these differences. In terms of the contrasts seen between NICE's decisions and HAS's ASMR rating, two drugs (eltrombopag and imatinib) with a high ASMR (I-III) were considered cost-ineffective by NICE. However, this was not only a consequence of the costs being accounted for, but also of the clinical endpoints considered relevant in the decision problem (which were not necessarily the primary trial endpoints for NICE), resulting in high ICERs. Findings from this thesis additionally highlighted three contrasting cases that received the lowest ASMR rating (V), but were positively received by NICE (mifamurtide, mannitol dry, trabectedin). This was because of the lack of comparative evidence presented due to early marketing authorisations or early scientific advice received, where no added benefit was demonstrated. Given the particular circumstances of these three drugs, NICE was more lenient in accepting uncertainty and positively received the drug. The authors also observe that HAS's ASMR assessment is less transparent than NICE's process. This is supported by findings from this thesis, where the criteria (considered as soft endpoints) and their weight on the ASMR decision was often unclear.

The recent study by Cerri and colleagues (2015) compared the factors that contribute to explaining differences in coverage decisions in England, Scotland, the Netherlands and France. When comparing the same drugs on a wide range of variables that may affect the decisions, only 30% of the variability across countries could be explained by the model possibly because of the range of factors that are highly specific to each agency. Such factors could relate to what went on during the decision process and the combination of factors that rendered the decision positive or negative. Such subtleties would not be captured when exploring the role of a set of pre-defined criteria, but would require more qualitative exploratory approaches, as was adopted in this thesis. The authors also found that the strength of the evidence had a significant role in the decisions, while some variables such as the number of observational studies did not show any effect. This is congruent to results from this thesis highlighting the limited role of non-primary non-phase III trials. A number of variables showed to have a significant impact on the decision (e.g. population size associated with a decreased probability of recommendation, indication). These were not captured within this thesis

as they are characteristics of the disease and treatment and require a larger size and heterogeneous sample to be able to measure their effects.

Another recent study by Maynou-Pujolras and Cairns (2015) explored the extent to which a number of variables contributed to explaining different HTA recommendations in ten European countries comparing 199 drugs. These results have been integrated with results from this thesis in order to assess whether findings from these two qualitative and quantitative studies respectfully are congruent, complementary or discrepant. The paper is soon to be submitted for peer review (Nicod, Maynou-Pujolras, Visintin, & Cairns, 2016). Results show that the two approaches are often complementary, and capture different aspects of the decision-making process. For example, quantitatively, a cost-effective ICER was more likely to receive a positive recommendation, whereas this thesis showed that a cost-effective ICER likely to receive a positive recommendation is often a modulated one. There were also a number of variables that were excluded from the quantitative analysis as not explaining sufficient variability in the model. Many of these, however, were captured qualitatively in this thesis and their role in the decision was explained (e.g. influence of severity in Sweden). There were also some contradictory findings across these two studies, where, for example, higher levels of stakeholder involvement were associated with a lower probability of recommendation. This thesis showed the contrary, where stakeholder involvement often contributed to confirming or not the plausibility of a claim being made.

These last three studies adopted quantitative techniques to analyse larger samples compared to the comparative studies previously discussed in Chapter 3 and the findings from this thesis. Generally, the findings are in line with each other. However, there are certain aspects that can only be captured using one technique or another. For example, the qualitative approach used in this thesis did not allow to analyse a large sample or assess the impact of certain disease or treatment characteristics on the final decision (e.g. orphan status, population size). By contrast, it did allow to capture a more comprehensive range of criteria and the interaction amongst criteria in the decisions, many of which may not have been captured quantitatively. This includes the subjective component of these decisions, which may

be highly specific to the agency and/or to the decision-maker. This clearly emphasises the added value of adopting mixed methods approaches, where quantitative findings allow to enhance and build upon qualitative findings, and qualitative findings allow to explain, illustrate and foster the credibility of quantitative findings (Mertens, 2011). The limited amount of variability captured using quantitative approaches seen (Cerri et al., 2015) also emphasises the added value of adopting an exploratory approach, as was done in this thesis.

There were also a number of studies that investigated the drivers of the HTA recommendations in one specific country (Devlin & Parkin, 2004; Dakin, Devlin, & Odeyemi, 2006; Cerri, Knapp, & Fernandez, 2014; Svensson, Nilsson, & Arnberg, 2015; Dakin, Devlin, Feng, Rice, O'Neill, & Parkin, 2015).

The first study by Devlin & Parker (2004) investigated whether NICE has a cost-effectiveness threshold and other factors influence its decisions by conducting a binary choice analysis. They assess the effect of six independent variables (e.g. ICER, uncertainty) on the HTA recommendation. Results support the notion of a threshold, where the probability of rejection increases as the cost/QALY increases, with an additional effect of the burden of disease and uncertainty on these decisions. They echo findings from this research by stating that decisions are based on imperfect and incomplete evidence and rely on the decision-makers' judgments. Similarly to the conclusions made for the comparative studies, the quantitative approach does not enable to capture the depth of what went on during these decisions, while the qualitative approach used within this thesis does not allow to interpret the relative contributions of each of the variables assessed.

A more recent study conducting a similar analysis for NICE focusing on a wider range of variables also support that cost-effectiveness is the main driver of their recommendations (Dakin et al., 2015). This concords with findings from this thesis, where all drugs above NICE's cost-effectiveness threshold were rejected. What this study adds is that the ICER is often a modulated one as already previously discussed. The other variable that may have influenced the decision is the type of disease, which was not captured in this thesis given the focus on orphan drugs and the small sample

size. An earlier study by the same authors also found that patient submissions were more likely to increase the probability of endorsement for routine care rather than restricted use (Dakin et al., 2006). This further supports findings on the influence of stakeholder involvement in these processes.

The study by Cerri and colleagues (2014) examined the impact of evidence, process and context on NICE decisions by assessing the relative contribution of 32 variables using a multinomial logistic regression. Four of their variables had a significant effect on the decisions (demonstration of superiority, ICER, number of pharmaceuticals appraised in the same HTA, and the appraisal year). Only the first two variables were captured in this thesis, as the others would require a larger sample and those technologies undergoing the multiple technology assessment process (which was not the scope of this thesis). Findings about the ICER concur with this thesis as previously discussed, whereas those about clinical superiority were not explicitly captured when assessing cross-country differences, as they may have been reflected in the ICER or ASMR rating. This further supports the conclusions on the complementarity of using different approaches to obtain a better understanding of the drivers for decisions.

Finally, the study by Svensson and colleagues (2015) analysed the impact of cost-effectiveness and severity on HTA decisions in Sweden for 102 decisions. Their findings elucidate the willingness to pay for a QALY and the fact that higher ICERs are accepted for the more severe conditions. This echoes the findings from this thesis, where higher ICERs were generally accepted for the sample of orphan drugs analysed. The main question remains about the definition of severity, where it was categorised as severe or non-severe. This thesis further contributed to better understanding the attributes of severity by coding how these were expressed during these decisions (Chapter 9).

11.2. Policy Implications and Recommendations

In the past decade, health expenditure has increased at a faster rate than national GDP (Figure 2-1), stressing the issue about its sustainability. The European Commission has recognized this issue as one of the cornerstones in their strategy on “investing in

health", where cost-efficiency is to be obtained through innovation and measuring health system performance (European Commission, 2013a). "Ensuring efficiency and making the provision of health services more cost-effective and efficient is crucial if countries are to ensure universal access to and equity in health services and their adequate and sustainable financing" (Directorate-General for Economic and Financial Affairs, 2012; European Commission, 2013a). Ensuring efficiency was also understood as a vital element for the sustainability of a health system's performance during the 67th WHO World Health Assembly, which "urges" Member States "to consider establishing national systems of health intervention and technology assessment, encouraging the systematic utilization of independent health intervention and technology assessment in support of universal health coverage to inform policy decisions…" (Sixty-seventh World Health Assembly, 2014).

Understanding the drivers of health expenditure is a way towards understanding where improvements are needed. These include the ageing population, increasing costs of care and labour, growing demand, as well as inefficiencies in the organisation and payment of care (Sorenson et al., 2013). In terms of the drivers for pharmaceutical expenditures, one systematic review identified these as increased drug quantities and the introduction of new and expensive drugs (Mousnad et al., 2014). New drugs generally carry high prices, which are justified by the investments in research and development and the added benefit of the treatment. A more recent study identified another potential cause for high drug prices in cancer that they referred to as the "market spiral pricing effects", where prices of last year's drugs are increased and prices of new drugs are set above the new market price (Light et al., 2013). The authors further argue that research and development costs are over-estimated (by about 10 times), that generally these drugs provide little or no added value, and that high prices are a consequence of monopolistic positions.

The complexity of these market dynamics, as also highlighted in Chapter 2, make it all the more challenging for policymakers to make the right resource allocation decisions, when already their objectives are competing in ensuring access to safe, efficacious and quality care while containing costs and incentivising innovation. Our systems should ensure that incentives for research and development target those drugs providing

additional benefits to patients and society, and outweigh their costs. It is generally recognised that HTA is one way to achieve this. However, we've seen that the successful implementation of HTA depends on various factors including the availability of evidence to demonstrate the value of these technologies (Sorenson et al., 2013).

This thesis contributed to better understanding the different ways in implementing HTA through cross-country comparisons and by elucidating the issues HTA bodies are facing during these assessments, including those that relate to dealing with rare conditions, and how they are dealing with them. As emphasised in Chapter 3, these differences are often legitimate because of the complexity of these processes and context in which they operate (Banta, 2003). A distinction should also be made between the assessment and appraisal phases of the processes, where it can be expected that the assessment is similar across countries while the appraisal differs according to context-specific considerations, which may frequently be accounted for ad hoc. In this context, decisions are inevitably based on the decision-makers' judgments about the evidence, influenced by their preferences and own knowledge and this will always be legitimate. However, the magnitude, and contradictory nature of the differences seen in HTA recommendations made across countries further emphasise the need to understand the reasons for these and understand when these are a consequence of weaknesses or limitations in the application of HTA approaches, and whether they occur during the assessment or appraisal phase. Policy implications and the recommendations from the findings of this thesis are discussed here.

11.2.1. Policy implications

The main contributions of this thesis discussed in the previous section can be summarised around three levels. First, the framework developed enables to capture a more comprehensive range of criteria accounted for during these decision processes, systematically and in a comparable manner. Second, it allows to raise awareness of those cases when the reasons for cross-country differences are a consequence of HTA methodological limitations and highlight areas for potential methodological improvements. Finally, it also highlights the challenges HTA bodies are facing when

assessing orphan drugs and how they deal with these. Better understanding the application of HTA in different settings has implications for policy, which are discussed here.

The identification of a more comprehensive range of the criteria accounted for across all drugs, therapy areas, indications, and countries is useful in several ways. Identifying the type of criteria that contributed to decision-making at each stage of the process provided a more structured understanding of these decisions and improved their transparency. This may help improve their accountability for reasonableness, which is obtained when it is transparent and public, based on reasons that are relevant, revisable when new evidence is available, and with a decision-making process that allows for these conditions to be enforced. This was emphasised in the findings from Chapter 9 (Paper 4) suggesting that although the study countries were relatively transparent in reporting the rationale for their decisions, it was not always clear how the different criteria weighed on the decisions. For example, it was not clear how disease severity was accounted by TLV apart from accepting greater ICERs. The criteria identified also included the softer endpoints made as part of the deliberative process, referred to as the scientific and social value judgments (e.g. elicited/quantified or not), which had not been identified in such a systematic manner previously together with how they influenced the decisions. This, together with the agency-specific preferences observed, are a way forward to improving the consistent use of these "other considerations" or preferences, while better defining their attributes based on previous applications. This will also improve the understanding of the expectations HTA bodies have in the submissions, which will minimise unnecessary and resource-consuming rejections and re-submissions.

The identification of a more comprehensive set of criteria accounted for also has practical implications in other priority setting applications, such as MCDA or discreet choice experiment (DCE). MCDA is a technique supporting decision-making allowing to account for explicit criteria without quantitative modelling (Thokala, Devlin, Marsh, Baltussen, Boysen, Kalo, Longrenn, Mussen, Peacock, Watkins, Ijzerman, 2016). Different uses for MCDA exist together with different methodological approaches. For example, DCE is one way to involve patients in these

decisions by eliciting their preferences. This is done by asking them to choose between two different scenarios with different attributes and repeating the exercise changing the attributes to choose from. With regression analyses techniques, it is possible to elicit their preferences for the different attributes (Thokala et al. 2016). Additionally, MCDA techniques enable to give explicit weights or categories to criteria in decisions through various means (e.g. stakeholder participation) (Angelis & Kanavos 2016). The same authors also argue that it is a way to capture unexplained heterogeneity rather than wait until the appraisal phase to capture these. This is because standard cost-effectiveness approaches account for costs and effects during the assessment phase and additional criteria during the appraisal phase, whereas with MCDA, all these criteria can be accounted for in the assessment and appraised together. One illustrative example is the way unmet needs were defined in the context of MCDA within the EVIDEM project, which accounted for the type of therapeutic benefit (e.g. efficacy, safety, etc.), but did not capture the different levels of needs (e.g. lack of satisfactory alternatives). Findings and the application of the framework developed in this thesis may also be useful for EU-level collaboration initiatives, such as EUnetHTA, in highlighting the differences seen across countries that can then be discussed when facing similar scenarios in the future. Findings may also contribute to recent initiatives, such as PACE in Scotland and the highly specialised technologies programmes at NICE, which account for additional criteria beyond the ICER in their decision-making processes, such as disease severity or unmet need. Retrospectively identifying how these factors have emerged in practice and were accounted for may feed into better defining and accounting for them in future cases.

Identifying the reasons for cross-country differences pertaining to the varying application of HTA enabled to further the debate about some of the limitations around current HTA methodological approaches. Table 7-2 summarises the reasons for differences in HTA recommendations across countries at each step of the process in terms of the evidence appraised and its interpretation. These differences were used to further the debate on some of the limitations around current HTA methodological approaches, namely:

- RCT limitations. This thesis showed that for the range of treatments studied RCTs continue to be the main source of evidence accounted for in these HTA decisions and that other forms of evidence had a limited effect. RCT limitations are well-known and include issues around safety and generalizability to heterogeneous populations of clinical practice. These issues were raised in the decision processes analysed and different ways in dealing with them were seen across countries. In some instances, it had a negative influence on the decision, and in other instances, it was considered acceptable through various means, such as stakeholder input.

- Use of non-phase III trials. Results showed that there were only a few cases when non-phase III secondary evidence had any influence on the decision. Outcomes from these trials were generally not reported, and when reported, the type of data provided was around safety, dosage research or historical controls. The uptake of such forms of evidence is still modest and likely due to the lack of expertise around dealing with a variety of types of observational evidence including those based on real world data such as electronic patient records (Berger et al, 2014) or patient-reported outcomes (McClimans and Browne, 2011). More clarity and research is needed around the quality standards for when such types of evidence can be deemed acceptable or not.

- Differences in the evidence appraised. Differences were seen in the level of evidence reported (different trials, different endpoints from the same trials or different levels of analysis). In some instances, these differences were also one of the explanatory factors for diverging decisions. These included registry data as historical controls for trabectedin (NICE), different primary endpoints for mifamurtide (overall survival for NICE and progression-free survival for SMC and HAS), the secondary endpoint “severe bleeding events” for eltrombopag only reported by NICE, the lack of quality of life data in the assessment of eltrombopag for HAS and TLV. More clarity is needed on the expectations from HTA bodies on the reasons for including some of these endpoints, which were not included by others. Specifically, more clarity is needed around how the endpoints of interest are selected and their level of importance.

• Differences in the uncertainties raised. These refer to cases when one HTA body raised a concern about the evidence, which was not raised by the others. This thesis showed that these are likely to be a consequence of agency-specific risk preferences or of the interpretation of the decision-making Committee, showcasing the subjectivity of these decisions that rely on the decision-maker's own preferences and experience. This is legitimate given the uncertain nature of evidence in general within which they operate. Policymakers should be aware of these happenings and query the different methodological approaches to minimising or better dealing with uncertainty (e.g. real world evidence, patient input). It is also important to continue to query and shed light about these preferences (as was done in Chapter 10) such as to improve transparency around the expectations from HTA bodies.

• Dealing with the same uncertainties. Agreement was generally low to less than expected by chance in dealing with the same uncertainty. In other words, when one agency addressed an uncertainty by various means, there was moderate to less than zero chance that the same uncertainty was addressed by another agency. A number of different ways were seen in dealing with these (Table 11-1). Such cases could be a good starting point for policymakers to discuss alternative scenarios in view of establishing criteria for accepting greater uncertainty in future cases. For example, in the case of trabectedin, due to the early marketing authorisation granted, the assessment relied on a phase II non-comparative study. Under certain circumstances historical control data could be acceptable.

Table 11-1. Dealing with uncertainty

Type of uncertainty	Criteria for acceptability	Examples
Non-significant improvement in clinical benefit (overall survival)	Orphan status	SMC for imatinib
	Expecting results from an on-going trial	HAS for imatinib
Risk of severe adverse event (risk of bronchospasm)	Expert opinion	NICE for mannitol dry
Risk of interaction between treatments	Expert opinion	NICE for mifamurtide
Lack of comparative	Rarity, early market	NICE for trabectedin

evidence (pivotal phase II trial)	authorisation, use of historical controls	
	Rarity, investigational nature of the treatment	SMC for trabectedin

• Value judgments. It is generally recognised that more transparency and consistency in the social value judgments made across cases are needed (Daniels, 2000; Daniels and Sabin, 2008; Earnshaw and Lewis, 2008). This can be achieved by categorising the types of value judgments made based on the classification framework developed, and on this basis, ensure that they are accounted for consistently across cases. It also contributes to identifying those value judgments made where further elicitation of societal preferences are needed, and to better defining these based on previous decisions. Better understanding the value judgments made can contribute to informing new initiatives such as the value-based pricing system or the highly specialised technology processes in England, or the ultra-orphan approach in Scotland, which account for additional criteria such as disease severity, unmet need, or the added value of the medicine for the patient, their carer or family (Brown, 2014).

• Societal perspective at TLV. None of the criteria identified pertained to adopting a societal perspective. More clarity around how this is accounted for during the decision-making process is needed.

Finally, the challenges HTA bodies are facing when assessing orphan drugs were also identified together with the different ways of dealing with these across countries. This is a consequence of the particularities in assessing orphan drugs due to the small patient numbers involved, where they are hardly cost-effective (Drummond et al., 2007). Despite this fact, evidence suggests that orphan drugs are prone to receive a similar or greater level of acceptance for reimbursement than other more common conditions (Dupont et al., 2011; Simoens et al., 2011; Stolk et al., 2009). The types of concerns raised that related to the rarity of these conditions were identified. This included issues in recruiting sufficient patient numbers or cases when results relied on subgroup data. The assessments also relied on (validated or non-validated) surrogate endpoints and frequently lacked quality of life data. These scenarios can be considered

common for rare diseases given the scarcity of knowledge and expertise, and the difficulties in generating robust evidence. The main question was to understand whether HTA bodies are accepting greater uncertainty in the evidence, higher ICER estimates, or are negatively assessing these drugs because of this. In some cases, this uncertainty relating to orphan drugs (but not necessarily specific to), was considered acceptable because of the condition's rarity or the recognised difficulties in recruiting sufficient patient numbers. In contrast, certain concerns were more commonly negatively appraised. This related to, for example, dealing with subgroup data (seen in 50% of cases), where the concerns remained inconclusive because of their lack of statistical power or retrospective nature. Another contrast was seen in the lack of comparative data, characterising 30% of the study drugs (likely due to the rare nature of these conditions where few or no alternatives exist, and therefore for ethical or early marketing authorisation, they did not include placebo comparisons). Given that HAS relies on the added benefit of the drug, this evidence was insufficient to demonstrate any added benefit and therefore received the lowest ASMR ratings (V). More flexibility was granted to this lack of comparative data in the other countries because of their rarity or investigational nature. More consensus in HTA processes is needed in dealing with these specific and common issues related to rarity.

In summary, the concerns raised often related to the well-known evidentiary issues seen in orphan drugs from the small patient populations, scarcity of existing knowledge about the disease, unmet need, or heterogeneity of these conditions. Different patterns were observed in dealing with these, where some accepted greater uncertainty for certain types of concerns (e.g. lack of statistical significance) under certain circumstances (e.g. post-marketing surveillance and planned future re-assessment), others used various means to modulate the ICER to an acceptable level (e.g. PAS, restriction to the most cost-effective subgroup, NICE end-of-life criteria, SMC modifiers, TLV and disease severity), additional considerations may have been accounted for particularly in those cases where it was recognised that the measures of clinical benefit provided did not capture the full effects of living with the disease and taking the treatments, or finally, they may have simply a negative impact on the decisions (e.g. lack of comparative data in France). Policymakers should be aware of these limitations and account for, in a more consistent way across cases, the different

approaches in dealing with these situations. Identifying these special cases and how they were dealt with is a good starting point to generating a set of criteria when greater uncertainty may be accepted in future cases.

11.2.2. Varying approaches to using HTA

When examining the different uses of appraising the clinical and health economic evidence within HTA, different approaches across countries were identified despite being driven by the same fundamental objective of clinical benefit and/or cost-effectiveness. Those countries with a formal (or informal) ICER threshold, such as NICE or SMC, would tend to be more instrumental, where anything above the threshold would not be considered cost-effective. For example, in the case of NICE, the starting point is always the ICER derived from the cost-utility analysis, its magnitude and certainty. If the most plausible ICER is reasonable (within the £20,000/QALY threshold), it would be considered cost-effective. However, if it were based on unreasonable assumptions, the drug would not be recommended or additional information would be requested (if possible and plausible). If the ICER is greater than the acceptable threshold of £20,000/QALY, additional considerations would be accounted for to understand whether the ICER fully captured the effects of the disease and the treatment, or whether the treatment can be considered an end-of-life treatment, or whether the company accepts to decrease the price to make it cost-effective for that promised effectiveness. This takes place during the deliberative process, where the ICER would be used for enlightenment during the decision.

In the case of TLV, which does not have a formal cost-effectiveness threshold, the starting points are the clinical claim made (e.g. superiority or non-inferiority) and the consequence of the decision (severity of the consequences if patients were not to receive the treatment, which would be greater for those with a higher unmet need). The ICER is therefore considered within this context, where higher evidentiary requirements are seen for higher clinical claims and price premiums, and where greater flexibility would be accepted for uncertain evidence for non-inferiority claims. Additionally, the greater the severity and unmet need, the more high ICERs would be acceptable.

