Elena Nicod, Panos Kanavos

Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: a mixed methods study

Article (Accepted version) (Refereed)

Original citation:
Nicod, Elena and Kanavos, Panos (2016) *Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: a mixed methods study.* Health Policy, 120 (1). pp. 35-45. ISSN 0168-8510

DOI: 10.1016/j.healthpol.2015.11.007

Health Policy xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Health Policy

journal homepage: www.elsevier.com/locate/healthpol

Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: A mixed methods study

Elena Nicod*, Panos Kanavos

London School of Economics and Political Science, Social Policy, LSE Health, Houghton Street, WC2A 2AE London, UK

ARTICLE INFO

Article history:
Received 27 April 2015
Received in revised form 17 October 2015
Accepted 24 November 2015

Keywords:
Health technology assessment
England
France
Sweden
Scotland
Mixed methods
Orphan drugs
Comparative analysis
Thematic analysis

ABSTRACT

Health Technology Assessment (HTA) often results in different coverage recommendations across countries for a same medicine 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 medicinal products. The criteria accounted for at each stage were qualitatively analyzed through thematic analysis. Piloting the framework for two medicines, eight trials, 43 clinical endpoints and seven 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 ($\kappa = 0.183$) was partly explained by stakeholder input ($n_s = 48$ times), or by 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, medicines and countries is ensured by its iterative and flexible nature, and detailed description.

1. Introduction

Health technology assessment (HTA) is widely adopted to inform coverage decisions of medicines or health technologies by healthcare systems. It relies on evidence about comparative effectiveness of alternative treatments in a particular clinical setting and aims to ensure that those covered provide value for money (or are cost-effective) [1], ultimately, improving access to medicines. In practice, countries frequently issue different coverage recommendations despite appraising the same body of clinical evidence and using similar methodological approaches. These differences are inevitable due to the complexity of these processes and the context within which they operate, where each country sets its own objectives for conducting HTA reflecting its values, preferences and constraints [2–4]. Implications include uneven access to these medicines across (often neighbouring) countries, non-optimal use of healthcare resources, and the unpredictability of the pharmaceutical market. Better understanding the application of HTA in different settings and the reasons for diverging recommendations through cross-country learning and

* Corresponding author.
E-mail address: e.m.nicod@lse.ac.uk (E. Nicod).

http://dx.doi.org/10.1016/j.healthpol.2015.11.007
0168-8510/

sharing of expertise is high on European and supra-national agendas, and may contribute to identify ways to minimize these differences [5,6] or understand how innovation was rewarded [7,8]. This is all the more important given the recent appreciation of HTA as a means towards universal healthcare [9] and the commitment of European Member States in implementing cross-border HTA collaboration through the EUnetHTA Joint Action 2.

Nine studies [10–18] compared HTA coverage recommendations for medicines in more than one country and identified important variations, where agreement ranged from poor to moderate [10,11,13]. 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 [10]. Another study highlighted cross-country variations for seventeen of the most expensive medicines, but the extent of, and reasons for these differences were not explored [12]. A more recent study investigated oncologic medicines, where negative recommendations were largely due to the high costs outweighing the marginal benefits [14]. Possible reasons for variations included differences in interpreting the clinical endpoints or in levels of patient input, or issues around appropriate comparators [14]. Another study highlighted differences across therapy areas and countries, suggesting that preferences varied according to the therapy area being appraised [13]. 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 better understanding the full range of factors accounted for and determining the extent to which they explain differences in coverage recommendations. Doing so would require a methodological approach that decomposes these processes to identify the key drivers contributing to decision-making in a systematic way. While this approach may 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 in different stages of HTA processes. A better understanding of value assessment processes may help address some of the methodological challenges in conducting HTA and, potentially, minimize cross-country differences when these were a consequence of the review or interpretation of the evidence.

The framework proposed in this study is informed by evidence from medicines with a European Medicines Agency (EMA) orphan medicinal designation [19], which have undergone an HTA in different settings in Europe. Orphan medicinal products are often characterized by significant inequalities in access [20] and are not always cost-effective [21]. 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.