In the case of HAS's coverage decision, the consequence of the SMR and ASMR ratings is somewhat different than in the other study countries (where a drug with a negative HTA recommendation would not probably not be covered). Drugs that received the lowest SMR and ASMR ratings would be made available to patients, but with a lower level of coverage and at a price set equal to or below already existing alternatives. In February 2014, HAS has implemented the requirement for an economic evaluation to inform drug pricing negotiations for those drugs that received an ASMR I, II or III, and a significant impact on healthcare resources. The SMR and ASMR ratings drive the coverage rate and pricing scheme, while the economic evaluation enlightens the price negotiations in the pricing scheme applicable.

It can be concluded that in the study countries, the use of clinical and cost-effectiveness evidence can be considered as enlightening the decision process, rather than instrumental. Their use differ across countries, depending on the criteria accounted for, the existence of a threshold, and the consequence of the HTA recommendation. This may not be the case in other countries, where the HTA decisions are instrumental, and may be based solely on the ICER or on the HTA decisions in other countries. This is partially the case, for example, in Bulgaria, where their recent regulation implemented in December 2015 on the conditions and procedures for conducting HTA states that the "HTA procedure shall be terminated in the cases when a negative HTA assessment for the evaluated medicinal product is available by a state institution of the UK, France or Germany" (Bulgarian Ministry of Health, Article 17 of their Regulation No. 9 of the 01.12. 2015 on the conditions and procedures for conducting health technology assessment). A negative recommendation in one of the referenced countries would be used as instrumental in issuing a negative decision in Bulgaria.

11.2.3. Recommendations

The recommendations that emerge from this research relate to (1) the added value of applying this framework (e.g. multiplicity of criteria, breakdown of the decision-making process), (2) cross-country comparisons of the application of HTA across

settings and implications for HTA methodological approaches, (3) value judgments, and (4) dealing with rarity.

- This research sheds light on the multiplicity of criteria accounted for throughout the decision processes.

  Beyond clinical benefit, ICER and elicited societal preferences, a number of additional criteria that influenced these decision processes have been identified, namely the influence from other sources of evidence (e.g. registry data) or stakeholder input, or the value judgments made (e.g. unmet need, severity, innovativeness). These were generally accounted for to address uncertainty. A first recommendation would be to ensure that these additional criteria are accounted for in a more consistent and explicit manner in future decisions.

  The use of additional criteria also highlights current limitations in appraising imperfect or incomplete evidence (e.g. from phase III trials considered the gold standard, but generating evidence from controlled environments), and the different approaches, beyond sensitivity analysis, to deal with these. This often relates to uncertainty around the real world use of these treatments (e.g. population generalizability, clinical practice, trial length capturing treatment effects). A recommendation is for these assessments to capture more comprehensively the use of these treatments in real world settings, and to do so, consider additional sources of evidence in a more systematic manner, such as stakeholder expertise or observations studies.

  Accounting for the criteria used beyond standard HTA methods and understanding their weight in the decisions can help decision-makers apply similar approaches when faced with similar scenarios in the future. This will contribute to a more consistent use of these criteria and improved accountability for reasonableness. This can also feed into other research.

  The added value of this research is the breakdown of the decision-making process in an easily understandable and comparable manner by the structure given: the

evidence, its interpretation (including value judgments) and their influence on the final decision. This allows to differentiate cases when differences are legitimate and a consequence of contextual factors (e.g. comparators reflecting clinical practice, elicited societal preferences, costs, willingness to pay levels), from cases that are a consequence of the evidence and review of the evidence (e.g. different evidence, issues raised, ways of dealing with the same issues, consideration of other forms of evidence beyond primary trials (non-primary trials, stakeholder input, "other considerations"). The latter cases can be used to further the debate around the application of HTA, types of challenges being faced and how they are dealt with. When focusing on conducting or comparing HTA across several countries, this structure of the decision-making process may be useful (e.g. EUNetHTA).

The application of this framework can also be useful for different stakeholders aiming to understand how certain types of criteria influenced the process. For example, patients can retrospectively identify all cases where patient input was provided, and understand the type of input provided and how this influenced the decision.

- Identifying the reasons for cross-country differences and when these were a consequence of the application of HTA methods enabled to generate additional evidence about the application of HTA in different settings, the challenges faced and how they were dealt with. A number of policy implications were highlighted in the previous section, based on which the following recommendations are made:

  A recommendation would be to encourage the use of other forms of evidence to overcome some of the well-known limitations relating to RCTs (e.g. reflecting clinical practice), which are considered the gold standard of evidence. This relates to evidence about the use of the treatment in the real world (e.g. stakeholder input, observational studies) or that can help deal with uncertainty in addition to current methods (e.g. sensitivity analysis).

The parameters from the primary trials, which were in most cases RCTs, were mainly considered for the decisions. The influence of other evidence (non-primary and non-RCTs) on the decisions was quasi inexistent. More clarity and research is needed around the quality standards for these observational studies, and around cases when such types of evidence are considered acceptable and for what purposes. More clarity is also needed around how the relevant endpoints are selected and the weight quality of life or safety may have in the decision process, particularly when the latter is considered as a soft endpoint.

Better alignment of incentives across the drug development pipeline is needed when early marketing authorization is granted and the assessments rely on phase II trials, sometimes even non-comparative. This could be through the ability to implement patient access schemes (e.g. coverage with evidence development) allowing to share the risk of reimbursing the treatment with the manufacturer.

More clarity around the application of a societal perspective is needed (e.g. TLV), which can also feed into current developments at NICE or SMC pertaining to the special processes for ultra-orphan or highly specialized medicines.

- This research identified the social value judgments across countries and drugs.

Retrospectively identifying these can contribute towards better defining them for future cases. For example, disease severity is a prioritization criteria for TLV but no definition of attributes for severity is provided. Identifying the different ways that severity was recognized across cases can help defining this attribute. The same can be done for unmet need, innovation, etc. These are criteria that are often discussed and are of interest to account for explicitly in innovative models such as MCDA, or within the new value-based pricing system in England.

Identifying these retrospectively and recording them could help improve their accountability for reasonableness in order to account for these in future cases.

- When dealing with orphan drugs, a number of issues relating to rarity and evidence generation were identified. Policymakers should be aware of these limitations, how they were dealt with in order to generate a set of criteria when greater uncertainty may be accepted.

  Decision-makers should consider whether the best available evidence is of sufficient level considering prevalence and the number of recruitable conditions, and innovative trial designs (e.g. N of 1 trials) that may help dealing with small sample sizes.

  Decision-makers should understand the challenges around generating evidence and account for these during the appraisals, including how to deal with the following: subgroup analysis, surrogate outcomes, short trial duration, large confidence intervals due to small sample sizes, unknown treatment pathway.

  Given the higher level of uncertainty characterizing orphan drugs, a continuous collection of data about their use in real-world settings and future re-assessments may contribute towards addressing uncertainty over time. This could be done with registries, post-marketing surveillance for efficacy and safety, PROs, and so forth. This further also emphasizes the need to empower patients at each stage of the drug development pipeline in order to produce evidence that responds to their needs and preferences, and measures what is most relevant. This may also be a way to identify and account for the wider impacts of the disease on patients and their carers (e.g. particularly when diseases affect children, and the chronicity of these conditions).

This thesis contributed to highlighting that the current system is not suitable enough to tackle these more rare conditions, as highlighted not only be the magnitude of and conflicting differences in the HTA recommendations made, but also be the contrasts seen at various levels in dealing with these particular cases. This is all the more important in a changing pharmaceutical environment that is shifting towards more niche and targeted therapies ("the right patient with the right drug at the right dose at the right time"), where such issues will soon become a daily reality. Policymakers

should be aware of this and work towards differentiating these processes for them to account for the particularities of orphan drugs, including agreeing on ways that will modulate the assessments towards greater acceptability of uncertainty (e.g. coverage with evidence development, continuous assessment and re-assessment).

11.3. Limitations

Each Chapter has outlined the limitations relating to the empirical work carried out, this section describes the overarching limitations throughout the thesis.

Data sources

The data sources used to understand the decision-making processes were the HTA reports issued by each agency summarising the decisions. The reports also do report the same level of detail when comparing countries. We assumed that these reports reflected the key determinants, defined as the main reasons for the final decisions, driving these decisions given that countries are required to be transparent about these (European Union, 1988). Nevertheless, all aspects of the decision, such as the context within which the decisions were made, were likely not captured within these reports, nor during the interviews given that these decisions were made a few years back and certain specific issues discussed during the deliberative process of HTA are probably not recorded in their memories. These contextual and other considerations, however, were not within the scope of this study but should be acknowledged as possibly having influenced these decisions. One of the HTA bodies also mentioned that when a positive decision is made, there may also be less reason for report in detail everything that was accounted for. This was indeed a limitation, even though the most relevant information for the purpose of this thesis were the main reasons for the final decision, which were usually reported in all decision reports.

A further limitation relating to the data collected when applying the methodological framework came from secondary sources. It would have been preferable to have full information about the submissions (e.g. manufacturer submission), but this was not feasible in the current scheme. The objective of the interview was to validate the

findings that arose from the interpretation of the researcher and obtain additional insights about the different approaches to valuing orphan drugs, which may have not been captured by even the most complete documentation.

Similarly, Chapter 9 explored and analysed the scientific and social value judgments made during the decision processes. Given that this is generally accounted for during the deliberative processes, it would have been preferable to record these discussions by taking part of the Committee meetings and interview the Committee Members about their individual judgments. However, these meetings are not open to the public and I relied on the assumption that the decision reports should be transparent and capture these judgments made.

The interviews were driven by scenarios derived from findings of the previous chapters. It would have been preferable to go through *each case study* in detail with the HTA bodies to ensure that the interpretation was correct and illustrate what happened in the other countries. This was not possible because of the limited availabilities of the interviewees and the fact that they were not the decision-makers of the study drugs. Therefore, the interviews focused on a number of scenarios identified, where contrasts were seen across countries.

Sampling and sample size

Another limitation relates to sampling issues arising from differences among the four agencies in the way they select topics for their assessments (e.g. highest need for guidance at NICE, outpatient drugs at TLV). Despite these differences, a suitable sample has been identified. The sample size analysed in Chapters 8 and 9 was relatively small consisting in ten orphan drugs. It would have been ideally preferable to have a greater sample size to conduct multivariate or regression analyses. However, given the depth of what was captured during these decisions, the focus of the analysis was to capture the nuance of the decision without being constrained to a limited number of pre-defined variables. On this basis, the decision was made to prioritise the qualitative strand of this mixed methods research project.

The sample size for TLV remained smaller (5 compounds), but was nevertheless included in order to explore the Swedish context and understand how the societal perspective adopted in the assessments influenced the decisions compared to the other countries, which adopt a health service perspective.

The study countries included in Chapters 7-10 (England, Scotland, Sweden and France) were a convenience sample with known and well-established HTA bodies that fulfil a set of pre-defined criteria that were considered relevant to ensure their comparability. Other countries, e.g. Germany, the Netherlands, would have fulfilled the criteria but were not selected either because of language barriers or lack of available data, or because of a lack of capacity. Including additional countries would have required more time and resources, and therefore, given the quantity of data collected with the current sample, I did not include any additional countries. Further, once the methodological framework was tested for its feasibility, it was applied to additional drugs and countries by colleagues from the Advance-HTA project under my supervision. The analysis is currently on-going.

A final limitation relates to the relatively small sample size, which does not allow for multivariate analysis. However, this research put greater focus on the qualitative component in order to get an in-depth understanding of the subtleties in each country, which would not have been possible if the analysis focused on more drugs. Now that we have a better understanding of the reasons for differences and trivialities in each country, it would be meaningful to now look at the drivers of these differences across a larger sample of drugs and therapy areas using multivariate regression analysis, and can be extended to other types of drugs to assess how different agencies assess different drug and disease characteristics. This is discussed in the next section.

Data analysis and generalisability of results

One of the main limitations of this thesis was around its external validity. The results of this thesis are not generalizable to all orphan drugs, but to the sample analysed. Despite this, results provide a more structured and fuller understanding of the potential reasons for differences. For results to be generalisable, the analysis should be

conducted over a larger sample of drugs. However, in order to maintain the depth and breadth of the analysis building on the methodological framework used in this study, it was highly recommended to begin by prioritising the qualitative strand to ensure that the depth of the processes is captured and comparable across settings. Further research could look at the drivers of these differences across a larger sample of drugs and therapy areas using multivariate regression analysis for a greater generalisation of the results, by extending it to other types of drugs to assess how different agencies assess different drug and disease characteristics.

Further, given the prioritization of the qualitative strand and the fact that the quantitative analyses conducted were more of descriptive nature to help with the interpretation of the results, one of the limitations in line with the lack of generalizability of the results is that it was not possible to interpret the relative contribution of each criteria identified in the decision, given that no adjustments were made with the other criteria identified. This would be possible to do over a larger sample of drugs by conducting a regression or multivariate analysis (Cerri et al., 2015).

The transferability of the methodological framework developed to other countries and therapy areas is limited to those cases where similar decision-making criteria are accounted for, from HTA entities that are arm's length, responsible for issuing coverage recommendations, and have a transparent process where sufficient detail about the appraisal process and reasons for the final decision are recorded in their decision reports.

11.4. Future Research Agenda

As more and more countries continue to implement HTA as drivers for drug coverage decisions and constantly aim to adapt and improve HTA methodologies, it is important to continue to learn from each other's experiences. Initiatives such as EUnetHTA are one way of doing this through collaborations, where the assessment of relative effectiveness are jointly conducted by several Member States on behalf of all participating Member States, on the basis of the common methodologies and

guidelines developed. Results from this common assessment are then used by each agency to build the cost-effectiveness model and make the final decision at jurisdiction-level. They also provide a platform where HTA bodies can exchange and learn from each other when questions or issues arise.

This initiative is currently being piloted, and individual HTA bodies are still responsible to conduct the full assessments on their territory. Differences are therefore inevitable, and it is even more important to understand the reasons for these given the important uptake of HTA in Europe and around the world. This was the objective of this PhD. Chapter 2 set the scene by emphasising the need to understand the reasons for differences across countries, partially reflecting HTA methodological weaknesses. Chapter 3 developed and tested a tool enabling to systematically compare these decision-making processes across countries and drugs in order to identify the reasons for different HTA outcomes. Chapter 4 successfully applied this tool to a sample of 35 orphan drug and indication pairs, and generated meaningful results, while Chapter 5 concentrated on one of the components ("other considerations") to further understand how they are accounted for during these processes. Finally, HTA body representatives provided their input in Chapter 6 about the meaningfulness of these findings and the implications for HTA.

These five Chapters focus on a number of issues related to orphan drugs both in-depth and across all study drugs. Results helped better understand trends and preferences across the study drugs, while at the same time scrutinizing the implications of some of these findings, as done in Chapter 5 with the "other considerations". Greater depth and breadth to these results can be added by applying the framework to additional drugs, therapy areas and countries. This is feasible given the flexible and iterative approach adopted when developing the framework, and is already being undertaken as part of the European project funding this thesis (Advance-HTA).

Subsequent indications being analysed are cancer and central nervous system (CNS) treatments. Reasons for selecting these three therapy areas are: cancer and orphan drugs are therapeutically more innovative and beneficial than CNS treatments, according to their ASMR ratings (Nicod et al., 2012). Cancer and rare diseases have

similar characteristics (e.g. disease severity) also given that many rare diseases are rare cancers. However, evidence suggests that different levels of acceptance rates for coverage decisions apply (Dupont et al., 2011). These two samples will be interesting to compare and examine on what basis orphan drugs receive higher acceptance levels than cancer (and CNS) drugs. CNS diseases have different characteristics than cancer and rare diseases, where a strong need for new treatment remains because of patient compliance or tolerance, despite many already existing. These three therapy areas will provide an appropriate basis for comparing the manner in which different agencies appraise different therapy areas with a special focus on disease and treatment characteristics. The analysis can further be broadened to additional therapies and countries. Other interesting comparisons would be between orphan and ultra-orphan drugs, cancer orphan and non-orphan, orphan and non-orphan cancer, and so forth.

This thesis provides a good basis on how to analyse the data and disseminate the findings in a meaningful manner. Once a larger sample size is available, explanatory and confirmatory factor analysis will be conducted to quantitatively understand the factors most contributing to these decisions. The latent variable will be the unobservable variable “decision-making”, and explanatory variables the different criteria identified through the methodological framework to the three therapy areas. Expected results are to measure the contributions of each of these criteria to the decision-making process through the factor loadings, including those that have a greater influence over others, and whether some of the criteria are associated. The residual correlation matrix (observed minus fitter variables) will then allow to assess whether one variable or another does not contribute to explaining the latent variable, and as such will be excluded from the analysis. The goodness of fit test will be conducted using a likelihood ratio test. Finally, the analysis will also assess how the contributions of the variables to decision-making differ depending on the therapy area being appraised.

The methodological framework can also be further developed and tailored for the needs of different stakeholders, such as HTA bodies, patient groups, clinical experts or manufacturers. For example, HTA bodies may use it to better understand how other HTA bodies deal with uncertainties for a sample of drugs, using only a specific

component of the framework. Patient or clinical experts could also retrospectively identify the type of information provided by experts and their influence on the decisions to tailor future inputs or advocate for a greater formalisation and consistency of their contributions. Manufacturers could use this framework to better understand the expectations of HTA bodies in the different countries on the basis of previous assessments, which could help them tailor their submissions in the different countries.

Another potential application of this framework is within the multiple MCDA models currently being developed (Angelis & Kanavos, 2014; Goetghebeur, Wagner, Khoury, Levitt, Erickson, & Rindress, 2012). MCDA models are being developed and tested to complement HTA processes in their capacity to account for a broader range of criteria. Criteria are usually accounted for by a range of stakeholders and their preferences in terms of the different weights that should be allocated to the different criteria. Applying the framework to identifying, in a retrospective manner, the criteria accounted for in previous decisions can be useful to add on an additional dimension around consistency, in accounting for what has been done in previous cases.

Building on another study that also investigated the reasons for cross-country differences in HTA recommendations using a quantitative approach (Maynou-Puljoras & Cairns, 2016), further research could be to explore the interface between their results and those from this thesis (more qualitative by nature). This is currently on-going.

The findings also have implications for new programmes dealing with (ultra-)rare conditions. A recent paper building on the results of this thesis is under review and aims to identify the challenges commonly encountered with orphan drugs, and how new programmes specifically for orphan and/or ultra-orphan drugs were established and whether they were done so to explicitly deal with the challenges highlighted in this thesis. On this basis, a conceptual framework summarising these challenges was developed and examined for each new HTA programme (Nicod, Annemans, Bucsics, Lee, Upadhyaya and Facey, 2016)

Findings may also be used for educational purposes to different stakeholders. For example, HTA was recently introduced in Bulgaria and is influenced by HTA recommendations from other countries. Indeed, their legislation state that “HTA shall be terminated in the case when a negative HTA assessment for the evaluated medicinal product is available by a state institution of the UK, France or Germany (Bulgarian Ministry of Health, Article 17 of their Regulation No. 9 of the 01.12.2015 on the conditions and procedures for conducting health technology assessment). Therefore, it is important for them to understand the processes in these different countries, the reasons for issuing different recommendations as well as the implications of these decisions.

References

Altman, D. G. (1991). *Practical Statistics for Medical Research*: London England: Chapman and Hall.

Angelis, A., & Kanavos, P. (2014). Applying Multiple Criteria Decision Analysis in the context of health technology assessement: An empirical case study. *Value in Health, 17*(7), A552-A552.

Angelis, A., & Kanavos, P. (2016). Value-based assessment of new medical technologies: towards a robust methodological framework for the application of multiple criteria decision analysis in the context of health technology assessment. *PharmacoEconomics, 34(5),* 1-12.

Bach, P. B. (2009). Limits on Medicare's ability to control rising spending on cancer drugs. *New England Journal of Medicine, 360*(6), 626-633. doi: 10.1056/NEJMhpr0807774

Baker, R., Bateman, I., Donaldson, C., Jones-Lee, M., Lancsar, E., Loomes, G., Mason, H., Odejar, M., Pinto Prades, J.L., Robinson, A., Ryan, M., Shackley, P., Smith, R., Sugden, R., Wildman, J., & SVQ Research Team. (2010). Weighting and valuing quality-adjusted life-years using stated preference methods: preliminary results from the Social Value of a QALY Project. *Health Technology Assessment, 14*(27), 1-162. doi: 10.3310/Hta14270

Banta, D. (2003). The development of health technology assessment. *Health Policy, 63*(2), 121-132. doi: 10.1016/S0168-8510(02)00059-3

Banta, D., & Jonsson, E. (2009). History of HTA: Introduction. *International Journal for Technology Assessment in Health Care, 25 Suppl 1*, 1-6. doi: 10.1017/S0266462309090321

Bartholomew, D., Steele, F., Moustaki, I., & Galbraith, J. (2008a). *Analysis of Multivariate Social Sciences Data* (2nd edition ed.). Florida: Taylor & Francis.

Bartholomew, D. J., Steele, F., Moustaki, I., & Galbraith, J. I. (2008b). "Correspondence analysis". In D. J. Bartholomew, F. Steele, I. Moustaki & J. I. Galbraith, eds. *Analysis of Multivariate Social Science Data* (2nd edition ed.) (pp. 83-116). Florida: Taylor & Francis.

Bell, S. A., & Tudur Smith, C. (2014). A comparison of interventional clinical trials in rare versus non-rare diseases: an analysis of ClinicalTrials.gov. *Orphanet Journal of Rare Diseases, 9*, 170. doi: 10.1186/s13023-014-0170-0

Berger, M. L., Martin, B. C., Husereau, D., Worley, K., Allen, J. D., Yang, W. N., Quon, N.C., Mullins, C.D., Kahler, K.H., & Crown, W. (2014). A questionnaire to assess the relevance and credibility of observational studies to inform health care decision making: An ISPOR-AMCP-NPC Good Practice Task Force report. *Value in Health, 17*(2), 143-156. doi: 10.1016/j.jval.2013.12.011

Bilton, D., Robinson, P., Cooper, P., Gallagher, C., Kolbe, J., Fox, H., Jaques, A., Charlton, B., CF301 Study Investigators. (2011). Inhaled dry powder mannitol in cystic fibrosis: an efficacy and safety study. *European Respiratory Journal, 38*(5), 1071-1080. doi: 10.1183/09031936.00187510

Biomarkers Definitions Working Group. (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics, 69*(3), 89-95. doi: 10.1067/mcp.2001.113989

Black, N. (1996). Why we need observational studies to evaluate the effectiveness of health care. *British Medical Journal, 312*(7040), 1215-1218.