2. Methods

2.1. 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 (Fig. 1) [23]. Both the depth and breadth of the HTA decision process were captured within the qualitative (stages I and II) and quantitative strands (stage III) [22,23]. A key characteristic of mixed methods design is the "iterative and cyclic approach used in the research" [24], 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 [25]. Priority was given to outline specifically the steps achieved in designing and piloting this methodological framework, while showcasing how the data collected can be analyzed quantitatively without drawing any conclusions due to the small sample size.

2.2. Sampling

Purposeful sampling was used to select the study countries with [27]: (a) well-established HTA agencies and processes, (b) similar decision-making criteria (clinical and/or cost-effectiveness), (c) adopting different approaches in HTA (e.g. clinical benefit versus clinical cost-effectiveness assessment, health service versus societal approach), and (d) publicly available HTA reports. The countries included were England, Scotland, Sweden and France (Box 1).

Medicine 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 orphan medicinal products – until December 2012 – and appraised in the four study countries. Excluded were those medicines that: (a) did not undergo the single technology assessment process at NICE, or the full submission process at SMC, and (b) did not receive diverging coverage recommendations. Coverage recommendations were either to list, restrict or reject the medicine under review, or in the case of France, to issue a ranking of clinical benefit (Service Médical Rendu, SMR) defining the coverage decision and rate, and one of improvement in clinical benefit (Amélioration du Service Médical 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 available.

This resulted in the selection of two compounds: eltrombopag (REVOLADE®) for the treatment idiopathic

Please cite this article in press as: Nicod E, Kanavos P. Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: A mixed methods study. Health Policy (2015), http://dx.doi.org/10.1016/j.healthpol.2015.11.007

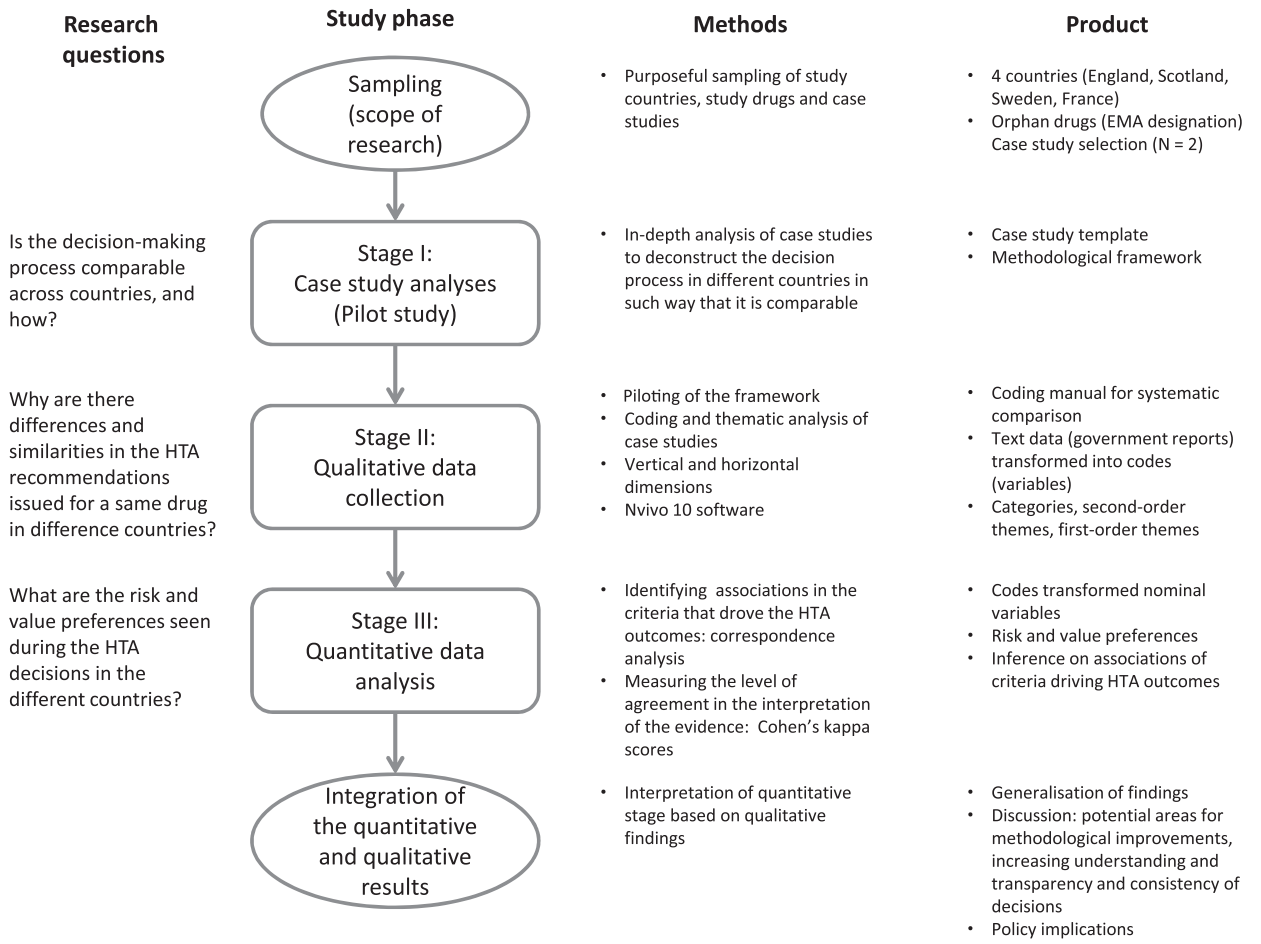

Fig. 1. Visual model of the mixed methods research design used. Illustrates the stages of this mixed methods study aiming to simplify the complex interrelationships by distinguishing between the different methods and data sources used to address the research questions and their integration in the final stage.

thrombocytopenic purpura (ITP) and everolimus (AFINITOR®) as second-line treatment for advanced renal-cell carcinoma (RCC) after failure of alternative therapies.

2.3. Data sources and analysis

Stage I (qualitative strand) involved case studies used as "an intensive study of a single unit for the purpose of defining a larger class of similar units" [29], to determine their comparability across countries. Previous research aiming at better understanding HTA decision-making was also used to outline the structure of the process [13]. 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).

A case study template and coding manual were developed. The case study template provides tables for the information to be extracted and coded, to ensure it is easily understandable and homogeneous across countries. Each line item represents one criterion and the countries considering it, and includes identifiers to ensure cross-country comparability. The coding manual exhaustively lists all the criteria included in the case study templates, organized into hierarchical levels clustered into common themes, referred to as first-order, second-order (clustering first-order themes) and third-order (clustering second-order themes) themes (Appendix A).

Stage II (qualitative strand) aimed to systematically capture the criteria accounted for during the HTA process through thematic analysis [25] and determine the similarities and differences across countries. Bottom-up coding was undertaken for each HTA report [30], where codes were created while examining the data to summarize and categorize the criteria identified within the case studies. The unit of coding, which is the section of text coded representing one criterion, was defined and illustrated with examples to avoid confusion or duplication in the results [31]. 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

Please cite this article in press as: Nicod E, Kanavos P. Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: A mixed methods study. Health Policy (2015), http://dx.doi.org/10.1016/j.healthpol.2015.11.007

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 (Appendix A).

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 external person, 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 [32].

In stage III, codes were quantitatively analyzed both vertically and horizontally through descriptive exploratory analyses [30] 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 [33]. 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 [35,36]. The horizontal dimension provided a measure of agreement between the HTA bodies in interpreting the same evidence using Cohen's kappa scores [37], allowing for a more robust evaluation of qualitative findings by comparing observed frequency of agreement with the probability of agreement occurring by chance.

2.4. 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.

3. Findings

The first two sections (qualitative strand) outline the information collected and coded based on the case study template, showcasing how the proposed structure was used to set up and pilot the methodological framework. Section 3.3 showcases how the data collected can be quantitatively analyzed, where the case studies were used as illustrative examples and results are by no means generalizable due to the small sample.