Blankart, C. R., Stargardt, T., & Schreyogg, J. (2011). Availability of and access to orphan drugs: an international comparison of pharmaceutical treatments for pulmonary arterial hypertension, Fabry disease, hereditary angioedema and chronic myeloid leukaemia. *Pharmacoeconomics, 29*(1), 63-82. doi: 10.2165/11539190-000000000-00000

Boggild, M., Palace, J., Barton, P., Ben-Shlomo, Y., Bregenzer, T., Dobson, C., & Gray, R. (2009). Multiple sclerosis risk sharing scheme: two year results of clinical cohort study with historical comparator. *British Medical Journal, 339*. doi: 10.1136/Bmj.B4677

Booth, C. M., & Eisenhauer, E. A. (2012). Progression-free survival: meaningful or simply measurable? *Journal of Clinical Oncology, 30*(10), 1030-1033. doi: 10.1200/JCO.2011.38.7571

Boudes, P. F. (2013). Clinical studies in lysosomal storage diseases: Past, present, and future. *Rare Diseases, 1*, e26690. doi: 10.4161/rdis.26690

Bouvy, J., & Vogler, S. (2013). Update on 2004 background paper, BP 8.3 pricing and reimbursement policies: priority medicines for Europe and the world "a public health approach to innovation". http://www.who.int/medicines/areas/priority_medicines/BP8_3_pricing.pdf Viewed October 2015.

Boyatzis, R. E. (1998). "Developing themes and codes". In R. Boyatzis (Ed.), *Transforming Qualitative Information: Thematic Analysis and Code Development* (pp. 29-53). Thousand Oaks, CA: Sage Publications.

Brazier, J. (2008). Valuing health states for use in cost-effectiveness analysis. *Pharmacoeconomics, 26*(9), 769-779. doi: 10.2165/00019053-200826090-00007

Brousselle, A., & Lessard, C. (2011). Economic evaluation to inform health care decision-making: Promise, pitfalls and a proposal for an alternative path. *Social Science & Medicine, 72*(6), 832-839. doi: 10.1016/j.socscimed.2011.01.008

Brouwer, W. B. F., Culyer, A. J., van Exel, N. J. A., & Rutten, F. F. H. (2008). Welfarism vs. extra-welfarism. *Journal of Health Economics, 27*(2), 325-338. doi: 10.1016/j.jhealeco.2007.07.003

Bryman, A. (2004). *Social Research Methods* (2nd edition ed.). Oxford: Oxford University Press.

Bulgarian Ministry of Health (2015). Regulation No. 9 of 01.12. 2015 on the conditions and procedures for conducting health technology assessment, Article 17.

Busse, R., Orvain, J., Velasco, M., Perleth, M., Drummond, M., Gurtner, F., Jorgensen, T., Jovell, A., Malone, J., Ruther, A., & Wild, C. (2002). Best practice in undertaking and reporting health technology assessments - Working group 4 report. *International Journal of Technology Assessment in Health Care, 18*(2), 361-422.

Cairns, J. (2006). Providing guidance to the NHS: The Scottish Medicines Consortium and the National Institute for Clinical Excellence compared. *Health Policy, 76*(2), 134-143. doi: 10.1016/j.healthpol.2005.05.006

Caputo, A. (2014). Exploring quality of life in Italian patients with rare disease: a computer-aided content analysis of illness stories. *Psychology Health & Medicine, 19*(2), 211-221. doi: 10.1080/13548506.2013.793372

Caro, J. J. (2009). Pursuing efficiency: A dead end for HTA? *Value in Health, 12*, S49-S49. doi: 10.1111/j.1524-4733.2009.00559.x

Carroll, C., Kaltenthaler, E., FitzGerald, P., Boland, A., & Dickson, R. (2011). A thematic analysis of the strengths and weaknesses of manufacturers' submissions to the NICE Single Technology Assessment (STA) process. *Health Policy, 102*(2-3), 136-144. doi: 10.1016/j.healthpol.2011.06.002

Cerri, K., Knapp, M., Fernandez, J.L. (2014). Untangling the complexity of funding recommendations: A comparative analysis of health technology assessment outcomes in four European countries. *Pharmaceutical Medicine,* 29:341-59.

Cerri, K., Knapp, M., Fernandez, J.L. (2015).Decision making by NICE: examining the influences of evidence, process and context. *Health Economics Policy and Law*, 9(2):119-41.

CDC. Centers for Disease Control and Prevention (2016). Health related quality of life Retrieved from: http://www.cdc.gov/hrqol/concept.htm Viewed September 2016.

CDR. (2006 3rd edition). Guidelines for the economic evaluation of health technology assessment. Canadian Agency for Drugs and Technologies in Health (CADTH). https://www.cadth.ca/media/pdf/186_EconomicGuidelines_e.pdf. Viewed November 2007.

CDR. (2008a). Common Drug Review - Clinical Review Report Template. Canadian Agency for Drugs and Technologies in Health (CADTH) (Ed.) https://www.cadth.ca/media/cdr/process/CDR%20-%20Sample%20Template%20for%20Clinical%20Report.pdf. Viewed November 2015.

CDR. (2008b). Notice of CEDAC final recommendation- Paliperidone (Invega - Janssen Ortho Inc.). http://www.cadth.ca/en/products/cdr/search?q=paliperidone. Viewed December 2012.

Chalkidou, K., Tunis, S., Lopert, R., Rochaix, L., Sawicki, P. T., Nasser, M., & Xerri, B. (2009). Comparative effectiveness research and evidence-based health policy: experience from four countries. *Milbank Quarterly, 87*(2), 339-367. doi: 10.1111/j.1468-0009.2009.00560.x

Ciani, O., Buyse, M., Garside, R., Pavey, T., Stein, K., Sterne, J. A. C., & Taylor, R. S. (2013). Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. *British Medical Journal, 346*. doi: 10.1136/bmj.f457

Clarke, J. P. (2003). "How to peer review a qualitative manuscript". In F. Godlee & T. Jefferson (Eds.), *Peer Review in Health Science* (2nd ed., pp. 219-235). London: BMJ books.

Clarke, J. T. R. (2006). Is the current approach to reviewing new drugs condemning the victims of rare diseases to death? A call for a national orphan drug review policy. *Canadian Medical Association Journal, 174*(2), 189-190. doi: 10.1503/Cmaj.050706

Claxton, K. (2008). Exploring uncertainty in cost-effectiveness analysis. *Pharmacoeconomics, 26*(9), 781-798. doi: 10.2165/00019053-200826090-00008

Claxton, K., Briggs, A., Buxton, M. J., Culyer, A. J., McCabe, C., Walker, S., & Sculpher, M. J. (2008). Value based pricing for NHS drugs: an opportunity not to be missed? *British Medical Journal, 336*(7638), 251-254. doi: 10.1136/bmj.39434.500185.25

Clement, F. M., Harris, A., Li, J. J., Yong, K., Lee, K. M., & Manns, B. J. (2009). Using effectiveness and cost-effectiveness to make drug coverage decisions: A comparison of Britain, Australia, and Canada. *Journal of the American Medical Association, 302*(13), 1437-1443.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46.

Commission of the European Communities. (2008). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on Rare Disease's: Europe's challenges. http://ec.europa.eu/health/archive/ph_threats/non_com/docs/rare_com_en.pdf. Viewed November 2015.

Congressional Budget Office. (October 2006). Research and development in the pharmaceutical industry. *A CBO Report*: Congressional Budget Office, Congress of the United States. https://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/76xx/doc7615/10-02-drugr-d.pdf. Viewed October 2015.

Creswell, J. W. (2003). "The selection of a research approach". In T. Oaks (Ed.), *Research Design: Qualitative, Quantitative and Mixed Methods Approaches* (pp. 1-24). CA: Sage.

Creswell, J. W., & Plano Clark, V. L. (2011a). "Analyzing and interpreting data in mixed methods research". In J. W. Creswell & V. L. Plano Clark (Eds.), *Designing and Conducting Mixed Methods Research* (pp. 203-205). Los Angeles: SAGE Publications.

Creswell, J. W., & Plano Clark, V. L. (2011b). "Collecting data in mixed methods research". In J. W. Creswell & V. L. Plano Clark (Eds.), *Designing and Conducting Mixed Methods Research* (2nd ed., pp. 110-127). Los Angeles: SAGE Publications.

Creswell, J. W., & Plano Clark, V. L. (2011c). *Designing and conducting mixed methods research* (2nd ed.). Los Angeles: SAGE Publications.

Creswell, J. W., & Tashakkori, A. (2007). Developing publishable mixed methods manuscripts. *Journal of Mixed Methods Research, 1*(2), 107-111. doi: 10.1177/1558689806298644

Daniels, N. (2000). Accountability for reasonableness - Establishing a fair process for priority setting is easier than agreeing on principles. *British Medical Journal, 321*(7272), 1300-1301. doi: 10.1136/bmj.321.7272.1300

Dakin, H., Devlin, N., Odeyemi, I. (2006). "Yes", "No" or "Yes, but"? Multinomial modelling of NICE decision-making. *Health Policy*, 77:352-67.

Dakin, H., Devlin, N., Feng, Y., Rice, N., O'Neill, P., & Parkin, D.. (2015). The influence of cost-effectiveness and other factors on NICE decisions. *Health Economics, 24: 1256-71*.

Daniels, N., & Sabin, J. E. (2008). Accountability for reasonableness - an update. *British Medical Journal, 337*(7675). doi: 10.1136/Bmj.A1850

Denis, A., Mergaert, L., Fostier, C., Cleemput, I., & Simoens, S. (2010a). A comparative study of European rare disease and orphan drug markets. *Health Policy, 97*(2-3), 173-179. doi: 10.1016/j.healthpol.2010.05.017

Denis, A., Mergaert, L., Fostier, C., Cleemput, I., & Simoens, S. (2010b). Issues surrounding orphan disease and orphan drug policies in Europe. *Applied Health Economics and Health Policy, 8*(5), 343-350. doi: 10.2165/11536990-000000000-00000

Department of Health, Medicines, Pharmacy & Industry Group. (2010). *A new value-based pricing approach to the pricing of branded medicines: a consultation*. http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_122793.pdf. Viewed March 2011.

Desser, A. S., Gyrd-Hansen, D., Olsen, J. A., Grepperud, S., & Kristiansen, I. S. (2010). Societal views on orphan drugs: cross sectional survey of Norwegians aged 40 to 67. *British Medical Journal, 341*. doi: 10.1136/Bmj.C4715

Devlin, N., Parker, D. (2004). Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Economics,* 13:437-52.

DG Research and Innovation. European Commission. (2011). *Orientation paper: proposed priorities for innovative health research 2012 - FP7-HEALTH-2012-INNOVATION*. http://ec.europa.eu/research/participants/portal/page/cooperation?callIdentifier=FP7-HEALTH-2012-INNOVATION-1. Viewed January 2012.

Dickenson, W. (2010). "Visual displays for mixed methods findings". In A. Tashakkori & C. Teddlie (Eds.), *Sage Handbook of Mixed Methods in Social & Behavioral Research* (2nd ed., pp. 469-504). Los Angeles: SAGE Publications.

DiMasi, J., & Grabowski, H. (2007). The cost of biopharmaceutical R&D: Is biotech different? *Managerial and Decision Economics, 28*, 469-479.

DiMasi, J. A., Hansen, R. W., & Grabowski, H. G. (2003). The price of innovation: new estimates of drug development costs. *Journal for Health Economics, 22*(2), 151-185. doi: 10.1016/S0167-6296(02)00126-1

Directorate-General for Economic and Financial Affairs. (2012). The quality of public expenditures in the EU, European Economy. *Occasional Papers 125*.

Dolan, P., & Shaw, R. (2004). A note on a discussion group study of public preferences regarding priorities in the allocation of donor kidneys. *Health Policy, 68*(1), 31-36. doi: 10.1016/j.healthpol.2003.07.006

Dolan, P., Shaw, R., Tsuchiya, A., & Williams, A. (2005). QALY maximisation and people's preferences: a methodological review of the literature. *Health Economics, 14*(2), 197-208. doi: 10.1002/Hec.924

Doward, L. C., Gnanasakthy, A., & Baker, M. G. (2010). Patient reported outcomes: looking beyond the label claim. *Health and Quality of Life Outcomes, 8*. doi: Doi 10.1186/1477-7525-8-89

Dranitsaris, G., Truter, I., Lubbe, M., Amir, E., & Evans, W. (2011). Advances in cancer therapeutics and patient access to new drugs. *Pharmacoeconomics, 29*(3), 213-224.

Drummond, M. (2009). More haste, less speed? The emerging practice of HTA in the United Kingdom. *EuroObserver, 11*(1), 9-10.

Drummond, M., de Pouvourville, G., Jones, E., Haig, J., Saba, G., & Cawston, H. (2014a). A comparative analysis of two contrasting european approaches for rewarding the value added by drugs for cancer: England versus France. *Pharmacoeconomics, 32*(5), 509-520. doi: 10.1007/s40273-014-0144-z.

Drummond, M., Evans, B., LeLorier, J., Karakiewicz, P., Martin, D., Tugwell, P., & MacLeod, S. (2009a). Evidence and values: requirements for public reimbursement of drugs for rare diseases--a case study in oncology. *Canadian Journal of Clinical Pharmacology, 16*(2), e273-281; discussion e282-274.

Drummond, M., Schwartz, J. S., Jonsson, B., Luce, B. R., Neumann, P. J., Siebert, U., & Sullivan, S. D. (2008). Key principles for the improved conduct of health technology assessments for resource allocation decisions. *International Journal of Health Technology Assessment, 24*(3), 244-258.

Drummond, M., & Sorenson, C. (2009b). Nasty or Nice? A perspective on the use of Health Technology Assessment in the United Kingdom. *Value in Health, 12*, S8-S13. doi: 10.1111/j.1524-4733.2009.00552.x

Drummond, M., & Towse, A. (2014b). Orphan drug policies: a suitable case for treatment. *European Journal for Health Economics, 15*, 335-340.

Drummond, M. B. (2008). What are the HTA processes in the UK. *What is...? Series*, 1-9.

Drummond, M. F., Wilson, D. A., Kanavos, P., Ubel, P., & Rovira, J. (2007). Assessing the economic challenges posed by orphan drugs. *International Journal of Technology Assessment in Health Care, 23*(1), 36-42.

Dupont, A. G., & Van Wilder, P. B. (2011). Access to orphan drugs despite poor quality of clinical evidence. *British Journal of Clinical Pharmacology, 71*(4), 488-496. doi: 10.1111/j.1365-2125.2010.03877.x

Earnshaw, J., & Lewis, G. (2008). NICE Guide to the Methods of Technology Appraisal - Pharmaceutical industry perspective. *Pharmacoeconomics, 26*(9), 725-727. doi: 10.2165/00019053-200826090-00002

Eddy, D. (2009). Health technology assessment and evidence-based medicine: What are we talking about? *Value in Health, 12*, S6-S7. doi: 10.1111/j.1524-4733.2009.00551.x

Eichler, H. G., Oye, K., Baird, L. G., Abadie, E., Brown, J., Drum, C. L., Ferguson, J., Garner, S., Honig, P., Hukkelhoven, M., Lim, J.C., Lim, R., Lumpkin, M.M., Neil, G., O'Rourke, B., Pezalla, E., Shoda, D., Seyfert-Margolis, V., Sigal, E.V., Sobotka, J., Tan, D., Unger, T.F., Hirsch, G. (2012). Adaptive Licensing: Taking the next step in the evolution of drug approval. *Clinical Pharmacology & Therapeutics, 91*(3), 426-437. doi: 10.1038/clpt.2011.345

EMA. (2013). Human medicines, rare disease (orphan) designations. http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/landing/orphan_search.jsp&mid=WC0b01ac058001d12b. Viewed November 2013.

EPIRARE. (2013). European Platform for Rare Disease Registries. from http://www.epirare.eu/ Viewed December 2013.

EUCERD. (2013). Report on the state of the art of rare disease activities in Europe In European Union Committee of Experts on Rare Diseases. http://www.eucerd.eu/upload/file/Reports/2013ReportStateofArtRDActivities.pdf. Viewed March 2015.

EUnetHTA. (2013). The EUnetHTA Core Model, Joint Action 2, Work Package 8 (Vol. version 2.0). http://meka.thl.fi/htacore/model/HTACoreModel2.0.pdf. Viewed December 2015.

EUnetHTA. (2015). EUnetHTA: Mission, vision and values. http://www.eunethta.eu/about-us/mission-vision-values. Viewed November 2014.

European Commission. (1989). *Transparency Directive 89/105/EEC*. http://ec.europa.eu/enterprise/sectors/healthcare/competitiveness/pricing-reimbursement/transparency/index_en.htm#h2-2. Viewed January 2013.

European Commission. (2011). *Directive 2011/24/EU on patients' rights in cross-border healthcare*. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:088:0045:0065:EN:PDF. Viewed March 2013.

European Commission. (2012). *The 2012 Ageing Report. Economic and budgetary projections for the 27 EU Member States (2010-2060).*

http://ec.europa.eu/economy_finance/publications/european_economy/2015/pdf/ee3_en.pdf. Viewed November 2014.

European Commission. (2013a). *Commission staff working document. Investing in Health. Towards social investiment for growth and cohesion – including implementing the European Social Fund 2014-2020*. http://ec.europa.eu/social/BlobServlet?docId=9771&langId=en. Viewed November 2014.

European Commission. (2013b). *Guiding principles for good practices implementing a pricing and reimbursement policy*. http://ec.europa.eu/enterprise/sectors/healthcare/files/docs/pricing_principles_en.pdf. Viewed November 2014.

European Commission. (2014). Pharmaceutical pricing and reimbursement information sheet. http://ec.europa.eu/health/technology_assessment/policy/index_en.htm. Viewed February 2014.

European Commission. (2015a). Health Technology Assessment - policy. http://ec.europa.eu/health/technology_assessment/policy/index_en.htm. Viewed November 2014.

European Commission. (2015b). Rare diseases - policy. http://ec.europa.eu/health/rare_diseases/policy/index_en.htm. Viewed November 2014.

European Medicines Agency. (2012). European public assessment reports - orphan medicines.

European Medicines Agency. (2014). Pilot project on adaptive licensing.

Council Directive of 21 December 1988 relating to the transparency of measures regulating the pricing of medicinal products for human use and their inclusion in the scope of national health insurance systems (1988).

Regulation EC No 141/2000 of the European Parliament and of the Council of December 16 1999 on orphan medicinal products (1999).

EVIDEM. (2015). Decision Criteria. Conceptual background, definitions, design & instructions (Vol. V3.0): EVIDEM Collaboration. https://www.evidem.org/docs/2015/EVIDEM-v3-0-Decision-criteria-conceptual-background-definitions-and-instructions-June-2015b.pdf. Viewed December 2015.

Faulkner, E., Matuszewski, K., & Niziol, C. (2009). ISPOR global health care systems road map. Sweden - Pharmaceutical. www.ispor.org/HTARoadMaps/Sweden.asp. Viewed March 2014.

Ferrario, A., & Kanavos, P. (2013). Managed entry agreements for pharmaceuticals: the European experience. Brussels: EMINET.

Fischer, K. E. (2012). A systematic review of coverage decision-making on health technologies-Evidence from the real world. *Health Policy, 107*(2-3), 218-230. doi: 10.1016/j.healthpol.2012.07.005

Folland, S., Goodman, A. C., & Stano, M. (2013). Demand and supply of health insurance. *The Economics of Health and Health Care* (7th ed., pp. 143-166). Upper Saddle River, N.J.: Pearson.

Friendly, M. (1991). Statistical graphics for multivariate data. *Sugi 16 - Proceedings of the Sixteenth Annual Sas Users Group International Conference*, 1157-1162.

Gagne, J. J., Thompson, L., O'Keefe, K., & Kesselheim, A. S. (2014). Innovative research methods for studying treatments for rare diseases: methodological review. *British Medical Journal, 349*. doi: 10.1136/Bmj.G6802

Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology, 13*. Doi 10.1186/1471-2288-13-117

Garau, M., & Mestre-Ferrandiz, J. (2009a). Access mechanisms for orphan drugs: a comparative study of selected European countries. Office for Health Economics. https://www.ohe.org/publications/access-mechanisms-orphan-drugs-comparative-study-selected-european-countries. Viewed November 2014.

Garau, M., Shah, K., Towse, A., Wang, Q., Drummond, M., & Mason, A. (2009b). Assessment and appraisal of oncology medicines: NICE's approach and international HTA experience. *Report for the Pharmaceutical Oncology Initiative Group (POI).* https://www.ohe.org/publications/assessment-and-appraisal-oncology-medicines-nices-approach-and-international-hta. Viewed November 2014.

Gauvin, F. P., Abelson, J., Giacomini, M., Eyles, J., & Lavis, J. N. (2010). "It all depends": Conceptualizing public involvement in the context of health technology assessment agencies. *Social Science & Medicine, 70*(10), 1518-1526. doi: 10.1016/j.socscimed.2010.01.036

Gerring, J. (2007). What Is a case study good for? Case study versus large-N cross-case analysis. *Case Study Research: Principles and Practices*, 37-63. doi: 10.2277/ 0521676568

Ghislandi, S., Apolone, G., Garattini, L., & Ghislandi, I. (2002). Is EQ-5D a valid measure of HRQoL in patients with movement disorders? A comparison with SF-36 and FIM questionnaires. *European Journal for Health Economics, 3*(2), 125-130. doi: 10.1007/s10198-002-0105-x

Goetghebeur, M. M., Wagner, M., Khoury, H., Levitt, R. J., Erickson, L. J., & Rindress, D. (2012). Bridging Health Technology Assessment (HTA) and efficient health care decision making with Multicriteria Decision Analysis (MCDA): Applying the EVIDEM framework to medicines appraisal. *Medical Decision Making, 32*(2), 376-388. doi: 10.1177/0272989X11416870

Haffner, M. E. (1998). Designing clinical trials to study rare disease treatment. *Drug Information Journal, 32*(4), 957-960.