3.1. Qualitative strand: Developing the methodological framework

The structure of the methodological framework was based on the three key stages of the decision-making process identified: (a) the clinical and cost-effectiveness evidence, comprising clinical trials and endpoints, safety, economic models, and comparators, (b) the interpretation of this evidence, relating to uncertainty, "other considerations" and stakeholder input, and (c) the final HTA recommendation, and how the previous two stages influenced the recommendation formulation (Fig. 2). These criteria formed the basis for the case study template and coding manual, both tools forming the methodological framework (Appendix A).

3.2. Qualitative strand: Testing the methodological framework

Eltrombopag and everolimus received diverging HTA recommendations for the treatment of ITP and RCC, respectively. Eltrombopag was rejected in England, restricted in Scotland to patients with severe symptomatic ITP or a high risk of bleeding, 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 considered to provide an important medical benefit in France with a low added benefit (ASMR IV).

Please cite this article in press as: Nicod E, Kanavos P. Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: A mixed methods study. Health Policy (2015), http://dx.doi.org/10.1016/j.healthpol.2015.11.007

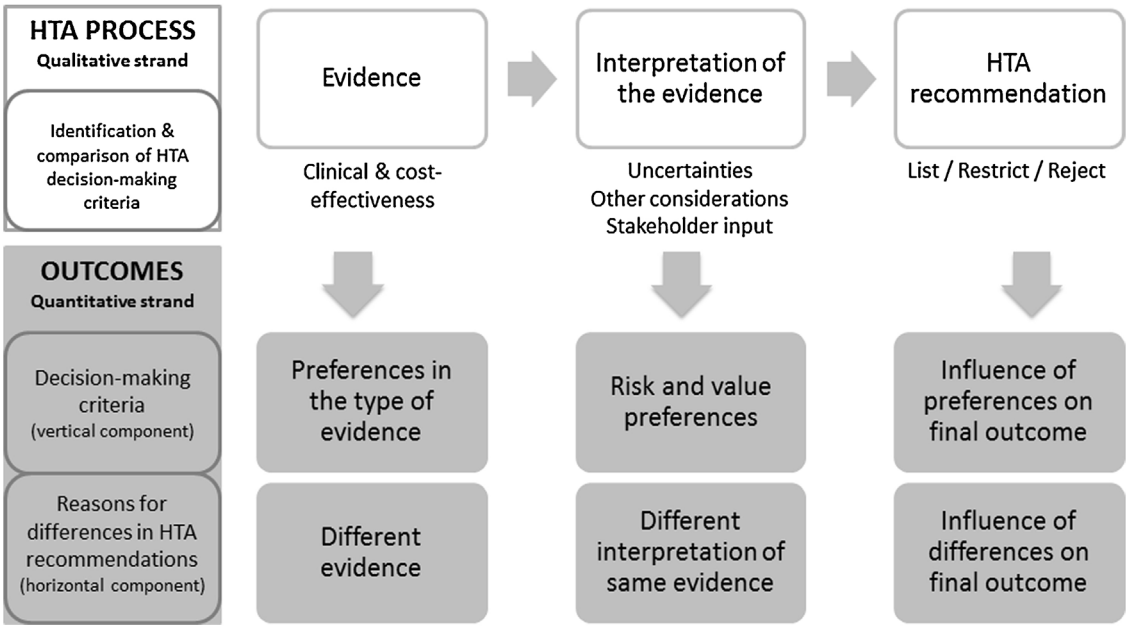

Fig. 2. Methodological framework for the systematic comparison of HTA processes. Illustrates the three key stages identified and used as a basis for the methodological framework, together with the outcomes from quantitatively analysing the data collected both vertically (e.g. agency-specific preferences)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

3.2.1. Clinical and cost effectiveness evidence

The same phase III primary trials were considered for both medicines. Additional trials were considered for eltrombopag, one of which was an indirect comparison with romiplostim appraised by all except HAS. For everolimus, subgroup analyses of the primary trial by prognostic categories were also considered by SMC and HAS, whereas NICE additionally conducted a meta-analysis of 28 studies (Table 1).