Hailey, D. (2009). Development of the International Network of Agencies for Health Technology. *International Journal of Technology Assessment in Health Care, 25*(S1), 24-27.

HAS. (2012). *Choices in methods for economic evaluation: a methodological guide*. France: Haute Autorite de Sante.

HAS. (2015). Règlement intérieur de la Commission de la Transparence. France: Haute Autorite de Sante.

Hoffman, D. L., & Franke, G. R. (1986). Correspondence analysis - Graphical representation of categorical-data in marketing-research. *Journal of Marketing Research, 23*(3), 213-227. doi: 10.2307/3151480

Howie, L., Hirsch, B., Locklear, T., & Abernethy, A. P. (2014). Assessing the value of patient-generated data to comparative effectiveness research. *Health Affairs, 33*(7), 1220-1228. doi: 10.1377/hlthaff.2014.0225

HTAglossary.net. (2015). Evidence-based medicine. http://www.htai.org. Viewed November 2014.

HTAI. (2011). Health Technology Assessment International: international society for the promotion of health technology assessment. *Mission.* http://www.htai.org/index.php?id=427. Viewed November 2014.

HTAI Policy Forum Meeting. (2013). HTA and Value: Assessing value, making value-based decisions, and sustaining innovation. Background paper. Barcelona, Spain: HTAI.

Hughes, D. A., Tilson, L., & Drummond, M. (2009). Estimating drug costs in economic evaluations in Ireland and the UK: An analysis of practice and research recommendations. *Pharmacoeconomics, 27*(8), 635-643.

Hurley, J. (2000). "An overview of the normative economics of the health care sector". In A. Culyer & J. Newhouse (Eds.), *Handbook of Health Economics* (Vol. 1A, pp. 56-118). Amsterdam, The Netherlands: Elsevier.

Huskamp, H. A. (2006). Prices, profits, and innovation: examining criticisms of new psychotropic drugs' value. *Health Affairs (Millwood), 25*(3), 635-646.

ICD10. (2007). International statistical classification od diseases and related health problems, 10th revision (ICD 10). http://apps.who.int/classifications/apps/icd/icd10online/. Viewed November 2014.

IMF. (2010). Gross domestic product based on purchasing-power-parity (PPP) per capita GDP in current international dollar for Canada, Australia, United Kingdom, Sweden and France (2010). http://www.imf.org/external/pubs/ft/weo/2011/02/weodata/weorept.aspx?pr.x=37&pr.y=12&sy=2009&ey=2016&scsm=1&ssd=1&sort=country&ds=.&br=1

&c=193%2C156%2C132%2C144%2C112&s=PPPPC&grp=0&a=#download. Viewed November 2014.

ISPOR. (2003). *Health care costs, quality and outcomes: ISPOR Book of Terms*. Lawrenceville, NJ.

ISPOR. (2012). ISPOR Global Health Case Systems Road Map - Canada, pharmaceuticals. from http://www.ispor.org/htaroadmaps/CanadaPharm.asp. Viewed November 2012.

Joppi, R., Bertele', V., & Garattini, S. (2013). Orphan drugs, orphan diseases. The first decade of orphan drug legislation in the EU. *European Journal of Clinical Pharmacology, 69*(4), 1009-1024. doi: DOI 10.1007/s00228-012-1423-2

Joppi, R., Bertele, V., & Garattini, S. (2009). Orphan drug development is not taking off. *British Journal of Clinical Pharmacology, 67*(5), 494-502. doi: 10.1111/j.1365-2125.2009.03369.x

Kanavos, P., Costa-Font, J., & Seeley, E. (2008). Competition in off-patent drug markets: Issues, regulation and evidence. *Economic Policy* (55), 500-544.

Kanavos, P., Manning, J., Taylor, D., Schurer, W., & Checchi, K. (2010a). Implementing value-based pricing for pharmaceuticals in the UK. *2020 Health*.

Kanavos, P., & Nicod, E. (2012). What Is Wrong with Orphan Drug Policies? Suggestions for Ways Forward. *Value in Health, 15*(8), 1182-1184. doi: DOI 10.1016/j.jval.2012.08.2202

Kanavos, P., Nicod, E., Van den Aardweg, S., & Pomedli, S. (2010b). The impact of health technology assessment: an international comparison. *EuroObserver, 12*(4), 1-7.

Kanters, T. A., Hakkaart, L., Rutten-van Molken, M. P., & Redekop, W. K. (2015). Access to orphan drugs in western Europe: can more systematic policymaking really help to avoid different decisions about the same drug? *Expert Review Pharmacoeconomics and Outcomes Research, 15*(4), 557-559. doi: 10.1586/14737167.2015.1045882

Kaplan, W., Wirtz, V., Mantel-Teeuwisse, A., Stolk, P., Duthey, B., & Laing, R. (2013). Priority medicines for Europe and the world 2013 update. http://www.who.int/medicines/areas/priority_medicines/MasterDocJune28_FINAL_Web.pdf. Viewed February 2012.

Kasenda, B., Schandelmaier, S., Sun, X., von Elm, E., You, J., Blumle, A., . . . Briel, M. (2014). Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications (vol 349, g4539, 2014). *British Medical Journal, 349*. Doi 10.1136/Bmj.G4921

Kesselheim, A. S. (2011). An empirical review of major legislation affecting drug development: Past experiences, effects, and unintended consequences. *Milbank Quarterly, 89*(3), 450-502. doi: 10.1111/j.1468-0009.2011.00636.x

Kesselheim, A. S., Myers, J. A., & Avorn, J. (2011). Characteristics of clinical trials to support approval of orphan vs nonorphan drugs for cancer. *Journal of the American Medical Association, 305*(22), 2320-2326. doi: 10.1001/jama.2011.769

Korn, E. L., Freidlin, B., Abrams, J. S., & Halabi, S. (2012). Design issues in randomized phase II/III trials. *Journal of Clinical Oncology, 30*(6), 667-671. doi: 10.1200/JCO.2011.38.5732

Kristensen, F. B., & Gerhardus, A. (2010). Health technology assessments: what do differing conclusions tell us? *British Medical Journal, 341*, c5236. doi: 10.1136/bmj.c5236

Kristensen, F. B., Makela, M., Neikter, S. A., Rehnqvist, N., Haheim, L. L., Morland, B., Milne, R., Nielsen, C.P., Busse, R., Lee-Robin, S.H., Wild, C., Espallargues, M., Chamova, J. (2009). European Network for Health Technology Assessment, EUnetHTA: Planning, development, and implementation of a sustainable European network for Health Technology Assessment. *International Journal of Health Technology Assessment in Health Care, 25*(S2), 107-116.

Le Cam, Y. (2010). Inventory of access and prices of orphan drugs across Europe: a collaborative work between national alliances on rare diseases & Eurordis: Eurordis Rare Diseases Europe.

Lexchin, J., & Mintzes, B. (2008). Medicine reimbursement recommendations in Canada, Australia, and Scotland. *American Journal of Managed Care, 14*(9), 581-588.

Light, D. W., & Kantarjian, H. (2013). Market Spiral Pricing of cancer drugs. *Cancer, 119*(22), 3900-3902. doi: 10.1002/Cncr.28321

Long, G., & Works, J. (2013). Innovation in the pharmaceutical pipeline: a multidimensional view. USA: Analysis Group. Economic, financial and strategy consultants. http://phrma.org/sites/default/files/pdf/2013innovationinthebiopharmaceuticalpipeline-analysisgroupfinal.pdf. Viewed November 2014.

Lu, G., & Ades, A. E. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine, 23*(20), 3105-3124. doi: 10.1002/sim.1875

Luce, B., Drummond, M., Jonsson, B., Neumann, P., Schwartz, J., Siebert, U., & Sullivan, S. (2010). EBM, HTA, and CER: clearing the confusion. *Milbank Quarterly, 88*(2), 256-276. doi: 10.1111/j.1468-0009.2010.00598.x

Madan, J., Stevenson, M. D., Cooper, K. L., Whyte, S., & Akehurst, R. (2011). Consistency between direct and indirect trial evidence: Is direct evidence always more reliable? *Value in Health, 14*(6), 953-960. doi: 10.1016/j.jval.2011.05.042

Malottki, K., Biswas, M., Deeks, J. J., Riley, R. D., Craddock, C., Johnson, P., & Billingham, L. (2014). Stratified medicine in European Medicines Agency licensing: a systematic review of predictive biomarkers. *British Medical Journal Open, 4*(1), e004188. doi: 10.1136/bmjopen-2013-004188

McCabe, C. (2010). Balancing economic, ethical and equity concerns in orphan drugs and rare diseases. *European Journal of Hospital Pharmacy and Practice, 16*(4), 22-25.

McCabe, C., Claxton, K., & Culyer, A. J. (2008). The NICE cost-effectiveness threshold: what it is and what that means. *Pharmacoeconomics, 26*(9), 733-744.

McCabe, C., Claxton, K., & Tsuchiya, A. (2005). Orphan drugs and the NHS: should we value rarity? *British Medical Journal, 331*(7523), 1016-1019. doi: 10.1136/bmj.331.7523.1016

McCabe, C., Stafinski, T., & Menon, D. (2010). Is it time to revisit orphan drug policies? *British Medical Journal, 341*. doi: 10.1136/Bmj.C4777

McCabe, C., Tsuchiya, A., Claxton, K., & Raftery, J. (2006). Orphan drugs revisited. *International Journal of Medicine, 99*(5), 341-345. doi: Doi 10.1093/Qjmed/Hcl033

McCabe, C., Tsuchiya, A., Claxton, K., & Raftery, J. (2007). Assessing the economic challenges posed by orphan drugs: A comment on Drummond et al.

International Journal of Technology Assessment in Health Care, 23(3), 397-401.

McClimans, L. M., & Browne, J. (2011). Choosing a patient-reported outcome measure. *Theoretical Medicine and Bioethics, 32*(1), 47-60. doi: 10.1007/s11017-010-9163-8

McGuire, M., Drummond, M., & Rutten, F. (2002). "Reimbursement policies in the European Union". In E. Mossialos, M. Mrazek & T. Walley (Eds.), *Regulating pharmaceuticals in Europe: striving for efficiency, equity, and quality*.

McMahon, M., Morgan, S., & Mitton, C. (2006). The Common Drug Review: A NICE start for Canada? *Health Policy, 77*(3), 339-351. doi: 10.1016/j.healthpol.2005.08.006

Merten, D. (2011). Publishing mixed methods research. Journal of mixed methods. *Journal of Mixed Methods Research, 5*(1), 3-6.

Michel, M., & Toumi, M. (2012). Access to orphan drugs in Europe: current and future issues. *Expert Review of Pharmacoeconomics and Outcomes Research, 12*(1), 23-29. doi: 10.1586/erp.11.95

Ministere des Affaires Sociales et de Sante. (2004). *La loi relative à la politique de santé publique: objectifs de santé, plans et programmes d'action.* http://www.sante.gouv.fr/cdrom_lpsp/pdf/587_596.pdf. Viewed November 2014.

Miyamoto, B. E., & Kakkis, E. D. (2011). The potential investment impact of improved access to accelerated approval on the development of treatments for low prevalence rare diseases. *Orphanet Journal of Rare Diseases, 6*. doi: 10.1186/1750-1172-6-49

Morgan, S. G., McMahon, M., Mitton, C., Roughead, E., Kirk, R., Kanavos, P., & Menon, D. (2006). Centralized drug review processes in Australia, Canada, New Zealand, and the United Kingdom. *Health Affairs, 25*(2), 337-347. doi: 10.1377/hlthaff.25.2.337

Mousnad, M. A., Shafie, A. A., & Ibrahim, M. I. (2014). Systematic review of factors affecting pharmaceutical expenditures. *Health Policy, 116*(2-3), 137-146. doi: 10.1016/j.healthpol.2014.03.010

NICE. National Institute for Health and Clinical Excellence - Who are we. http://www.nice.org.uk/aboutnice/whoweare/who_we_are.jsp. Viewed May 2012.

NICE. (2004). *Public involvement programme (PIP).* http://www.nice.org.uk/about/nice-communities/public-involvement/public-involvement-programme.Viewed May 2012.

NICE. (2008). Guide to the methods of technology appraisal. National Institute for Health and Care Excellence,.

NICE. (2009a). Briefing paper for methods review workshop on exploring uncertainty. National Institute for Health and Care Excellence.

NICE. (2009b). Cetuximab for the treatment of metastatic colorectal cancer - single technology appraisal. NICE technology appraisal guidance 176. http://publications.nice.org.uk/cetuximab-for-the-first-line-treatment-of-metastatic-colorectal-cancer-ta176. Viewed December 2011.

NICE. (2009c). Supplementary advice to the Appraisal Committees. http://www.nice.org.uk/media/E4A/79/SupplementaryAdviceTACEoL.pdf. Viewed December 2011.

NICE Citizens Council. (2014). Citizens Council. from https://www.nice.org.uk/Get-Involved/Citizens-Council. Viewed March 2015.

NICE Citizens Council Report. (2004). *Ultra orphan drugs*. London: Retrieved from http://www.nice.org.uk/media/065/85/Citizens_Council_Ultraorphan.pdf. Viewed March 2012.

NICE (2007). Briefing paper for methods review workshop on exploring uncertainty. Retrieved from: https://www.nice.org.uk/media/4A6/52/FinalBriefingPaperUncertaintyForDistribution080807.pdf Viewed September 2016.

NICE (2013). Guide to the methods of technology appraisal 2013. Retrieved from: https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781 Viewed September 2016.

Nicod, E. (2010). Harmonizing the outcomes of HTA: an application to central nervous system treatments. *EuroObserver, 12*(4), 9-14.

Nicod, E. (2016a). Why do health technology assessment coverage recommendations for the same drugs differ across settings? Applying a mixed methods framework to systematically compare orphan drug decisions in four European countries. *European Journal of Health Economics (in press, accepted for publication on 2 August 2016)*

Nicod, E., & Kanavos, P. (2012). Commonalities and differences in HTA outcomes: A comparative analysis of five countries and implications for coverage decisions. *Health Policy, 108*(2-3), 167-177. doi: DOI 10.1016/j.healthpol.2012.09.012

Nicod, E., & Kanavos, P. (2015). Developing an evidence-based methodological framework to systematically compare HTA coverage decisions across

countries: a mixed methods study. *Health Policy, Accepted for publication, November 2015*. doi: 10.1016/j.healthpol.2015.11.007

Nicod, E., & Kanavos, P. (2016b). Scientific and social value judgments for orphan drugs in health technology assessment. *International Journal for Technology Assessment in Healthcare, 32(4), 1-15*.

Nicod, E. Annemans, L., Bucsics, A., Lee, A., Upadhyaya, S., Facey, K. (2016c). HTA programme response to the challenges of dealing with orphan medicinal products: Process evaluation in selected European Countries. *Health Policy, a*ccepted pending revisions (September 2016).

NIH. (2011). Global health and ageing. National Institute on Ageing. National Institute for Health. U.S. Department of Health and Human Services. *NIH Publication* no. 11-7737. http://www.nia.nih.gov/sites/default/files/global_health_and_aging.pdf. Viewed November 2014.

O'Donnell, J. C., Pham, S. V., Pashos, C. L., Miller, D. W., & Smith, M. D. (2009). Health Technology Assessment: Lessons learned from around the world-an overview. *Value in Health, 12*, S1-S5. doi: 10.1111/j.1524-4733.2009.00550.x

OECD. (2014). Pharmaceutical expenditure, as a percentage expenditure on total health. http://www.oecd-ilibrary.org/sites/pharmexp-table-2014-1-en/index.html;jsessionid=5ce2sr0ojqb2n.x-oecd-live-02?contentType=%2fns%2fStatisticalPublication%2c%2fns%2fTable%2c%2fns%2fKeyTableEdition&itemId=%2fcontent%2ftable%2fpharmexp-table-2014-1-en&mimeType=text%2fhtml&containerItemId=%2fcontent%2ftable%2fpharmexp-table-en&accessItemIds=%2fcontent%2ftablecollection%2f20758480. Viewed March 2015.

OECD Health Statistics. (2013). Retrieved from: http://dx.doi.org/10.1787/health-data-en. Viewed November 2014.

Onwuegbuzie, A., & Combs, J. (2010). "Emergent data analysis techniques in mixed methods research: a synthesis". In A. Tashakkori & C. Teddlie (Eds.), *Sage Handbook of Mixed Methods in Social & Behavioral Research* (2nd ed., pp. pp.397-430). Los Angeles: SAGE Publications.

Orphanet. (2015). About orphan drugs. http://www.orpha.net/consor/cgi-bin/Education_AboutOrphanDrugs.php?lng=EN&stapage=ST_EDUCATION_EDUCATION_ABOUTORPHANDRUGS_EUR. Viewed November 2015.

Orr, D. (2014). Breast cancer drug Kadcyla is too expensive for the NHS - not to pay for it is the correct decision, *The Guardian*. http://www.theguardian.com/commentisfree/2014/apr/25/kadcyla-cancer-drug-too-expensive-nhs. Viewed November 2014.

Panju, A. H., & Bell, C. M. (2010). Policy alternatives for treatments for rare diseases. *Canadian Medical Association Journal, 182*(17), E787-E792. doi: 10.1503/Cmaj.081429

PBAC. (2007). Public summary document - Paliperidone, prolonged release tablets, 3mg, 6mg, 9mg, 12mg, Invega. Sponsor: Janssen-Cilag Pty-Ltd. http://www.health.gov.au/internet/main/publishing.nsf/Content/pbac-psd-paliperidone-nov07. Viewed December 2011.

PBAC. (2008). Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Board (version 4.3). Pharmaceutical Benefits Advisory Committee (PBAC). Department of Health and Ageing. Australian Government.

PBAC. (2010). Cetuximab, solution for IV infusion, 100 mg in 20 mL, 100 mg in 50 mL and 500 mg in 100 mL, Erbitux.

http://www.health.gov.au/internet/main/publishing.nsf/Content/pbac-psd-cetuximab-mar-10. Viewed December 2011.

Pearson, S. D., & Rawlins, M. D. (2005). Quality, innovation, and value for money: NICE and the British National Health Service. *Journal of American Medical Association, 294*(20), 2618-2622. doi: 10.1001/jama.294.20.2618

Persson, U. (2012). Value based pricing in Sweden: lessons for design? *Office for health economics, seminar briefing, 12*, 1-11.

Pharmaceutical Forum. (2013). *Final conclusions and recommendations of the Pharmaceutical Forum: improving access to orphan medicines for all affected EU citizens*. European Commission, Entreprise and Industry. http://ec.europa.eu/enterprise/sectors/healthcare/files/docs/pricing_orphans_en.pdf. Viewed November 2014.

Pomedli, S. (2010). HTA and access to cancer medicines. *EuroObserver, 12*(4), 7-8.

Portfolio Management Solutions. Attrition. http://www.portfoliomanagementsolutions.com/the-organization-of-pharmaceutical-rd/attrition/. Viewed January 2015.

PPRS. Pharmaceutical Price Regulation Scheme. http://www.dh.gov.uk/en/Healthcare/Medicinespharmacyandindustry/Pharmaceuticalpriceregulationscheme/index.htm. Viewed May 2012.

Price, V. E., Klaassen, R. J., Bolton-Maggs, P. H. B., Grainger, J. D., Curtis, C., Wakefield, C., Dufort, G., Riedlinger, A., Soltner, C., Blanchette, V.S., Young, N. L. (2009). Measuring disease-specific quality of life in rare populations: a practical approach to cross-cultural translation. *Health and Quality of Life Outcomes, 7*. doi: 10.1186/1477-7525-7-92

Prostate Cancer UK. (2014, 15 August 2014). NICE decision on prostate cancer drug is a "kick in the teeth" for patients, *The Guardian*. http://www.theguardian.com/society/2014/aug/15/nice-prostate-cancer-drug-abiraterone. Viewed November 2014.

QSR International Pty Ltd. (2012). NVivo qualitative data software (Version 10).

Rare Best-Practices. (2013). A platform for sharing best practices for the management of rare diseases. http://www.rarebestpractices.eu/documenti/setup/1_Attachment_171.pdf. Viewed November 2014.

Rawlins, M. (2008). De testimonio: on the evidence for decisions about the use of therapeutic interventions. *Lancet, 372*(9656), 2152-2161. doi: 10.1016/S0140-6736(08)61930-3

Rawlins, M., Barnett, D., & Stevens, A. (2010). Pharmacoeconomics: NICE's approach to decision-making. *British Journal of Clinical Pharmacology, 70*(3), 346-349. doi: 10.1111/j.1365-2125.2009.03589.x

Rawlins, M. D. (2014). Evidence, values, and decision making. *International Journal of Technology Assessment in Health Care, 30*(2), 233-238. doi: 10.1017/S0266462314000154

Rawlins, M. D., & Culyer, A. J. (2004). National Institute for Clinical Excellence and its value judgments. *British Medical Journal, 329*(7459), 224-227. doi: 10.1136/bmj.329.7459.224

Richards, M. (2010). Extent and causes of international variation in drug usage: A report for the Secretary of State for Health. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216249/dh_117977.pdf. Viewed November 2012.

Russo, P., Mennini, F. S., Siviero, P. D., & Rasi, G. (2010). Time to market and patient access to new oncology products in Italy: a multistep pathway from European context to regional health care providers. *Annals of Oncology, 21*(10), 2081-2087. doi: 10.1093/annonc/mdq097

Scherer, F. M. (2000). "Handbook of health economics". In A. J. Culver & J. P. Newhouse (Eds.), *Handbooks in economics* (pp. 1298-1336). Amsterdam, Oxford: Elsevier.

Schwappach, D. L. (2002). Resource allocation, social values and the QALY: a review of the debate and empirical evidence. *Health Expectations, 5*(3), 210-222.

Schwarzer, R., & Siebert, U. (2009). Methods, procedures, and contextual characteristics of health technology assessment and health policy decision making: Comparison of health technology assessment agencies in Germany, United Kingdom, France, and Sweden. *International Journal of Technology Assessment in Health Care, 25*(3), 305-314. doi: 10.1017/S0266462309990092

Sculpher, M. (2008). NICE's 2008 Methods Guide - Sensible consolidation or opportunities missed? *Pharmacoeconomics, 26*(9), 721-724. doi: 10.2165/00019053-200826090-00001

Sculpher, M. J., Claxton, K., Drummond, M., & McCabe, C. (2006). Whither trial-based economic evaluation for health care decision making? *Health Economics, 15*(7), 677-687. doi: 10.1002/Hec.1093

Severens, J.L., Milne, R.J. (2004). Discounting health outcomes in economic evaluation: The ongoing debate. *Value in Health, 7(4),* 397-401.

Shah, K. K., Mestre-Ferrandiz, J., Towse, A., & Smyth, E. N. (2013). A review of health technology appraisals: Case studies in oncology. *International Journal*

of Technology Assessment in Health Care, 29(1), 101-109. doi: 10.1017/S0266462312000669

Simoens, S. (2011). Pricing and reimbursement of orphan drugs: the need for more transparency. *Orphanet Journal of Rare Diseases, 6*. doi: 10.1186/1750-1172-6-42

Simoens, S., & Dooms, M. (2011). How much is the life of a cancer patient worth? A pharmaco-economic perspective. *Journal of Clinical Pharmacy and Therapeutics, 36*(3), 249-256. doi: 10.1111/j.1365-2710.2010.01181.x

Sixty-seventh World Health Assembly. (2014). Health intervention and technology assessment in support of universal health coverage (Vol. WHA67.23, Agenda item 15.7).

SMC. (2008). Paliperidone 3, 6 and 9mg prolonged release tablets (Invega). Advice no. 453/08. http://www.scottishmedicines.org.uk/SMC_Advice/Advice/453_08_paliperidone_tablets/paliperidone_prolonged-release_tablets__INVEGA_. Viewed December 2011.

SMC. (2009). Advice, Cetuximab 100mg/20mL and 500mg/100mL solution for intraveneous infusion (Erbitux). no. 543/09. http://www.scottishmedicines.org.uk/SMC_Advice/Advice/543_09_cetuximab__Erbitux_/cetuximab__Erbitux_. Viewed December 2011.

SMC. (2011a). *Access to newly licensed medicines*. http://www.scottish.parliament.uk/S4_HealthandSportCommittee/Inquiries/ANM023_Scottish_Medicines_Consortium.pdf. Viewed December 2011.

SMC. (2011b). Guidance to manufacturers from completion of new product assessment form (NPAF). Scottish Medicines Consortium.

SMC. (2011c). Policy Statement on Orphan drugs. http://www.scottishmedicines.org.uk/About_SMC/Policy_Statements/Orphan_Drugs Viewed December 2011.

SMC. (2012). SMC Modifiers used in Appraising New Medicines. https://www.scottishmedicines.org.uk/About_SMC/Policy_statements/SMC_Modifiers_used_in_Appraising_New_Medicines. Viewed December 2012.

SMC. (2013a). Public Involvement. from http://www.scottishmedicines.org.uk/Public_Involvement/Public_Involvement. Viewed December 2014.

SMC. (2013b). *Working with SMC - a guide for manufacturers*. https://www.scottishmedicines.org.uk/files/submissionprocess/Working_with_SMC_July_2014.pdf. Viewed December 2014.

SMC. (2015a). About SMC. https://www.scottishmedicines.org.uk/About_SMC/What_we_do/index. Viewed December 2012.

SMC. (2015b). SMC assessment process flow chart. https://www.scottishmedicines.org.uk/files/SMC_Assessment_Process_Flow_Chart.doc. Viewed December 2015.

Sorenson, C. (2009). The role of health technology assessment coverage and pricing decisions: A cross-country comparison. *EuroObserver, 11*(1), 1-2.

Sorenson, C., Drummond, M., & Bhuiyan Khan, B. (2013). Medical technology as a key driver of rising health expenditure: disentangling the relationship. *Clinicoeconomics and Outcomes Research, 5,* 223-234. doi: 10.2147/CEOR.S39634

Sorenson, C., Drummond, M., & Kanavos, P. (2008). Ensuring value for money in health care: the role of helath technology assessment in the European Union. *Observatory Studies Series 11*.

StataCorp. (2013). Stata Statistical Software: Release 13. TX: StataCorp LP.

Stolk, P., Heemstra, H. E., Leufkens, H. G., Bloechl-Daum, B., & Heerdink, E. R. (2009). No difference in between-country variability in use of newly approved orphan and non- orphan medicinal products--a pilot study. *Orphanet Journal of Rare Diseases, 4*, 27. doi: 10.1186/1750-1172-4-27

Sun, X., Briel, M., Busse, J. W., You, J. J., Akl, E. A., Mejza, F., . . . Guyatt, G. H. (2011). The influence of study characteristics on reporting of subgroup analyses in randomised controlled trials: systematic review. *British Medical Journal, 342*. Doi 10.1136/Bmj.D1569

Sutton, A., Ades, A. E., Cooper, N., & Abrams, K. (2008). Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics, 26*(9), 753-767. doi: 10.2165/00019053-200826090-00006

Svensson, M., Nilsson, F., Arnberg, K. (2015). Reimbursement decisions for pharmaceuticals in Sweden: The impact of disease severity and cost-effectiveness. *Pharmacoeconomics,* 33:1229.36.

Teddlie, C., & Tashakkori, A. (2010). "Overview of contemporary issues in mixed methods research". In A. Tashakkori & C. Teddlie (Eds.), *Sage Handbook of Mixed Methods in Social & Behavioral Research* (2nd ed., pp. 1-42). Los Angeles: SAGE Publications.

Praveen, T., Devlin, N. Marsh, K. Baltussen, R. Boysen, M., Kalo, Z. Longrenn, T. Mussen, Peacock, F., Watkins, J., Ijzerman, M. (2016). ISPOR Task Force Reprot. Multiple Criteria Decision Analysis for health care decision making—

An introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force.*Value in Health,* 13, 1-13.

TLV. (2003). General guidelines for economic evaluations from the Pharmaceutical Benefits Board (LFNAR 2003:2). http://www.tlv.se/Upload/English/Guidelines-for-economic-evaluations-LFNAR-2003-2.pdf. Viewed December 2011.

TLV. (2008). Sökande - Janssen-Cliag AB - Invega, depottablet 3, 6, 9 and 12mg. www.tlv.se. Viewed December 2011.

TLV. (2012). *Guide for companies when applying for subsidies and pricing for pharmaceutical products, Version 2.0* http://www.tlv.se/Upload/English/ENG-guide-for-companies.pdf. Viewed December 2011.

TLV. (2013). *Handbook – reviewing the reimbursement status of pharmaceuticals decision*. http://www.tlv.se/Upload/English/ENG-handbook.pdf. Viewed December 2013.

Tordrup, D., Mossman, J., & Kanavos, P. (2014). Responsiveness of the Eq-5d to Clinical Change: Is the Patient Experience Adequately Represented? *International Journal of Technology Assessment in Health Care, 30*(1), 10-19. doi: 10.1017/S0266462313000640

Towse, A. (2010). Value based pricing, research and development, and patient access schemes. Will the United Kingdom get it right or wrong? *British Journal of Clinical Pharmacology, 70*(3), 360-366. doi: 10.1111/j.1365-2125.2010.03740.x

United Nations Population Fund. (2011). World population to exceed 9 billion by 2050: developing countries to add 2.3 billion inhabitants with 1.1 billion aged

over 60 and 1.2 billion of working age. http://www.unfpa.org/pds/trends.htm. Viewed November 2014.

United States Congress. (1976). *The development of medical technology: opportunties for assessment*. http://govinfo.library.unt.edu/ota/Ota_5/DATA/1976/7617.PDF. Viewed November 2014.

US Food and Drug Administration. (2015). About FDA. Office of Orphan Products Development. http://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/OfficeofScienceandHealthCoordination/ucm2018190.htm. Viewed December 2015.

Van den Aardweg, S. (2010). HTA of orphan drugs across six countries: flair, flawed or failing? *EuroObserver, 12*(4), 14-19.

Vegter, S., Rozenbaum, M. H., Postema, R., Tolley, K., & Postma, M. J. (2010). Review of regulatory recommendations for orphan drug submissions in the Netherlands and Scotland: Focus on the underlying pharmacoeconomic evaluations. *Clinical Therapeutics, 32*(9), 1651-1661. doi: 10.1016/j.clinthera.2010.08.006

Velasco-Garrido, M., Borlum Kristensen, F., Palmhoj Nielsen, C., & Busse, R. (2008). Health technology assessment and helath policy-making in Europe: current status, challenges and potential. In European Observatory on Health Systems and Policy (Ed.), *Policy brief*.

Velasco-Garrido, M., & Busse, R. (2005). Health technology assessment, an introduction to objectives, role of evidence, and structure in Europe. In European Observatory on Health Systems and Policy (Ed.), *Policy Brief*.

Vernon, J. A., Golec, J. H., & Dimasi, J. A. (2010). Drug development costs when financial risk is measured using the Fama-French Three-Factor model. *Health Economics, 19*(8), 1002-1005. doi: 10.1002/Hec.1538

Vickers, P. J. (2013). Challenges and opportunities in the treatment of rare diseases. *Drug Discovery World, Spring*.

Wanless, D. (2002). Securing our future health: Taking a long-term view: HM Treasury. http://si.easp.es/derechosciudadania/wp-content/uploads/2009/10/4.Informe-Wanless.pdf. Viewed December 2012.

Webb, D. J. (2009). Early assessment of cost-effectiveness. *Basic & Clinical Pharmacology & Toxicology, 105*, 11-12.

Wieand, H. S. (2005). Randomized phase II trials: what does randomization gain? *Journal of Clinical Oncology, 23*(9), 1794-1795. doi: 10.1200/JCO.2005.10.956

Wonder, M. J., Neville, A. M., & Parsons, R. (2006). Are Australians able to access new medicines on the pharmaceutical benefits scheme in a more or less timely manner? An analysis of Pharmaceutical Benefits Advisory Committee recommendations, 1999-2003. *Value in Health, 9*(4), 205-212. doi: 10.1111/j.1524-4733.2006.00104.x

World Health Organisation. (2012). Disorders management: mental health. http://www.who.int/mental_health/management/schizophrenia/en/. Viewed September 2012.

World Health Organisation. (2014). *Health intervention and technology assessment in support of universal health coverage*. http://apps.who.int/gb/ebwha/pdf_files/WHA67/A67_R23-en.pdf. Viewed March 2015.

Appendix A: Selected HTA Countries

England & Wales: National Institute for Health and Care Excellence (NICE)

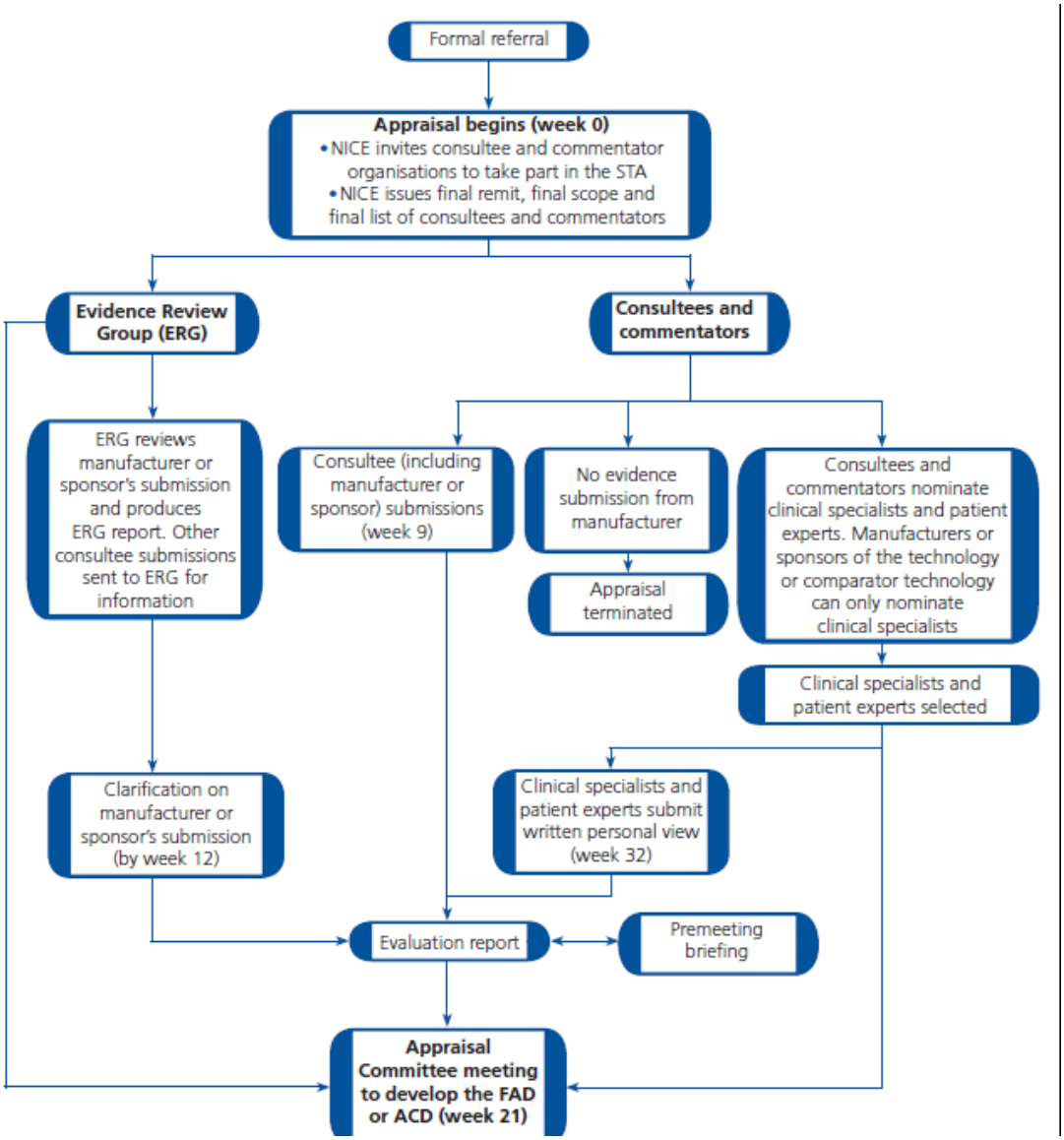
The National Institute for Health and Clinical Excellence (NICE) was established in 1999 to reduce regional variations in the drugs that are subsidized and acts as an arm's length organization funded by the Department of Health (NICE). NICE has a regulatory role in deciding on the reimbursement of medicines, and provides four types of guidance on the use of medicines: 1) technology appraisals, 2) clinical guidelines, 3) public health guidance, and 4) reports on interventional procedures (Drummond, 2008). Recommendations are therefore legally binding, and those receiving a negative decision are listed in a negative formulary. Those drugs that do not undergo the NICE process, decisions about whether or not they are provided are taken at local level by the Clinical Commissioning Groups (CCGs), who are responsible for ensuring that the services provided by the NHS meet patient needs.

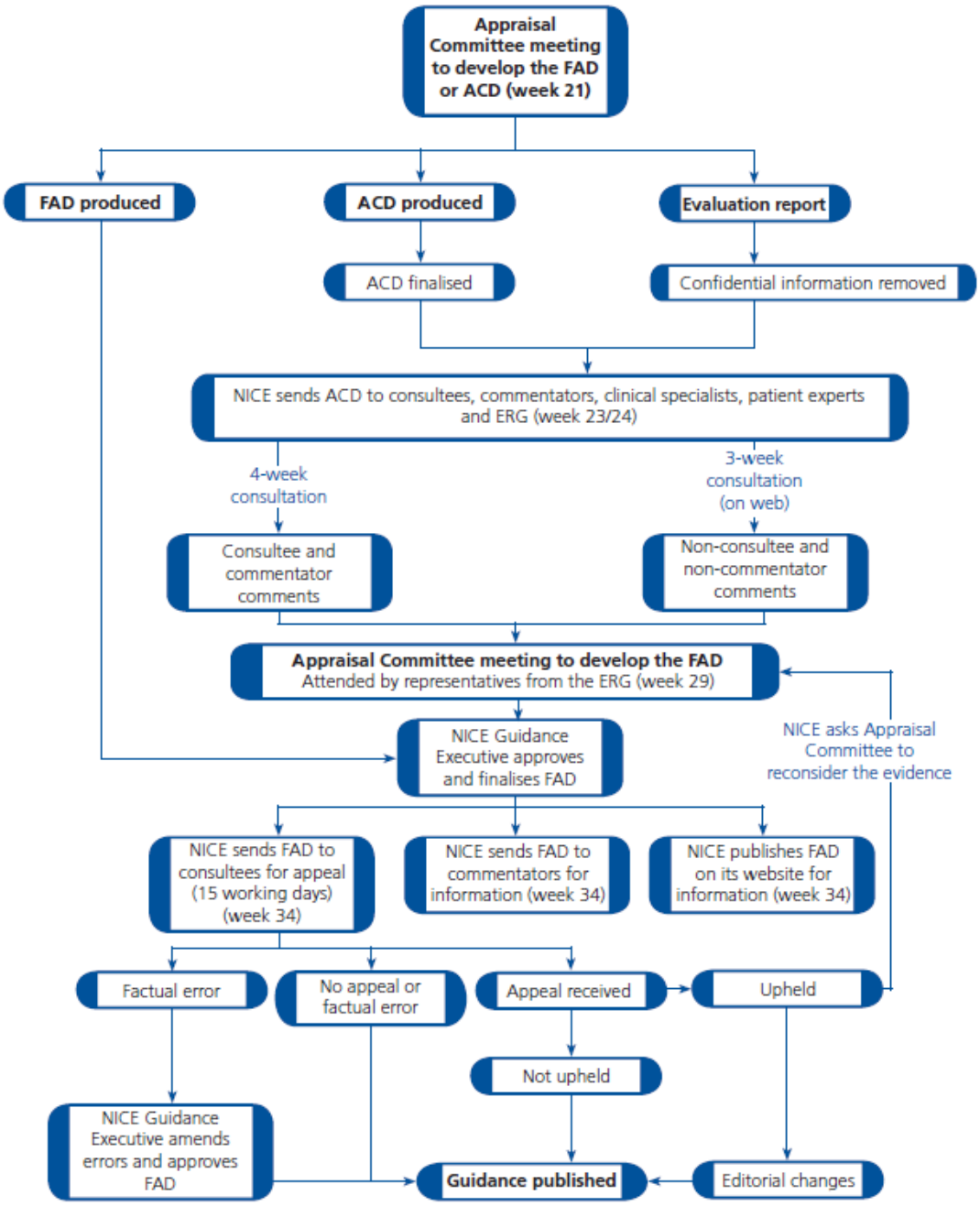
In terms of pricing, free-pricing of branded products is applicable under the voluntary PPRS scheme, where profits are regulated by a fixed allowable return on capital invested (PPRS). In terms of reimbursement decisions by NICE, two approaches exist to appraising pharmaceuticals. During the multiple technology appraisal approach (MTA), evidence for a group of competing technologies treating a specific condition is appraised based on the manufacturer's submission and evidence produced by the review team; the duration of the process is 52 weeks. The single technology appraisal (STA) was established at a later stage to speed up the process, where only one drug is appraised based mainly on evidence submitted by the manufacturer; the process lasts about 39 weeks depending on the number of appeals (Drummond, 2008). For comparability reasons, this thesis focuses solely on STA appraisals.

Initiation of the HTA process is done through the topic selection phase. Evidence is provided from several sources. The manufacturer or sponsor provides the main evidence. An independent expert committee (Technology Appraisal Committee (TAC)) represented by many stakeholders is commissioned by NICE to appraise the evidence and issue a report based on the technologies' relative benefit and cost-

effectiveness. On this basis, NICE issues its guidance. Stakeholders (manufacturers, patient groups) have the right to comment or appeal. Once the guidance is finalized it is sent to the NHS and the PCTs have three months to adopt it (Drummond & Sorenson, 2009b). Timelines for the STA process depend on the complexity of the cases, but generally are limited to 34 weeks.

NICE Single Technology Appraisal Process



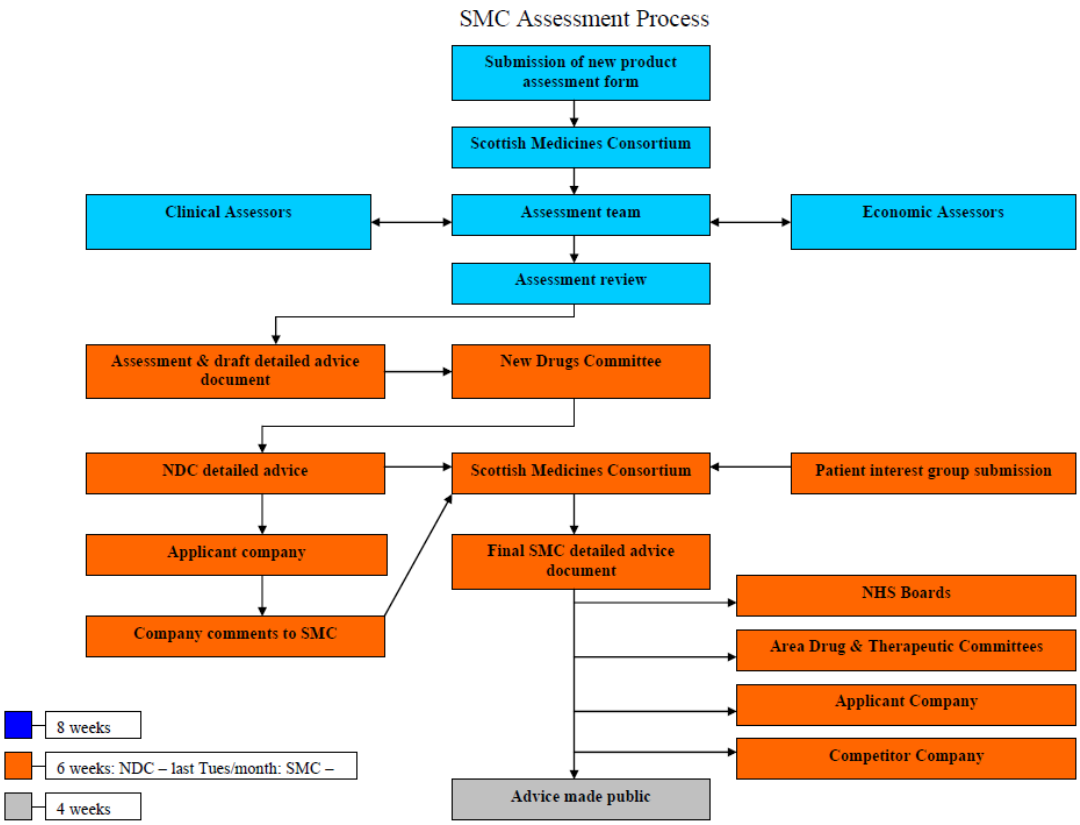


Source: Guide to the single technology appraisal process, NICE, October 2009

Scotland: Scottish Medicines Consortium (SMC)

The Scottish Medicines Consortium (SMC) was established in 2001 in order to issue an advice for all of Scotland about the value of each new medicine and the patients who would most benefit from these. The SMC is a consortium of NHSScotland's 14 Health Boards. The SMC issues an advice to the Health Boards and their Area Drug and Therapeutics Committees on the use of all newly licensed drugs. The Health Boards then recommend or not the use of these new drugs in their area, and it is up to the clinicians to decide whether or not to prescribe them (SMC, 2015a).