The clinical endpoints from the same trials were reported in a variety of ways (Table 1). For example, WHO 3–4 bleeding events were only recorded for eltrombopag by NICE, and health-related quality of life (HRQoL) 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 when considering only patient subgroups (as reported only by NICE). A similar scenario was seen for everolimus, where HRQoL was not reported by TLV and neither were objective response rate or progression-free survival from 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. The most common and clinically significant adverse events and treatment discontinuation rates were reported homogeneously. HAS usually provided more detail about the percentage of patients affected and deaths (even if not associated with the treatment).

The clinical evidence identified included: (a) the clinical trials, comprising eight trials (three primary trials) and five subgroup analyses with their respective comparators, coded 51 times across countries; (b) 43 different clinical endpoints 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. For example, each trial was coded according to the type of trial (e.g. phase III, phase II), the type of comparator (e.g. placebo, standard care), and whether it was a primary trial.

Economic models were considered by all HTA agencies except HAS. Those considered for eltrombopag differed from the comparator and/or type of model considered. NICE appraised three models with different comparators: (1) "watch and rescue", (2) romiplostim, and (3) a sequence of treatments. 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 at the time was under review at NICE), or were not valid, respectively. In contrast, SMC considered a cost-utility model comparing eltrombopag to romiplostim in splenectomised and non-splenectomised patients, and TLV a cost-minimization analysis with the same comparator based on a non-inferiority claim. For everolimus, the same cost-utility models were considered. This resulted in a cost per QALY ranging from £49,000 to £51,700 for NICE, £61,330 for SMC, and was not specified in TLV's HTA report. The cost-effectiveness evidence consisted in the economic model and its comparator, both coded 7 times (Appendix A).

3.2.2. Interpretation of the evidence

A total cumulative number of 18 clinical uncertainties were raised by the agencies (10 for eltrombopag and 8 for everolimus), coded 28 times (as some may have been raised by more than one agency) (Table 2). 56% (of 18 uncertainties) were based on evidence commonly appraised by all,

Please cite this article in press as: Nicod E, Kanavos P. Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: A mixed methods study. Health Policy (2015), http://dx.doi.org/10.1016/j.healthpol.2015.11.007

Table 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	√	√	√	√	Primary endpoint
Rescue treatment	√	√		√	
Bleeding (WHO1-4)	√	√	√	√	
Bleeding (WHO2-4)	√	√		√	Clinically significant bleeding
Bleeding (WHO3-4)	×				Gross (grade 3) and debilitating (grade 4) blood loss
Main reduction in bleeding	×	√			Seen in grade WHO2
HRQOL (SF36)	√ 4 ×6	√ 4			Significant over 4 domains, and not over 6
KUTER	Indirect comparison with RAISE				
Platelet response	√	NR	NR	NA	Primary endpoint
Platelet response	×				Splenectomised patients
Platelet response	×				Non-splenectomised patients
Everolimus					
RECORD-1	Phase III, placebo-controlled ($N = 416$)				
Progression-free survival	√	√	√	√	Primary endpoint
Overall survival	×	×	×	×	Blinded phase
HRQOL	*R**	×*		×**	*EORTC, FKSI-DRS **QLQ-C30
Objective response		×		×	
Progression-free survival (subgroup analysis)		√		√	Per risk stratification group

NICE: National Institute for Health and Care Excellence; SMC: Scottish Medicines Consortium; TLV: Dental and Pharmaceutical Benefits Board; HAS: Haute Autorité de Santé; √: Statistically significant; ×: Non-statistically significant; R: reported; NR: not reported; NA: not applicable; 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; Primary endpoint in bold.

while the remaining 44% on evidence included in only some of the appraisal reports. 38% (of 18 uncertainties) were also put forward as one of the main reasons for the final recommendation.

Table 2 illustrates the different phases in interpreting the clinical evidence. The first column (e.g. evidence, considered by) reports, per agency, the evidence interpreted and whether it was primary evidence. The second column (e.g. interpretation, raised by) reports cases when this evidence was considered uncertain, and whether it was nevertheless deemed acceptable or not (e.g. addressed or not). The last column (e.g. outcome, main reason for recommendation) reports cases when this issue was also one of the main reasons put forward for the final recommendation.

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 and one of the main reasons for the final recommendation for TLV and HAS, and the primary trial's small sample size was a concern only for TLV, but considered acceptable given the treatment's orphan status. Another example is NICE's concern that no improvement was seen in the lower 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 the endpoint was not specifically appraised or was not identified as being relevant to the decision. In the case of everolimus, the primary trial was terminated early due to the superior efficacy stopping rule, and results were biased due to cross-overs (81% of patients). Nevertheless, overall survival estimates were deemed plausible for NICE based on clinical expertise and results from a meta-analysis, and for SMC and TLV based on a specific tool used to derive these estimates. In contrast for HAS, no benefit was demonstrated (Table 2).

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 example, the model appraised by NICE was considered highly uncertain and driven by costs rather than benefit. One of the main concerns was about the assumption that differences in treatment arise because of bleeding events, which was used as the main effectiveness endpoint despite platelet response being the trial's primary endpoint. SMC was also concerned about the relative risk of bleeding events, but with a different comparator (romiplostim).

A number of "other considerations" were identified in the HTA reports and coded 62 times, differentiated by whether they pertained to the treatment (e.g. type of benefit, innovativeness) or to the disease (e.g. severity, unmet need) (Table 3). 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

Please cite this article in press as: Nicod E, Kanavos P. Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: A mixed methods study. Health Policy (2015), http://dx.doi.org/10.1016/j.healthpol.2015.11.007

Table 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 for recommendation			
		√ Evidence considered √*evidence considered within the pivotal trial				√ Positive influence (addressed)		× Negative influence (not addressed)		√ Positive influence (addressed)		× Negative influence (not addressed)	
	Clinical uncertainties	England NICE	Scotland SMC	Sweden TLV	France HAS	England NICE	Scotland SMC	Sweden TLV	France HAS	England NICE	Scotland SMC	Sweden TLV	France HAS
Eltrombopag	Lack of comparator	√*	√*	√*	√*	×	×	×	×			×	×
	Short duration of trial	√*	√*	√*	√*	×	×		×	×			
	Sample size	√*	√*	√*	√*			√				√	
	Trial population, indication under review	√*	√*	√*	√*	×				×			
	Trial population, generalizability	√*	√*	√*	√*	√	×						
	Trial population, low platelet count patients instead of those with severe risk of bleeding	√*	√*	√*	√*	√							
	Significant bleeding events (WHO3-4)	√*				×				×			
	Quality of life estimate	√	√						×				×
	Liver function monitoring		√				×						
	Uncertain nature of the indirect comparison	√	√√	√		×	×						
Everolimus	Bias in overall survival (cross-overs)	√*	√*	√*	√*	√	√	√	√	√	√	√	×
	Weak overall and partial response		√*		√*		×						
	Lack of comparative safety evidence	√	√	√	√		×						
	Trial population, patients with co-morbidities excluded	√*	√*	√*	√*		×						
	Trial population, generalizability	√*	√*	√*	√*	√							
	Risk of pneumonitis, immunosuppression	√	√	√	√	√							
	Quality of life estimate	√	√		√				×				
	Risk stratification method not applicable (subgroup analysis)		√	√	√		×						

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

Please cite this article in press as: Nicod E, Kanavos P. Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: A mixed methods study. Health Policy (2015), http://dx.doi.org/10.1016/j.healthpol.2015.11.007

Table 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 characteristics	Type of benefit	Curative, life-extending				√	√ End-of-life			√
	Innovativeness	Innovative, new class of drugs	√ Clinicians		√ Main		√ Clinicians/patients			
	Adverse effects	Similar across treatment arms	√							
		Patients willing to cope, manageable, tolerated, transient, reversible					√ Patients			
	Administration	Oral administration		√ Main	√ Main					
Disease characteristics	Unmet need	Unmet need, no treatment alternatives, few alternatives, need for treatment options	√Clinicians	√ Main	√ Main	√	√ Clinicians/patients			√
	Nature of the condition	Disease severity, serious condition			√ Main	√			√ Main	
		Life-threatening	√ Clinicians/patients		√ Main		√ Clinicians/patients/ End-of-life		√	√
		Short life-expectancy					√ End-of-life		√	
		Impact on quality of life, functional capacity, impact on daily activities	√ Clinicians/patients		√Main	√				
		Social stigma, limiting of life-style choices, ability to work, travel or undertake leisure activities, limiting life-style choices	√ Clinicians/patients							
	Rare disease	Orphan status, small population, rarity		√ Main	√ Main	√	√			
	Clinical practice	No routine standard pathway, complex clinical practice, tailored to the patient	√ Clinicians	√ Clinicians						
		Comparator unlicensed for indication, and associated with important adverse events and anxiety	√ Clinicians/patients							
		No long term evidence and unknown correct dosage of comparator	√ Clinicians							
		Later diagnosis when disease advanced							√	
		Licensed over non-licensed preferred	√ Clinicians							
	National priority	Plan Maladies Rares (2004)				√				