SMC assessment process flow chart



Source: (SMC, 2015b)

For all newly licensed medicines, manufacturers are requested to provide an HTA application before the drug in question is marketed in Scotland, or within 3 months of marketing authorisation. The independent committee of the SMC, the New Drug Committee (NDC), reviews this and issues a provisional advice on reimbursement.

During the SMC meeting, this advice, together with feedback from the manufacturer, is considered by the SMC. Patient Interest Groups are also given the opportunity to provide their input, to ensure that the patient and carer perspective is accounted for during the review process. The SMC issues a Detailed Advice Document (DAD), communicated to the NHS Boards and the pharmaceutical company making the submission. The advice for reimbursement is either (a) accepted for use, (b) accepted for restricted use, or (c) not recommended for use. Four weeks later, the advice is made public. Timelines for the SMC Assessment process is approximately 18 weeks.
The assessment relies on evidence about the drug's relative clinical and cost effectiveness using a health service perspective (e.g. NHS Scotland and social work), in order to demonstrate that the drug provides value for money based on a robust clinical and economic case.

Greater flexibility and uncertainty in the economic case is accepted in certain specific circumstances, where a greater uncertainty or higher cost per QALYs is accepted. This applies to orphan drugs as per the EMA definition of orphan designation. SMC recognises that trial patient populations may be smaller and accepts greater uncertainty in the economic case in this respect. Additional factors are also: life-threatening disease, substantial increase in life-expectancy or quality of life, reverses rather than stabilises the disease, or bridges the gap to a definitive therapy. More detail is required for other parts of the submission, such as the relevance of surrogate markers and the theoretical basis for their selection. Additionally, SMC requires that long term monitoring be implemented about the use of the drug (e.g. patient registries).

A higher ICER is accepted when the economic case is robust & additional factors pertaining to orphan drugs are true:

- Substantial improvement in life expectancy with sufficient quality of life (e.g. median gain of 3 months survival)
- Substantial improvement in quality of life (with or without survival benefit)
- Specific and extra benefit for a subgroup of patients
- Absence of therapeutic options provided by the NHS
- Possible bridging to another therapy (e.g. bone marrow transplantation)

- Licensed medicine as an alternative to an unlicensed product used in NHS Scottish clinical practice.
- Additional other drug and disease characteristics highlighted by the manufacturer and by clinical or patient experts (e.g. PAPIG).

Since October 2013, a new procedure has been implemented at SMC for end of life (for patients with less than 3 years life expectancy) and ultra-orphan drugs (for conditions affecting less than 1 in 50,000). This does not apply to our study given than it does not apply to our study period.

Sweden: Dental and Pharmaceutical Benefits Board (TLV)

The TLV is a regulatory body deciding on the pricing and reimbursement of drugs in Sweden. During the study period, TLV appraised primarily outpatient drugs. Three different processes exist: the new drug, the re-assessment, and hospital drug process. This study will focus solely on the new drug process. It is up to the company to apply for reimbursement. Drugs that haven't been assessed by TLV are in general not reimbursed. Reimbursement decisions are based on three principles: cost-effectiveness, human value, and need and solidarity, adopting a holistic perspective where medical, humanitarian and socio-economic aspects are considered. The company carries the burden of proof, and TLV's decision is based on the manufacturers' submission together other available evidence, and advice from scientific advisors and medical experts. A memorandum is written with the decision and course of action. Companies have a few weeks to comment, and if the decision is restrictive, companies also have an opportunity to discuss with the committee. The final decision is then made, and a written decision is issued to the public; this is the report analysed in this research.

TLV New Drug Appraisal process

Application from company
Questions of clarification to the company
Investigation by TLV
- Contact with external experts (if needed)

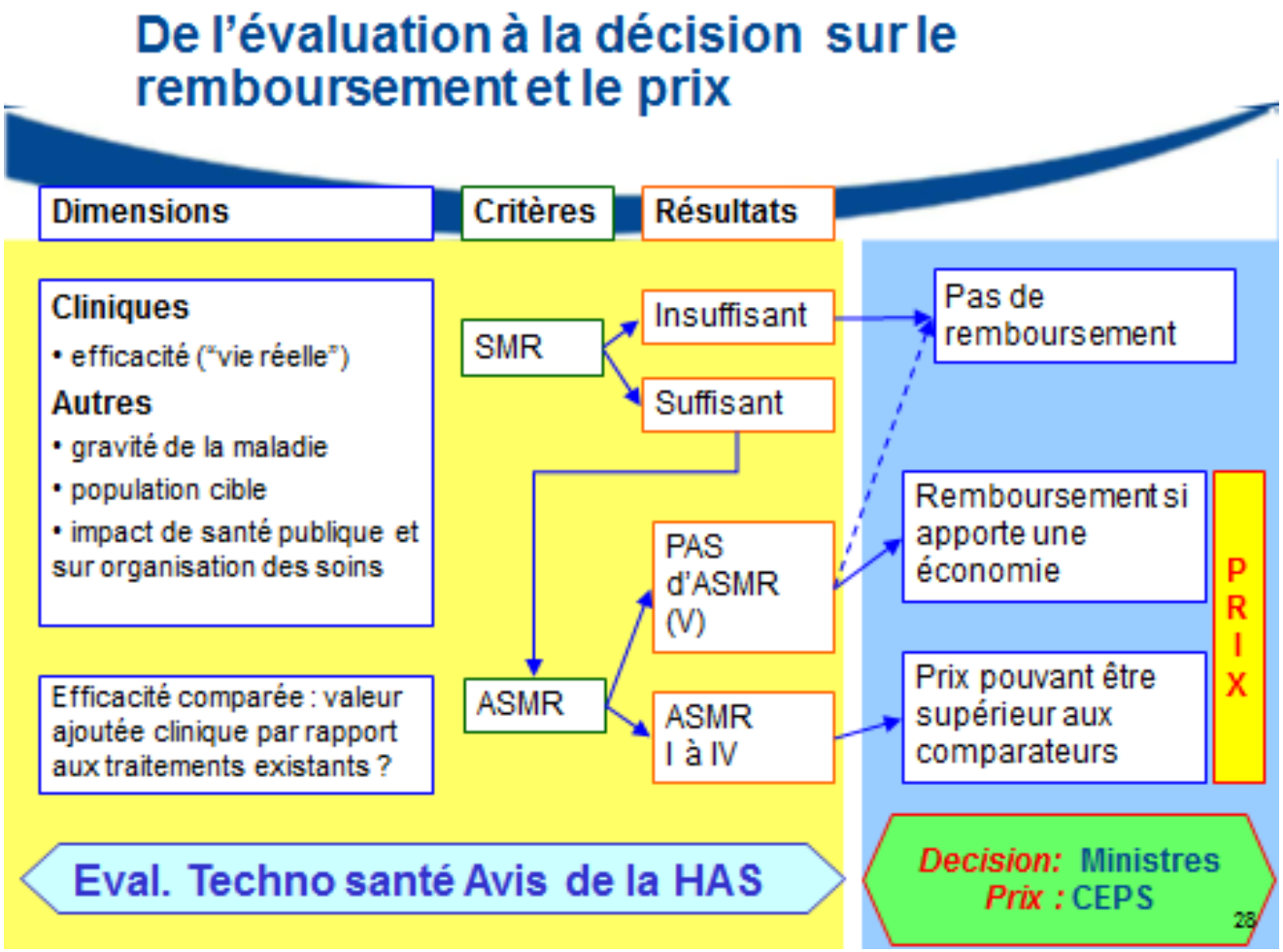
Memorandum to the company – with the suggested decision form the working group
Memorandum to the deciding committee
Company has the possibility of having a discussion with the committee if the decision is unfavorable for the company
Decision by the committee
Appeal (decided by the company)

Source: personal communication with TLV

France: Haute Autorité de Santé (HAS)

The HAS is an advisory body to the Minister of Health in France, assessing a drug's medical benefit using a health service perspective. The Commission nationale d'évaluation des dispositifs médicaux et des technologies de santé (CNEDiMTS) is an independent scientific body within the HAS that appraises the evidence. Two different processes exist: new drugs, and re-assessments; the first is the focus of this research. As of October 2013, economic evaluations are being conducted by HAS, but will not be included in this research given that the study drugs were appraised prior to this. All drugs undergo the assessment at HAS, where manufacturers are invited to submit an HTA application; the CNEDiMTS will also review the literature and in some instances request input from stakeholder experts. Focusing on the drugs that are approved for the first time, HAS focuses solely on the drug's clinical effectiveness, and assesses its medical benefit ("Service Médical Rendu") and relative improvement in medical benefit ("Amélioration du Service Médical Rendu"), ranking treatments from major to insufficient medical benefit and major innovations to no improvement through SMR and ASMR ratings respectively. SMR ratings drive the general coverage rate (e.g. a major or important SMR with a high disease severity is covered at 65% of the cost, for the remainder of cases it is covered at 35%). Patients with a rare disease are eligible for 100% coverage, regardless of the coverage for the drug. ASMR will drive reimbursement decisions based on improvements in therapeutic benefit in relation to the current standard of care. Assuming the SMR rating is positive, the ASMR assessment has been used to provide an appraisal of the drug's perception by the decision maker relative to its comparator(s). A technology receiving an ASMR V rating is classed as providing no additional therapeutic benefit and coverage occurs if the price of the technology is equal to, or lower than its comparator(s). The process lasts 180 days.

HAS new drug assessment process



Source: L'évaluation des médicaments à la HAS, Powerpoint presentation by Dr Francois Meyer, March 2011

Appendix B. Case Study example

Indication & HTA recommendation

Eltrombopag (Revolade©) received EMA marketing authorisation to treat chronic immune (idiopathic) thrombocytopenic purpura:

- in splenectomised adults whose condition is refractory to other treatments (corticosteroids, immunoglobulins), or
- as 2nd line treatment in non-splenectomised adults where surgery is contraindicated.

It underwent the HTA process in the four study countries for this same indication and received diverging recommendations: rejected by NICE, restricted by SMC, TLV, and HAS (Table 1).

Table B-1. HTA recommendations and restrictions

HTA body	NICE	SMC	TLV	HAS
Date of recommendation	10.2010	07.2010	05.2011	06.2010
Recommendation	Reject	Restrict	Restrict	SMR important (65% reimbursed)* ASMR II
Restrictions				
Patient population with severe ITP and high risk of bleeding		✓		
Re-assessment (October 2013)			✓	
Hospital use			✓	

* Coverage is 100% under the ALD programme (Affectation de Longue Duree)

Evidence submitted

Efficacy & Safety

The primary phase III RCT trial, RAISE, was appraised by all alongside a number of additional secondary trials. In the HTA report, NICE was the only agency to report results from subgroup analyses for two groups of patients: splenectomised (whose spleen has been removed) and non-splenectomised patients (whose spleen has not been removed). The additional trials considered included: (1) a dosage research phase II trial (TRA100773A) considered by NICE and HAS; (2) a phase III trial (TRA100773B) considered by all (despite the trial name not being specified in SMC's report, it was assumed to correspond based on the same trial design and outcomes); (3) a meta-analysis including three trials only considered by NICE (RAISE, TRA100773A, and TRA100773B); (4) an open-label (REPEAT) and an open extension (EXTEND) trial only considered by HAS; and (5) an indirect comparison of eltrombopag (RAISE) with romiplostim (Kuter 2008) considered by NICE, SMC, TLV (Table B-2).

Table B-3 summarises the clinical endpoints for each of the trials considered and included in the HTA reports. A variety of ways to report the clinical endpoints were seen. Some of the endpoints were not reported by all agencies, and in other cases the outcomes depended on the population or instrument used. For instance, WHO 3-4 bleeding events, which correspond to the more severe events, were only reported by NICE. Quality of life relied on the number of domains reported, where it was significant over 4 and not over 6. Results from the indirect comparison were significant across the whole population, but non-significant when considering the subgroups of patients separately, and so forth (Table B-2).

Table B-2. Clinical trials and endpoints

	NICE	SMC	TLV	HAS
RAISE Phase III, N 197, 6 months, placebo-controlled				
Platelet response	✓*	✓*	✓	✓
Rescue treatment	✓	✓		✓
Bleeding (WHO 1-4)	✓	✓	✓	✓

Bleeding (WHO 2-4) (= clinically significant bleeding)	✓	✓		✓
Bleeding (WHO 3-4)	×			
Concomitant medication (% reducing or stopping concomitant medication)		✓	✓	✓
HRQOL (SF36) (4) = 4 domains; (6) = 6 domains	✓ (4) × (6)	✓ (4)		
Response duration				◆
75% minimum response				✓
Improvement activities and concerns associated with Thrombocytopenia		✓		
Main reduction in bleeding in group WHO 2	×			
Subgroup RAISE Splenectomised (1/3)				
Platelet count	✓	✓		
Rescue treatment	×			
Bleeding (WHO1-4)	×			
Bleeding (WHO2-4)	✓			
Subgroup RAISE Non-splenectomised (2/3)				
Platelet count	✓	✓		
Rescue treatment	×			
Bleeding (WHO1-4)	✓			
Bleeding (WHO2-4)	✓			
TRA100773A Phase II, N 118, 6 weeks, placebo-controlled				
Platelet response	❖			✓
TRA100773B Phase III, N 114, 6 weeks, placebo-controlled				
Platelet response	❖	✓	✓	✓
Bleeding	✓	✓	✓	✓
HRQOL		×		
REPEAT (TRA 108057) Open-label, N 66, 26 weeks				
Platelet response				❖
EXTEND (TRA 105325) Open extension, N 207, Average 91.5 days (range 2-523 days)				
Platelet response				◆
Secondary endpoints				◆
Indirect comparison RAISE-KUTER 2008				
Platelet response	✓	◆	◆	
Platelet response, splenectomised	×			
Platelet response, non-splenectomised	×			

Legend. ✓: statistically significant, ×: not statistically significant, ◆: reported, ❖: considered within a meta-analysis (combined result).

Safety

The safety profile of eltrombopag was appraised based on a number of types of adverse events. These were reported by all agencies, except TLV. Table B-3 suggests that SMC and HAS conducted a more thorough toxicity assessment, according to the number of more severe adverse events reported. Some of these risks were a major concern for HAS, whose HTA recommendation was conditional until reassessment of results from the risk management plan to be implemented by the manufacturer.

Table B-**3.** Adverse events reported in the different countries

	NICE	SMC	TLV	HAS
Type of adverse events (AE) reported				
Common AE	✓	✓		✓
Clinically significant or severe AE	✓	✓		✓
% of cases with AE				✓
Deaths				✓
Discontinuation		✓		✓
Similar to SPC				✓
Reporting of specific more severe or significant events				
Liver dysfunction		✓		
Cataracts		✓		✓
Reticulin deposits in bone marrow		✓		✓
Hepatobiliary laboratory abnormalities	✓			✓
Thromboembolic events	✓	✓		✓
Risk of malignant haematological diseases				✓
Recurrence of thrombocytopenia on treatment cessation		✓		✓
Risk of bone marrow fibrosis				✓

Economic models

A variety of economic models were submitted in the different study countries. Differences were seen in the comparators and types of economic models were used. Three cost-utility models were reported for NICE: a) the "watch and rescue model" that compared eltrombopag and "watch and rescue" to "watch and rescue" alone, b) the "long-term continuous model" that looked at eltrombopag as part of a long-term sequence of treatments, and c) a cost-effectiveness model comparing eltrombopag to romiplostim. "Watch and rescue" treatment consists of immunosuppressive agents and if necessary rescue treatments for bleeding with anti-D or intravenous immunoglobulin. In contrast, the SMC submission included a cost-utility model comparing eltrombopag with romiplostim, and TLV a cost-minimisation analysis with the same comparator.

Interpretation of the evidence

Clinical uncertainties

A number of concerns about the clinical evidence were raised by the different agencies (Table B-4). NICE and SMC were concerned about the **trial duration** (6 months for RAISE, and 6 weeks for the other trials), which were deemed too short to reflect clinical practice and long-term safety. This was further confirmed by clinical experts at NICE. This concern was also raised by HAS, but because the treatment was considered to partially address an important public health need, the benefit of eltrombopag was nevertheless considered similar to its comparator romiplostim despite the lacking of direct comparative data.

TLV and NICE were also concerned about the **lack of direct comparative data** with romiplostim. For SMC, the main issue was that the effect of eltrombopag as a 2nd line treatment is unknown given that only placebo comparisons were presented.

NICE, SMC and TLV considered the indirect comparison of eltrombopag (RAISE) with romiplostim (Kuter 2008), but only NICE reported the detailed results in the HTA

report. NICE and SMC raised a number of issues related to these. The nature of this **indirect comparison was deemed uncertain**, which considered both alternatives to have similar effect. Reasons for these uncertainties were due to the differences between the study populations and the specification of the clinical endpoint (overall response). In the case of NICE, the manufacturer and ERG revised the comparison, which showed little impact on the initial results and was statistically significantly in favour of romiplostim. The sensitivity analysis showed no differences between treatment groups. For SMC, sensitivity analysis showed no difference in efficacy between both treatments; it also indicated that a mixed treatment comparison may have been more appropriate to account for trial differences.

Although eltrombopag demonstrated improved outcomes in platelet response and need for rescue medication, NICE was concerned by the fact that this improvement was **not significant in the more severely affected** patient population (Bleeding events WHO3-4); this was not raised by the other agencies, and what is more, this clinical endpoint was not recorded in their HTA report.

HAS was concerned that the impact of eltrompobag on **quality of life** was not adequately documented, although it was considered to be severely affected by ITP. This was not raised by TLV, and was not a concern for NICE and SMC, given that quality of life data was included in the submission.

Eltrombopag is indicated in "splenectomised adults whose condition is refractory to other treatments, and as 2nd line treatment in non-splenectomised adults where surgery is contraindicated". NICE however raised concerns that the **trial population was not in line with the indication being assessed**. First, a 17% response rate in RAISE was seen in the comparator arm, together with a minority of patients received treatment with intravenous immunoglobulin, which goes against the assumption that these patients are refractory to other treatments; second, the study population included in the non-splenectomised group did not have any contraindications to splenectomy. Further, NICE was concerned that RAISE included **only patients with low platelet count** where the risk of bleeding is variable**,** and not those with a persistent risk of bleeding.

Trial representativeness was raised by NICE and the SMC, where the trial population was deemed not to be representative of the UK and Scottish population respectively; clinical experts at NICE, however, confirmed the plausibility of this concern. For the SMC, initiation of treatment with eltrombopag based on platelet count is not deemed to **reflect clinical practice** that would also consider symptoms and whether an invasive procedure is planned. Finally, SMC was concerned that eltrombopag requires **liver function monitoring.**

Table B-**4.** Clinical uncertainties, and whether they relate to the same evidence and were put forward as one of the main reasons for the final recommendation.

	Evidence Considered by				**Interpretation** Uncertainty raised by				**Outcome** Main reason for recommendation			
	✓evidence considered ✓*evidence considered within the primary trial				✓ positive influence (addressed)		× negative influence (not addressed)		✓ positive influence (addressed)		× negative influence (not addressed)	
Clinical uncertainties	NICE	SMC	TLV	HAS	NICE	SMC	TLV	HAS	NICE	SMC	TLV	HAS
Lack of comparator (RAISE)	✓*	✓*	✓*	✓*	×	×	×	×			×	×
Short duration of trial (RAISE, 6 months)	✓*	✓*	✓*	✓*	×	×		✓	×			✓
Sample size (RAISE)	✓*	✓*	✓*	✓*			✓				✓	
Trial population, indication under review not corresponding (for patients refractory to other treatments, contraindication to splenectomy)	✓*	✓*	✓*	✓*	×				×			
Trial population generalizability	✓*	✓*	✓*	✓*	✓	×						
Trial population, low platelet count patients instead of those with severe risk of bleeding	✓*	✓*	✓*	✓*	✓							
Low benefit for populations in the lower	✓*				×				×			

incidence of the more severe bleeding events (WHO 3-4)										
Lack of quality of life data	✓	✓						×		×
Liver function monitoring required		✓					×			
Uncertain nature of the indirect comparison (different trial populations, overall response not pre-specified in RAISE)	✓	✓	✓			×	×			

Other considerations

A number of "other considerations" were identified in the reports, some of which were also put forward as one of the main reasons for the final decision (Table 5). **Unmet need** was highlighted by NICE for specific subgroups of the population for which no evidence was submitted, and by TLV and HAS who acknowledged that generally limited treatment options are available for this population. HAS mentioned that the treatment is **curative**. Clinical experts in England qualify eltrombopag as being **innovative,** with a new mechanism of action, together with romiplostim, but no evidence demonstrating distinctive benefits was presented and therefore this had no impact on the final outcome. TLV recognised that eltrombopag was part of a new class of drugs, and SMC and TLV recognised its advantage from oral administration. The drug's **orphan status** seems to provide favourable outcomes for the size of the trial population for TLV. For SMC, the orphan status of eltrombopag resulted in accepting greater uncertainty in the evidence. For HAS, the rarity of the condition signifies that the public health burden is small. NICE, TLV and HAS have highlighted the important impact of the disease on **quality of life** to the patient. NICE also accounted for the impact of the disease on patients' daily activities, lifestyles, ability to work, travel and conduct leisure activities, social stigma, or the fact that the disease is life threatening. For NICE although patients may have some anxiety to undergo surgery, clinical specialists indicate that surgery is today routinely performed with very few medical contraindications. Similarly, TLV acknowledged the severe and life threatening condition, and the fact that patients' functional capacities are affected from having the disease. HAS also recognised the severe condition, and impact on daily activities.