NICE: National Institute for Health and Care Excellence; SMC: Scottish Medicines Consortium; TLV: Dental and Pharmaceutical Benefits Board; HAS: Haute Autorité de Santé; Main: considered one of the main reason for the final recommendation; X: considered during the assessment; Clinicians: put forward by clinicians; Patients: put forward by patients; End-of-life: eligible as "end-of-life treatment" for NICE.

Please cite this article in press as: Nicod E, Kanavos P. Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: A mixed methods study. Health Policy (2015), http://dx.doi.org/10.1016/j.healthpol.2015.11.007

eltrombopag was one of the main reasons for the final decision by SMC and TLV, 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 treatment as an end-of-life treatment.

Stakeholder input was identified 46 times in the NICE reports and twice in the SMC reports; none was identified for TLV or HAS. For NICE, patient input providing information about "other considerations" (e.g. impact on daily activities) was seen in 30% of cases (14 out of 46), and clinical input about "other considerations" (e.g. limited treatment options) or commenting on clinical and economic uncertainties (e.g. clinical practice) was seen in 70% of cases (32 out of 46). For SMC, clinical experts provided input about "other considerations" (e.g. clinical practice) and commented on uncertainty (e.g. generalizability) (Table 3).

3.3. Quantitative strand (Stage III): Outcomes from the methodological framework

3.3.1. Clinical evidence and its interpretation

Reasons for differing HTA recommendations include instances when different evidence was considered, or when the same evidence was interpreted differently. Table 2 illustrates cases when uncertainties were raised based on *different clinical evidence*. For example, HAS was concerned about the lack of HRQoL data given the impact of the disease on the patient. This was not an issue for NICE and SMC since it was included in the assessment.

When considering *the same evidence*, agreement was poor using Cohen's kappa scores ($\kappa = 0.183$, 95% CI [0.015; 0.35]) [38]. These differences may relate to a subjective (unexplained) component of the decision or to different risk or value preferences. Risk preferences, when one agency is relatively more likely to raise a concern when appraising *the same evidence* compared to the other agencies, were identified through correspondence analysis. Although the chi-squared probability of independence was non-significant given the small sample size ($\chi^2 = 22.49$; $p = 0.550$), the 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, illustrated in the correspondence analysis biplot (Appendix B). 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 medicines ($\chi^2 = 30.97$; $p = 0.029$) (Appendix C). One example illustrating how preferences influenced the acceptability of an uncertainty was for eltrombopag, where the small sample size was deemed acceptable for TLV given its orphan status (Tables 2 and 3, Appendixes B and C).

3.3.2. Criteria driving the decisions

The criteria that drove the HTA decision-making process consist of the main reasons for the recommendation identified at each stage of the process (Tables 2 and 3), potentially influenced by agency-specific risk or value preferences or because of the subjectivity in the interpretation of the results (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 medicinal product, 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 HRQoL were also put forward, with a follow-up of effectiveness to take place in October 2013. For HAS, while the trial duration was limited and comparative data was 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. For NICE, overall survival was considered superior to 3 months and the treatment was eligible as "end-of-life treatment". Despite this, sensitivity analysis showed only a very low probability of the medicine 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 HRQoL relative to alternatives.

4. Discussion and policy implications

4.1. 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.