Clinical experts in England highlight the fact that **no routine practice** for ITP exists, and that treatment pathways depend on the patient's characteristics. SMC also recognised the complexity of these treatment pathways. ITP is recognised as a **national priority** in France because it is an orphan drug ("Plan Maladies Rares, 2004). Finally, more flexibility was given to issues relating to the clinical evidence (lack of direct comparisons, short duration of trials) based on the fact that an important **public health need** was recognised in France (Table B-5).

Table B-**5.** "Other considerations" identified across the study countries

TREATMENT CHARACTERISTICS	NICE	SMC	TLV	HAS	Quotations
Type of therapeutic benefit					
• Curative				✓*	**HAS: "**eltrombopag is a curative treatment" (translated)
Innovativeness of treatment	✓C				**NICE**: "Innovative treatment that mimics the action of the hormone thrombopoietin and stimulates platelet release from the bone marrow"
• Novel mechanism of action	✓				**NICE**: "with romiplostim, eltrombopag represents a new approach to therapy"
• New class of drugs			✓*		**TLV**: "the drug belongs to a new class of drugs" (translated)
• Oral administration benefit		✓*	✓*		**SMC: "**benefit from oral administration" **TLV**: "eltrombopag is the only oral treatment available" (translated)
Adverse effects					
• Adverse effects similar across treatment arms	✓				**NICE:** "adverse events were similar between the eltrombopag and placebo groups in RAISE"
DISEASE CHARACTERISTICS					

Unmet need	✓ C				**NICE:** "greatest unmet need for treatment for patients with a low platelet count and persistent risk of uncontrolled bleeding, and that a sequence of treatments, in which eltrombopag might have an appropriate place, would be used to control the persistent risk of bleeding"
• No treatment alternatives exist		✓	✓*		**SMC**: "Oral corticosteroids and intravenous immunoglobulin are recommended as first-line treatments but the recommendations for second-line therapy and treatment of refractory patients are less clear. In these patients treatment should be tailored to suit the individual." **TLV**: "no satisfactory treatment options for patients with chronic refractory ITP exist (only romiplostim and eltrombopag)." (translated)
• Few treatment alternatives exist	✓		✓*	✓	**NICE:** "limited treatment options when conventional treatments fail to reduce risk of bleeding" **TLV:** "treatment options are limited" (translated) **HAS**: "limited therapeutic options available" (translated)
• Need for treatment options		✓*			**SMC**: "Eltrombopag will provide an additional treatment option"

Nature of the condition					
• Disease severity			✓*		**TLV**: "patients with severe disease may lead to severe haemorrhage" (translated)
• Serious condition				✓*	**HAS**: " ITS is a serious condition" (translated)
• Quality of life	✓ C, P		✓*	✓*	**NICE: "**affects quality of life" **HAS: "**important impact on quality of life of patients that have a risk of bleeding" (translated) **TLV:** "quality of life affected" (translated)
• Functional capacity affected			✓*		**TLV**: "functional capacity affected" (translated)
• Impact on daily activities	✓ C, P			✓	**NICE**: "bleeding and bruising can have significant impact on patients' daily activities " **HAS**: "daily activities are limited from having the disease" (translated)
• Social stigma	✓ C, P				**NICE**: "bruising can lead to social stigma"
• Limiting of life-style choices	✓ C, P				**NICE**: "bleeding can prevent or delay surgery and limit life style choices"
• Anxiety from risk of bleeding	✓				**NICE**: "anxiety about the risk of bleeding can also affect quality of life"
• Ability to work, travel or leisure	✓ C, P				**NICE:** "impact on person's ability to travel and take part

activities					in leisure activities”
• Life threatening	✓ C, P		✓*		**NICE: “**Spontaneous bleeding is an important but rare cause of premature death in people with ITP” **TLV**: “severe disease that may lead to death in the worst cases” (translated)
Rare disease characteristics					
• Orphan		✓*			**SMC**: “weaknesses in the clinical evidence were deemed acceptable partly because of the treatment’s orphan status”
• Rarity			✓*	✓*	**TLV:** “evidence is consistent and based on a relatively large number of patients given that ITP is rare” (translated) **HAS:** “minor public health burden because of the rarity of the condition” (translated)
Clinical practice/management					
• No standard treatment pathway used in routine practice	✓ C				**NICE**: “no single treatment pathway could be defined as routine practice”
• Unlicensed comparator	✓ C				**NICE:** “rituximab does not have a marketing authorisation for the treatment of chronic ITP”

• Complex and tailored to patient	✓C	✓ C			**NICE: "**the pathway for ITP would vary depending on the individual person's circumstances" **SMC**: "the approach of treatment has become more conservative and its initiation is based on clinical symptoms and whether invasive surgery is planned, in addition to platelet count"
• Adverse events of comparator treatments important	✓ C, P				**NICE: "**current treatment options may be associated with adverse effects"
• No long term evidence with comparative treatments	✓ C				**NICE**: "no long term safety data are currently available for people treated with rituximab"
• Correct dosage of treatment alternative unknown	✓				**NICE:** "correct dosage of rituximab is unknown"
• Preference for licensed over non-licensed	✓C				**NICE:** "preference to use licensed treatments before unlicensed options"
• Anxiety related to treatment alternatives	✓ P				**NICE**: "anxiety about the risk of contracting a hospital-acquired infection and the increased risk of infection following spleen removal, which requires life-long treatment"
National priority				✓*	**HAS**: "Plan Maladies Rares" (translated)

Legend: C: information provided from clinical experts; P: information provided from patient representatives; *: put forward as main reason for decision

Economic model uncertainties

A number of concerns relating to the economic modelling were raised, and are discussed here (Table B-6). In the NICE submission, the first model ("watch and rescue model") was rejected because of its high and uncertain cost-effectiveness. Concerns about the assumptions included: (1) the trial population not corresponding to the indication under review, (2) the assumption that differences in treatment arise because of bleeding events, (3) the assumption that the 26-week time horizon was similar in the future, and (4) the issues in the comparability of the evidence with romiplostim. Further, sensitivity analysis showed that the model was mainly driven by costs and not by clinical benefit, and multivariate exploratory analysis showed that there was a high degree of uncertainty in the model results. Consultation consultees suggested to consider the low budget impact of this drug, but the Committee responded that they do not account for this in their assessment. The challenges in modelling because of unknown clinical practice were also argued, though based on STA guidelines, the Committee does not give more flexibility to this type of scenario. The Committee did not account for the innovative nature of the treatment since no evidence demonstrating distinctive benefits was submitted.

For the long-term continuous model, similar issues were raised: (1) the trial population not corresponding to the indication under review, (2) the treatment sequences used not likely to reflect clinical practice, (3) the fact that it is uncertain where the place of eltrombopag in this sequence should be, and (4) the uncertainties around the relative effectiveness with comparator treatments. Further, there were large variations in the ICER with changes in the order of treatments with rituximab, and was driven by large variations in cost and little variations in health benefit. Consequently, the model was considered not to be valid given the assumptions and the cost per QALY that was greater than the acceptable 100 thousand pounds per QALY.

Finally, the model comparing eltrombopag to romiplostim was considered not to be relevant to the decision problem being considered, and as such rejected. This was mainly because of a number of issues raised with the model and the fact that

romiplostim was under review with NICE and therefore not routinely used in clinical practice.

In the model appraised by SMC (cost-utility model comparing eltrombopag with romiplostim), the nature of the indirect comparison led to uncertainties in the drug's relative effectiveness. Despite this, the treatment was considered cost-effective given the benefit from oral administration and its orphan status.

TLV considered a cost-minimisation analysis comparing eltrombopag with romiplostim. Eltrombopag was considered cost-effective, given that clinical benefits in both drugs were deemed comparable, and that the cost of treatment for eltrombopag was lower than romiplostim, which was already previously considered cost-effective for TLV.

Table B-**6.** Economic models and final ICER considered for decision, clinical uncertainty

	NICE	**SMC**	**TLV**
Watch and rescue model (eltrombopag versus standard care)	**Cost-utility model** Splenectomised: £104,100/QALY Non-splenectomised: £116,800/QALY		
Short duration of trial (RAISE)	×*		
Low benefit for populations in the lower incidence of the most severe bleeding events	×*		
Lack of adverse event data for comparator treatments	×		
Assumption made that differences between treatments arise because of bleeding events	×*		
Sensitivity analysis suggested that the model driven by costs and very small difference in effect	×*		
Exploratory multivariate analysis: high degree of uncertainty in the model results	×		

Long-term continuous treatment model	**Cost-utility model** Not reported		
Suitability of trial population for modelled population	×		
No survival benefit across treatment sequences	×		
Sequences presented do not reflect clinical practice, according to clinical specialists	×*		
Eltrombopag versus romiplostim	**Cost-utility model** Not reported	**Cost-utility model** Splenectomised: Savings £12,641 and 0.039 QALY gain (dominant) Non-splenectomised: Savings £2,094 and 0.028 QALY gain (dominant)	**Cost-minimisation model** Similar benefit to romiplostim at lower cost
Eltrombopag was superior to romiplostim using this crude comparison of bleeding events, but there is no robust evidence to support this assumption as a formal indirect comparison based on this outcome was not conducted		×	
Choice of clinical endpoint: bleeds instead of primary endpoint platelet response		×	
Sensitivity analysis showed no difference with romiplostim		×	
The relative risk of bleeding events would have to fall substantially from the base case values before eltrombopag would no longer be considered cost-effective		×	
Budget impact	No impact	No impact	

Legend: × uncertainty not addressed; * uncertainty put forward as one of the main reasons for the final decision.

Final HTA recommendation

The reasons for the HTA recommendations issued are summarised in Table B-7, and includes the overall conclusion about the treatment's clinical effectiveness and main reasons for this, as well as the conclusion about the treatment's cost-effectiveness and main reasons for this.

Table B-7. Conclusions about clinical effectiveness and cost-effectiveness

HTA body	**NICE**	**SMC**	**TLV**	**HAS**
Recommendation	DNL	LWC	L	ASMR II - LWC
Clinical effectiveness	×*	✓*	✓*	✓*
Main reasons for clinical effectiveness assessment	Platelet response was better for eltrombopag than placebo, but results were deemed uncertain because of the long-term effect	Although clinical evidence weak, significantly more effective than placebo in raising and maintaining platelet count – benefit from oral administration and orphan status.	Indirect comparison shows that drugs are likely to have similar effects, though small patient population, it is acceptable because of orphan status. Follow-up on clinical effectiveness and uncertainties in October 2013.	Limited trial duration and lack of comparative data. Benefit deemed similar to romiplostim until further evidence provided Risk assessment plan to monitor toxicity.
Cost-	×*	✓*	✓	NA

effectiveness			
Main reasons for cost-effectiveness assessment	Watch and rescue model: uncertain effects = unacceptable cost/QALY > £100k Continuous model: trial population not generalizable to ITT population, treatment sequence not consistent with clinical practice	Uncertainties in clinical evidence addressed as per above	Similar benefit to romiplostim at lower price (comparator having been assessed as cost-effective).

×*: clinical effectiveness uncertain and not demonstrated

✓: improved or positively assessed clinical effectiveness

Notes: HAS issued an ASMR II rating because it is deemed similar to romiplostim in terms of the expected public health benefit, although this expected is considered to be weak because of lack of evidence.

Appendix C. Interview Topic Guide

TOPIC GUIDE FOR INTERVIEWS

Haute Autorité de Santé

The objective of these interviews is to ask a number of open-ended questions around some of the findings from an on-going research project, which aimed: (a) to identify the criteria underlying the HTA opinions for a number of orphan drugs, and (b) to understand why, for a same drug and indication, different HTA bodies issued different opinions. This is to ensure that the interpretation of the researcher when analysing the opinions was accurate, particularly in cases when differences across countries were one of the explanatory factors for different HTA outcomes, to correct any misinterpretations and to obtain further insights into the drivers of these opinions in each agency. This research is being undertaken in the context of Advance_HTA, a project funded by the European Commission's Research Framework Programme (FP7). http://www.advance-hta.eu

Findings were derived from a mixed methods research project that enabled (1) the in-depth analysis of the decision processes through case study analyses, and (2) the generalisation of findings across all case studies. In the first stage, a case study template was set up and used for each drug-indication pair, from which the criteria driving the HTA opinions were identified and extracted from the HTA reports (publicly available on the agencies' websites). On the basis of this case study, thematic analysis was conducted to code the HTA reports according to a previously established coding manual listing the criteria driving these opinions. This allowed for the comparison across cases, which helped us to better understand how different agencies conduct value assessments as well as the reasons for differences across countries. The HTA bodies included are: the National Institute for Health and Clinical Excellence (NICE) in England, Scottish Medicines Consortium (SMC) in Scotland, Pharmaceutical and Dental Benefits Consortium (TLV) in Sweden, Haute Autorité de

Santé (HAS)[9] in France. Ten drug and indication pairs were included in the study (Appendix C-1).

Interviews will take place face-to-face (or if no convenient dates are found by skype or by phone), and may last approximately one to two hours, depending on your availability. Interviewees are competent HTA body representatives, which may include the Head of one of the Committees making the recommendations, a clinician Committee member, a pharmacist or health economist. The interview topic guide consists of this document and will be used as a basis for the discussion. Interviewees may send their responses in writing prior to the interview if desired. The interview questions are divided into four general themes: (a) General evidentiary requirements for orphan drugs; (2) Other considerations; (3) Dealing with uncertainty; (4) Stakeholder involvement. Each theme discusses a number of related issues that were derived from cases when differences were seen across countries, or when it was unclear how certain of the identified criteria influenced the opinions.

These interviews are a way forward to furthering the debate about a number of HTA methodological issues, while simultaneously raising awareness about the types of differences that are seen across countries, which may also be applicable within countries when more than one decision-making body exists.

Thank you for your participation.

---

[9] The analysis focuses on the Opinion issued by the Transparency Committee, and its conclusions about the assessment of actual benefit ("Service Médical Rendu (SMR)"), improvement in actual benefit ("Amélioration du Service Médical Rendu (ASMR)"), and final recommendations.

GENERAL EVIDENTIARY REQUIREMENTS

This section seeks to understand what scientific evidence is required, and to what extent it is different for orphan drugs compared to other drugs.

1. Primary evidence

For 8 of the 10 study drugs (Appendix C-1), the primary evidence considered was phase III trials; for the remaining two study drugs, it was phase II trials (e.g. trabectedin and ofatumumab). The latter cases were due to the early marketing authorisations granted by the EMA, and therefore the manufacturers were not required to conduct any RCTs. Additionally for ofatumumab, results relied on an interim analysis of subgroup data (Table C-1).

For trabectedin, the primary endpoint "time-to-progression" was significantly improved. Similarly for ofatumumab, the primary endpoint "response rate" was also significantly improved.

Table C-1. Phase II primary trials & recommendations

Trabectedin ***Soft tissue sarcoma***				**Ofatumumab** ***Chronic lymphocytic leukaemia***			
STS-201: phase II randomised trial (N 270) testing trabectedin at different doses (cross overs permitted)				Subgroup analysis of Hx-CD20-406 including patients refractory to ludarabine and alemtuzumab (N 59) Hx-CD20-406: prospective, non-randomised, non-comparative phase II trial (N 154) testing different groups of patients refractory to different therapies or non-refractory (3 arms).			
NICE	SMC	TLV	HAS	NICE	SMC	TLV	HAS
Restricted	Rejected	NA	SMR Important ASMR V	Rejected	Rejected	NA	SMR moderate ASMRV

All HTA bodies were concerned about the trial designs, including the lack of comparative evidence, but this was perceived in different ways (Table C-2). Additionally, for ofatumumab, all HTA bodies deemed the nature of the subgroup data insufficient to demonstrate the drug's clinical benefit (Table 2).

Table C-2. How the lack of comparative data was perceived by the HTA bodies (✓ acceptable, or ✗ not acceptable)

	NICE	SMC	HAS
Trabectedin	✔ rarity of the condition ✔ investigational nature ✔ registry data as historical controls ✔LWC (discount)	✔ rarity of the condition ✔ investigational nature ✗ DNL	✗ therapeutic contribution unknown given the lack of comparative evidence (ASMR V) ✔ SMR important (no validated alternative drug)

Ofatumumab	✗ non-comparative study ✗ difficulties in conducting RCTs could have been better addressed, where methylprednisone could have been used as a comparator ✗ only interim results were available ✗ small patient numbers => uncertain size of effect in the absence of robust comparative evidence and immaturity of the data ✗ DNL	✗ non-comparative study ✗ only interim results were available ✗ small patient numbers => the robustness of the response shown in this study and its ability to be translated into a clinical benefit is uncertain, due to the small patient numbers and interim nature of the analysis ✗ DNL	✗ non-comparative study ✗ only interim results were available => efficacy hard to assess in view of the methods used. Limited clinical data obtained from the interim analysis ✔ Provisional while awaiting complementary data: SMR moderate (no alternative drug) ✔ Provisional ASMR V while awaiting complementary data

Questions:

- Are phase III trials preferred as primary evidence for the appraisals (e.g. as seen in the results)?

- To what extent is evidence from phase II trials, subgroup analyses or ATU usage acceptable as primary evidence (e.g. criteria for acceptability? e.g. trabectedin and ofatumumab)?

- To what extent are indirect comparisons acceptable? (e.g. preference for indirect comparisons if direct comparative evidence weak? Define types of weaknesses)

- Is it common that the comparators included in the assessment are not deemed the appropriate ones?

- To what extent can registry data be acceptable to proxy historical controls when no comparative data is available (e.g. as seen in the NICE appraisal for trabectedin)?

- Sir Michael Rawlins discusses the criteria for acceptability of historical controls.[10] Would these be acceptable for your HTA agency and this approach in general for non-phase III evidence? Please explain.

2. Non-primary evidence

A number of non-primary non-phase III trials were reported in the HTA reports for the 10 study drugs. These, however, had very little influence on the final outcome. Outcomes from these trials were generally not reported, and when reported, the type of data provided was around safety (e.g. romiplostim, ofatumumab, eltrombopag), dosage research (e.g. eltrombopag) and historical controls (e.g. trabectedin).

Questions:

[10] In: Rawlins, M. (2008). De Testimonio: on the evidence for decisions about the use of therapeutic interventions. Clinical Medicine, 8, 579-588.
In the context of HIV trials, the following requirements should be met for historical controls:

- "there must be no other treatment appropriate to use as a control;
- there must be sufficient experience to ensure that the patients not receiving treatment will have a uniformly poor prognosis;
- the therapy must not be expected to have substantial side-effects that would compromise the potential benefit to the patient;
- there must be a justifiable expectation that the potential benefit to the patient will be sufficiently large to make interpretation of the results of a non-randomised trial unambiguous;
- and, the scientific rationale for the treatment must be sufficiently strong that a positive result would be widely accepted."

Cited from: Byar DP, Schoenfield DA, Green SB, et al. Design considerations for AIDs trials. N Engl J Med 1990; 323: 1343–48.

- ❖ How is non-primary evidence accounted for during the appraisal process, what type of information is provided, and are there any criteria in accepting this type of evidence (e.g. promise of another study)?

3. Trial length

In two cases the trial duration of the primary trials were deemed too short by some of the HTA bodies.

- For imatinib, this was an identified concern for NICE and SMC (negatively influencing the opinion), but was not mentioned in the HAS recommendation. The drug was not appraised by TLV.
- This issue was also raised by all agencies for eltrombopag. For HAS, eltrombopag received the same ASMR II rating as romiplostim (Nplate). For the others, this had a negative influence on the opinions, where additionally, it was also confirmed as not long enough by clinical experts at NICE.

Questions:

- ❖ How is the appropriate trial duration determined, and under what circumstances could a trial considered too short be deemed acceptable (e.g. same length as comparator)?

- ❖ Given that the appropriate trial length is discussed for the marketing authorisation (MA) stage, how are these MA determinants perceived within the HTA decision (e.g. degree of reliance, any differences in perspectives)?

4. Primary endpoints

Eltrombopag was appraised in all four countries based on the pivotal trial RAISE (phase III placebo-controlled RCT) and received diverging recommendations by the study countries (Table C-3).

Table C-3. HTA recommendations

NICE	SMC	TLV	HAS
☒ REJECT	☑ RESTRICT to patients with severe ITP and high risk of bleeding	☑ RESTRICT for hospital use & reassessment in 2013	☑ SMR important ASMR II

The primary endpoint was "platelet response" and secondary endpoints included "WHO bleeding events" of different grades of severity (Table C-4).

Table C-4. Bleeding events and their statistical significance in RAISE (pivotal trial for eltrombopag).

	NICE	SMC	TLV	HAS
WHO bleeding event 1-4 = *all events*	☑	☑	☑	☑
WHO bleeding event 2-4 = *clinically significant events*	☑	☑		
WHO bleeding event 3-4 = *gross and debilitating blood loss*	☒			

Legend: ☑ significantly improved; ☒ non-significantly improved.

One of the reasons for rejecting the drug by NICE was that no significant differences between treatment groups in the low incidence of the most serious bleeding events ("WHO bleeding events grades 3 and 4").

Questions:

- How is the relevant endpoint identified (e.g. trial's primary endpoint, proxy for survival, most relevant to the patient)?

- The CT minutes indicate that the members discussed the existence of supplementary data concerning the effect of Revolade on bleeding compared with Nplate. Given that both drugs were available under ATUs, one could speculate whether the outside expert had additional data from clinical use in France. Is this data and/or clinical experience with drugs based on ATUs considered in the evaluation process?

5. "Overall survival" versus progression-free survival

A preference for "overall survival" was seen in a NICE appraisal (e.g. mifamurtide), where it was considered as the endpoint of interest despite "progression-free survival" (PFS) being the primary endpoint in the trial. This had a negative influence in the opinion given it was not significantly improved (compared to PFS which was significantly improved positively influencing the assessments in other countries).

In another case (e.g. imatinib), the secondary endpoint "overall survival" was not significantly improved. This was one of the main reasons negatively influencing the opinion for NICE. In contrast, for SMC and HAS, this was also raised as a concern, but was acceptable for SMC given the orphan status of the treatment, and for HAS given an on-going trial collecting additional data.

Questions:

- Does the CT have a preference for overall survival over progression-free survival, even in cases when overall survival is a secondary endpoint in the trial?

- What criteria make that an uncertain clinical outcome is acceptable (e.g. on-going trial, rarity)?

6. Surrogate endpoints

The primary endpoints in 8 out of 10 cases were surrogate (e.g. substitute for the clinical endpoint of interest), which were predominantly validated with the exception

of “time-to-progression” for soft tissue sarcoma (trabectedin) and “platelet response” for ITP (eltrombopag, romiplostim).

Questions:

- Under what criteria for accepting surrogate endpoints (e.g. validated, when the relationship between surrogate and clinical endpoint is causal), which is a biomarker intended to substitute a clinical endpoint, the latter being ‘a characteristic or variable that reflects how a patient feels, functions, or survives’?

- Are surrogate (validated or non-validated) endpoints more acceptable for orphan indications?

PUBLIC HEALTH EVIDENCE

This section seeks to understand how other evidence and considerations have influenced the decision processes, and how.

7. Quality of life data

Quality of life data was not present in 6 out of 10 cases, and in three other cases, it was only present in the assessments by SMC and NICE (eltrombopag, romiplostim and everolimus).

Questions:

- What type of quality of life data is accepted (e.g. from trials, registries, expert opinion, etc.), and preferred?

- What are the implications for assessments when quality of life data is missing?

- How is quality of life data from different forms of evidence accounted for in the analysis of the public health value (intérêt de santé publique)?

8. Innovativeness of the technology

This research identified a number of instances where the treatment's innovativeness was highlighted during the assessment (Appendix C-2). No clear or uniform definition exists of the determinants of an innovative treatment. Identifying cases where a treatment was recognised as innovative in practice may be a way forward to identifying these determinants.

Questions:

- Appendix C-2 lists the quotations coded from the HTA reports as recognising an "innovation". Do you agree with what was coded as pertaining to the innovative nature of the treatment in your country? Any comments on how this was perceived in the other countries?

- How does the orphan status of the drug reflect on innovation, and how does this impact the assessment and the ASMR rating? Does this change depending on whether the drug is subject to fast-track consideration or to an ATU?

- What are the criteria for recognising a technology as innovative (e.g. patient benefit, new class of drugs, etc.)? Is administration benefit considered an innovation?

9. Unmet need

Unmet need in a given disease area may be considered among the determinants of a better ASMR (France) or price (UK). This research identified a number of instances where the disease's unmet need was highlighted during the assessment (Appendix C-3). Again, no clear or uniform definition exists of the determinants unmet need and one way of defining these could be by identifying previous cases.

Questions:

- Appendix C-3 lists the quotations coded from the HTA reports as recognising an "unmet need". Do you agree with what was coded as referring to an "unmet need" in your country? Any comments on how this was perceived in the other countries?

- What are the criteria for recognising the "unmet need" of a disease (e.g. no treatments, few treatments, no curative treatment, etc.) and how do these influence the HTA evaluation of a drug meeting that need?

- In Sweden, unmet need is contextualised by the degree of severity (e.g. the severity of a condition is defined by the consequences to the patient without the treatment under review), what do you think about that?

- We assume that few comparators will be translated into a better ASMR. How is unmet need accounted for and could there be double counting if accounted for separately from the ASMR.

10. Disease severity

A number of determinants of severity exist, which include the impact on quality of life and mobility, or considerations of life expectancy.[11] There are two ways to include severity into HTA: through a weighing of the QALY (or other measures of HTA), or during the deliberative process of HTA.[12] The second method would apply to our study countries as no specific weighing for severity in the outcome of HTA is seen.

---

[11] Dolan P, Shaw R. A note on a discussion group study of public preferences regarding priorities in the allocation of donor kidneys. Health Policy. 2004; 68: 31-36.

[12] Garau M, Shah K, Towse A, et al. Assessment and Appraisal of Oncology Medicines: NICE's Approach and International HTA Experience. Report for the Pharmaceutical Oncology Initiative Group (POI), 2009.

This research identified a number of instances where reference to the severity of the disease was made during the assessment.

Questions:

- Appendix C-4 lists the quotations coded from the HTA reports as recognising "disease severity". Do you agree with what was coded as pertaining to disease severity in your country? Any comments on how this was perceived in the other countries?

- What are the criteria for recognising a disease as severe (e.g. life-threatening) and how do these influence the HTA outcome?

- Are end-of-life treatments considered differently, and potentially qualify as severe cases (e.g. NICE end-of-life treatment)?

11. Consistency across decision

- Would it be conceivable for findings of innovativeness, unmet need and disease severity in the recommendations for the study drugs to constitute a precedent for how these matters are assessed in future recommendations?

- In the appeals, has the consistency of considering these factors been a concern, and if yes, how have they been addressed?

12. Public health value (ISP) and SMR

- How are the following factors, accounted for in assessing the SMR, used to assign a higher or lower SMR?

  - efficacy & adverse events

  - place in the therapeutic strategy

- disease severity
- preventive curative or symptomatic treatment
- public health value (intérêt santé publique)

❖ What is the weight of a public health value (ISP) in determining a drug's SMR? Could the finding of an ISP (important, moderate, weak) increase the SMR of a drug that otherwise would have a lower SMR based on its clinical benefit? Is there a framework with implicit weights for each criteria?

❖ How are the following factors, accounted for in assessing the ISP, used to assign a higher or lower ISP?
  - Public health need
  - Impact of the treatment on the population's health
  - Impact on the organisation of the health system
  - The particular ministerial plan under which the disease is designated as a priority

❖ It appears that there were insufficient data to support a finding of public health value (ISP) for a number of the drugs. Would better trial data on outcomes reflecting the impact of the treatment on the allocation of health system resources be valuable in the deliberative process for orphan drugs?

DEALING WITH UNCERTAINTY

We know that orphan indications are more difficult to appraise compared to more common indications, given it is more difficult to produce robust evidence due to the small sample sizes and heterogeneity of these conditions. This, together with the high prices of these drugs to recoup R&D investments, results in orphan drugs being hardly ever cost-effective. The question is whether more flexibility is given to accepting certain types of uncertainty when other types of evidence are presented, which is what we will try to understand in this section.

In a few cases, clinical uncertainties were deemed acceptable given the rarity of the condition:

For eltrombopag, given it is an orphan drug destined to treat rare disease, the number of trial participants is considered large by TLV. Quotation (translated): "The clinical studies of Revolade's clinical efficacy demonstrates consistent data and relatively large number of patients included in the studies, given that ITP is a rare disease."

For trabectedin, NICE and SMC accepted that comparative evidence was limited given it is a drug that treats a rare condition. Quotation: "The Committee appreciated that because soft tissue sarcoma is a rare condition, the evidence for the comparative effectiveness of trabectedin was limited."

For imatinib, SMC accepted the uncertain survival benefit given the rarity of the condition treated. Quotation: "Although there were some limitations in the economic analysis in terms of the likely estimate of overall survival benefit, the economic case was considered to be demonstrated when the SMC modifiers, in particular those relating to medicines for orphan diseases, and the anticipated survival benefit associated with imatinib, were applied."

Question:

- To what extent are certain types of uncertainties more acceptable given the rarity of the condition?

- Do "other considerations" compensate for greater uncertainty for orphan drugs? Does a finding of ISP, even weak, compensate for clinical uncertainty?

STAKEHOLDER INVOLVEMENT

Who has input into the assessment process other than the manufacturer and the HTA body members?

Stakeholder input was mainly identified in the NICE appraisals, and in a few instances in SMC's and is a direct result of the formal processes that exist (e.g. Public Involvement Programme (PIP) at NICE and the Patient and Public Involvement Group (PAPIG) at SMC). No formal process for stakeholder input exists at TLV. At HAS, participation of outside experts within the Transparency Committee was identified for all study drugs in the meeting minutes (procès-verbal), but the identification of the expert and the content of his/her advice was not recorded.

The main type of information provided from patient and clinical experts was about "other considerations" and about clinical uncertainties.

Questions:

- What type of input do clinical experts and patient experts provide (e.g. uncertainties, clinical practice, evidence, assumptions, nature of the disease, living with the disease)?

- How is their input accounted for (e.g. address uncertainty, participation in the deliberative process)?

- Do clinical specialists have different conflict of interest requirements when dealing with small patient populations and the scarcity of specialists?

Appendix C-1: Study drugs & countries

Generic name/Brand name	Indication	ICD10 code[1]	NICE England	SMC Scotland	TLV Sweden	HAS France[2]
Eltrombopag REVOLADE	Thrombocytopenic purpura	D2	DNL TA205 Oct 2010	LWC 625/10 July 2010	LWC 3731/2010 May 2011	SMR important ASMR II June 2010
Romiplostim NPLATE	Chronic idiopathic thrombocytopenic purpura	D2	LWC TA221 Apr 2011	LWC 553/09 May 2009	LWC 833/2010 Oct 2010	SMR important ASMR II June 2009
Everolimus AFINITOR	Renal cell carcinoma (2nd line, advanced)	C	DNL TA219 Apr 2011	DNL 595/10 Mar 2010	L 853/2010 Sep 2010	SMR important ASMR IV Jan 2010
Lenalidomide REVLIMID	Multiple myeloma (3rd line)	C	LWC TA171 Jun 2009	LWC 441/08 Apr 2010	L 410/2010 Jul 2010	SMR important ASMR III Oct 2007
Azacitidine VIDAZA	Myelodysplastic syndrome	D1	LWC TA218 Mar 2011	LWC 589/09 Aug 2011	NA	SMR important ASMR II Apr 2009
Imatinib GLIVEC	Gastro intestinal stromal tumour (adjuvant, after surgery)	C	DNL TA196 Aug 2010	LWC 584/09 Aug 2010	NA	SMR important ASMR III Sep 2009
Mannitol dry BRONCHITOL	Cystic fibrosis	E	LWC TA266 Nov 2012	DNL 837/13 Jan 2013	NA	SMR weak ASMR V Sep 2012
Mifamurtide MEPACT	Osteosarcoma	C	LWC TA235 Oct 2011	L 837/13 Jan 2013	NA	SMR insufficient DNL Nov 2010
Ofatumumab ARZERRA	Chronic lymphocytic leukemia	C	DNL TA202 Oct 2010	DNL 626/10 Jul 2010	NA	SMR moderate

						ASMR V Oct 2010
Trabectedin YONDELIS	Soft tissue sarcoma	C	LWC TA185 Feb 2010	DNL 452/08 Jun 2011	NA	SMR important ASMR V Apr 2008

Legend: NICE: National Institute for Health and Care Excellence (NICE); SMC: Scottish Medicines Consortium; TLV: Dental and Pharmaceutical Benefits Board; HAS: Haute Autorité de Santé; L: list; LWC: list with restrictions; DNL: do not list; NA: not applicable.

Appendix C-2. Quotations coded as an "innovation" in the HTA reports

	NICE	SMC	TLV quotations translated	HAS quotations translated
Eltrombopag	**Innovative**: "clinical experts say that it is an innovative treatment that mimics the action of a hormone…" **New class of drugs:** "represents a new approach to therapy"	**Oral administration:** "benefit of oral administration recognised"	**Oral administration:** "the only treatment in that class that can be taken orally"	**First in class and oral administration:** "the first oral thrombopotein receptor agonist"
Romiplostim	**Adverse event profile:** "it has a good adverse-effect profile, particularly in comparison with currently available treatments" **Improved clinical** benefit: "The Committee heard from the clinical specialists that romiplostim may have benefits over other active treatments because it produces a sustained platelet response during " **New mechanism of action:** "The Committee concluded that romiplostim has a novel mechanism of action " **Step change:** "romiplostim represents a step change for the treatment of ITP"		**New class of drugs:** "considered a new class of drugs" + "it is a new principle for the treatment of ITP"	
Everolimus				
Lenalidomide	**Important advance:** "lenalidomide is an important advance in the treatment of multiple myeloma and could be considered as an alternative to bortezomib"			**Provision:** "lenalidomide could decrease the use of the health system given its administration mode, although no evidence was available to support this"
Azacitidine	**Important advance:** "Committee recognised that azacitidine represents an important change in the treatment of patients with myelodysplastic syndromes, noting the substantial benefits associated with its use"			
Imatinib		**Important advance:** "imatinib treatment could represent a significant advance in therapy in patients who are at high risk of tumour recurrence."		
Mannitol dry	**Administration:** "treatment burden is substantially less for an inhaler than for a nebuliser according to clinical specialists and patient experts….. The Committee concluded that it provided practical advantages over treatment with nebulisers, but mannitol as an add-on therapy would not replace the use of nebulisers, and so could not be considered a step-change in treatment."	**Administration:** "it is administered as a dry powder inhalation, so may offer an important advantage over alternative treatments which require nebulisation."		
Mifamurtide	**Important advance:** "it is considered a significant innovation" **New mechanism of action**: "the mechanism of action was novel" **Potential valuable new therapy:** "it is a potentially valuable new therapy"			
Ofatumumab	**New mechanism of action:** "accordin to clinical experts, it may offer a slightly different mechanism of action because it targets a different epitope.			
Trabectedin				

Appendix C-3. Quotations coded as an “unmet need” in the HTA reports

	NICE	SMC	TLV quotations translated	HAS quotations translated
Eltrombopag	**Unmet need:** "..The Committee heard that patients with a low platelet count with a persistent risk of uncontrolled bleeding have the greatest unmet need for treatment..." **Few treatment alternatives exist:** "The Committee understood that options for treatment of chronic ITP are limited when conventional treatments fail to reduce the risk of bleeding"	**No treatment alternatives exist**: "The guideline states that there is no indication for treatment in adults in whom there are no signs and symptoms and platelets are greater than 30x109/L" **Need for treatment options**: "economic case was considered to be demonstrated as eltrombopag would offer an additional treatment option "	**No treatment alternatives exist:** "No satisfactory treatment options for patients with chronic refractory ITP. " **Few treatment alternatives exist:** "For these patients, treatment options today are limited."	**Few treatment alternatives exist:** "few therapeutic alternatives exist"
Romiplostim	**Unmet need:** "Clinical experts advised that that there is significant unmet need" **Few treatment alternatives**: "The Committee understood that there are few treatments licensed for the treatment of chronic ITP"	**Few treatment alternatives:** "limited therapeutic alternatives are available." **Need for treatment options**: "romiplostim would be a valuable addition to the range of available treatment options for ITP"	**Few treatment alternatives:** "limited therapeutic options"	**No treatment alternatives exist:** "There is no validated therapeutic alternative in these clinical situations."
Everolimus	**Few treatment alternatives exist:** "The Committee heard from clinical specialists and patient experts that there are limited treatment options for people with advanced... " **No treatment alternatives:** "no alternative medicinal products exist"			**No treatment alternatives exist:** "No alternative medicinal products exist"
Lenalidomide	**Need for treatment options:** "The Committee noted the importance that patients, their carers and physicians placed on having effective options to treat multiple myeloma at presentation and at subsequent relapses." **No alternatives with similar benefits**: "alternatives not routinely available on the NHS"			**Treatment alternatives exist**: "medicinal or non-medicinal treatment alternatives exist"
Azacitidine		**No alternatives with similar benefit: "**first medicine to be licensed specifically for the treatment of primary MDS"		
Imatinib	**No treatment alternatives:** "there are currently no adjuvant therapies available for people following resection of a GIST, and that watchful waiting is the current standard of care"	**No treatment alternatives exist:** "The current gold standard of treatment in primary resectable GIST is surgery with gross margin resection and there is no accepted adjuvant treatment."		**No treatment alternatives:** "there is no alternative medication"
Mannitol dry	**Unmet need:** "The Committee concluded that there is an unmet clinical need in patients with rapidly declining lung function, particularly if there are no other therapies appropriate to offer the patient." **Need for treatment options:** "the importance of treatment options for people with cystic fibrosis who have few alternative options."	**Unmet need: "**The company has suggested that there is an unmet need for an effective treatment in patients who have failed to respond to, or are intolerant of, other treatments" **Need for treatment options:** "high unmet need for effective therapies."		**Treatment alternatives exist**: "therapeutic alternatives exist"
Mifamurtide	**Few treatment alternatives exist:** "patient experts and clinical specialists stated that there had been few developments that had improved treatment outcomes for osteosarcoma over the past 20 years"			**Treatment alternatives exist**: "therapeutic alternatives exist"
Ofatumumab	**No treatment alternatives:** "lack of treatment options for patients at this stage of disease" **Few treatment alternatives:** "limited treatment options" **Need for treatment options: "** clinical specialists stated that it was important to have additional treatment options, such as ofatumumab, later in the treatment pathway"	**No treatment alternatives exist: "**given the absence of any alternative treatment validated in extensively pretreated patients who are refractory to fludarabine and alemtuzumab...."		**No treatment alternatives exist:** "absence of validated treatment alternatives" **Public health need:** "improving the therapeutic management of chronic lymphcytoctic leukemia is a public health need"
Trabectedin	**No alternatives with similar benefit:** "that treatment with trabectedin represents an option for those patients who would otherwise have no licensed treatment options."	**Unmet need:** "clinical experts advised that there is unmet need for effective second line treatments in patients with advanced soft tissue sarcoma who have good performance status."		**No treatment alternatives:** "no validated alternative drug treatment at this stage of the disease"

Appendix C-4. Quotations around disease severity and the impact of living with the disease on the patient & surrounding

	NICE	SMC	TLV quotations translated	HAS quotations translated
Eltrombopag	**Severity:** "severe disease may lead to severe hemorrhage" **Day-to-day activities:** "the risk of bleeding affects a person's ability to travel and take part in leisure activities" + "bleeding and bruising can have considerable impact on the daily activities of people with chronic ITP" **Quality of life:** "anxiety about the risk of bleeding can affect quality of life" **Life threatening:** "Spontaneous bleeding is an important but rare cause of premature death in people with ITP" **Limiting life style choices:** "bleeding limits lifestyle choices" **Social stigma:** "bruising may result in social stigma"		**Quality of life:** "quality of life is affected" **Functional capacity:** "functional capacity is affected" **Life treatening:** "worse case, bleeding may lead to death"	**Day-to-day activities:** "risk of bleeding affects daily activities" **Quality of life:** "ITP affects the quality of life of patients' with a risk of bleeding" **Serious condition**
Romiplostim	**Day-to-day activities:** "the risk of bleeding affects a person's ability to work, travel and undertake leisure activities" +"bleeding and bruising can have considerable impact on the daily activities of patients" **Quality of life:** "the risk of bleeding affects a person's quality of life" **Anxiety of symptoms:** "anxiety about the risk of bleeding" **Fatigue:** "many patients experience fatigue, but there is no clear relationship between fatigue and platelet count." **Limiting life style choices:** "bleeding limits lifestyle choices" **Need for rescue treatments and hospitalisations:** "patient experts sated that a bleed could result in a person seeking medical care to receive rescue therapies, and if the bleeding was severe the person could need hospitalisation" **Social stigma:** "bruising may result in social stigma"	**Life threatening**	**Severity:** given the high disease severity, the cost per output is reasonable" **Quality of life:** "quality of life is affected" **Functional capacity:** "functional capacity is affected" **Life threatening:** "can lead to severe hemorrhage and in worst case, death" **Chronic condition**	**Day-to-day activities:** "risk of bleeding forces patients to limit their activities" **Quality of life:** "severe impact on the quality of life of patients" **Serious condition**
Everolimus	**Severity:** "the views from patient and clinical experts on the severity of the disease" **Short life expectancy:** "potentially as low as 5 months"		**Severity:** "used to treat disease with high severity" **Late diagnosis:** "cancer often diagnosed at advanced stages given that symptoms are diffuse" **Life threatening:** "RCC is the 6th most common cancer diagnosis leading to death" **Short life expectancy:** "median survival 8-9 months"	**Life threatening:** potentially life-threatening condition" **Serious condition**
Lenalidomide	**Severity:** " **Incurable:** "multiple myeloma is an incurable disease" **Short life expectancy:** "potentially as low as 9 months"		**Incurable:** "the is no curative treatment" **Life threatening:** "approximately 25% of patients survive after 5 years of diagnosis" **Short life expectancy:** "the risk of early death for patients with multiple myeloma is great"	**Life threatening:** "multiple myeloma is almost always fatal" **Short life expectancy:** "short median survival (3 to 5 years)"
Azacitidine	**Day-to-day activities:** "ability to carry out day-to-day activities reduced" **Quality of life:** "symptoms from the disease have a negative impact on the patient's quality of life" **Fatigue:** "fatigue is common and has a negative impact on quality of life" **Need for rescue treatments and hospitalisations:** "patient groups reported that dependence on blood transfusions is an important aspect of these conditions and also has a negative impact on quality of life" **Short life expectancy:** "life expectancy with best supportive care alone was likely to be approximately 11.5 months"	**Life threatening disease**		**Poor prognosis** **Short life expectancy:** "median survival between 0.4 and 1.2 years"
Imatinib		**High risk of relapse or recurrence**		**Life threatening:** "GIST are serious life-threatening conditions" **Serious condition**
Mannitol dry	**Severity:** considered severe **Life expectancy:** "cystic fibrosis leads to early mortality" **Burden to carers:** in terms of assisting with treatment and helping patients to maintain normal lives" **Morbidity:** "cystic fibrosis leads to considerable morbidity according to clinical specialists" **Quality of life percieved as normal:** as patients have always lived with the disease	**Life expectancy:** "cystic fibrosis is life-limiting" **Incurable** **Life long treatment**	**Severity:** "disease considered to have a very high severity"	**Severity:** "severe disease" **Life expectancy:** "life expectancy is affected by the respiratory burden" **Quality of life:** "quality of life substantially affected" **Incurable** **Life long treatment required**
Mifamurtide	**Disruption of family life** **Financial pressures** **Negative effects on health of surroundings:** "negative effect on family and friends" + "Committee heard from patient experts that supporting patient has a profound impact on the healthrelated quality of life of the family and friends of the person affected, particularly when treatment is not successful. For example, parents and siblings may develop mental health problems and family relationships may be strained" **Strain on family relationships** **Stress at work**			**Life threatening:** "affects patient's vital prognosis"
Ofatumumab	**Poor prognosis:** "patients with double refractory chornic chronic lymphocytic leukaemia have a particularly poor prognosis, as per discussions with patient and clinical experts" **Short life expectancy:** "limited life expentancy" + "potentially as low as 8 months"	**Chronic condition** **Incurable**		**Life threatening**
Trabectedin	**Short life expectancy:** "approximately 6 months"	**Poor prognosis:** "outcomes in advanced soft tissue sarcoma are poor"		**Life threatening:** "affects patient's vital prognosis" **Serious condition**

Appendix D. Ethics Approval