4.2. How do our findings fit with existing evidence?

Comparing our results with existing studies that investigated 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

Please cite this article in press as: Nicod E, Kanavos P. Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: A mixed methods study. Health Policy (2015), http://dx.doi.org/10.1016/j.healthpol.2015.11.007

and their findings are consistent with ours. One review of quantitative studies identified existing empirical evidence about coverage decisions for a range of health technologies [39]. 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 of medicines made by the Evidence Review Group (ERG) at NICE to identify the strengths and weaknesses in the submissions [40]. 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. 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 reported in the introduction section. None, however, have compared the decision-making processes in a systematic manner.

4.3. The methodological framework

The added-value of this study is 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 [41]. 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 medicines. 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 interdisciplinary potential of applying such designs to novel areas, such as HTA.

4.4. 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 medicines, 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 medicine-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 to raise awareness on the reasons for cross-country differences. When 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 medicines, 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. 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).

5. Conclusion

Improving the level of access to medicines is a priority at both European and supra-national levels [7,8]. The substantial variations seen in HTA recommendations is one of the causes of differential access to medicines and possibly reflects weaknesses in HTA methodologies, where the implications are enormous in seeking to obtain value for money. The urgent need to better understand the reasons for variations in decision-making processes and improve the quality of assessments is recognized. In this study, we have proposed, developed and piloted a methodological framework aiming to account for (part of) the unexplained heterogeneity 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 currently being enhanced and applied to a larger sample of medicines, therapy areas and countries.

Conflict of interest statement

No conflicts of interest arise.

Acknowledgements

This research is funded under the European 7th Framework Programme with Advance-HTA. The results presented reflect the authors' views and not the views of the European Commission. The EC is not liable for any use of the information communicated. The authors thank all partners from the Advance-HTA consortium for their reviews, particularly (in alphabetical order) Karl Arnberg, Karen Brigham, Isabelle Durand-Zaleski, Jaime Espin, Karen Facey, Andres Freiburg, Niklas Hedberg, Douglas Lundin, Lise Rochaix, Francis Ruiz, Domenica Taruscio, and Thomas Wilkenson. We also thank Aude Bicquelet for her courses on qualitative research. Finally, we thank our colleagues

Please cite this article in press as: Nicod E, Kanavos P. Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: A mixed methods study. Health Policy (2015), http://dx.doi.org/10.1016/j.healthpol.2015.11.007

Paul Dolan, Emilie Courtin and Alessandra Ferrario for reviewing this work and providing valuable feedback.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.healthpol.2015.11.007.

References

[1] Luce B, Drummond M, Jonsson B, Neumann P, Schwartz J, Siebert U, et al. HTA, and CER: clearing the confusion. The Milbank Quarterly 2010;88:256–76.
[2] Banta D. The development of health technology assessment. Health Policy 2003;63:121–32.
[3] Schwarzer R, Siebert U. 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 2009;25:305–14.
[4] Busse R, Orvain J, Velasco M, Perleth M, Drummond M, Gurtner F, et al. Best practice in undertaking and reporting health technology assessments—working group 4 report. International Journal of Technology Assessment in Health Care 2002;18:361–422.
[5] Commission E. Directive 2011/24/EU on patients' rights in cross-border healthcare; 2011.
[6] European Commission, DG Health and Consumers, Public Health. Pharmaceutical pricing and reimbursement information sheet; 2014.
[7] European Commission, Enterprise and Industry, Pharmaceutical Forum. Final conclusions and recommendations of the Pharmaceutical Forum: improving access to orphan medicines for all affected EU citizens; 2013.
[8] Bouvy J, Vogler S. 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. WHO Collaborating Centre for Pharmaceutical Policy and Regulation; 2013.
[9] World Health Organisation. Health intervention and technology assessment in support of universal health coverage; 2014, http://apps.who.int/gb/ebwha/pdf_files/WHA67/A67_R23-en.pdf.
[10] Clement F, Harris A, Li J, Yong K, Lee K, Manns B. Using effectiveness and cost-effectiveness to make drug coverage decisions a comparison of Britain, Australia, and Canada. JAMA—The Journal of the American Medical Association 2009;302:1437–43.
[11] Lexchin J, Mintzes B. Medicine reimbursement recommendations in Canada, Australia, and Scotland. American Journal of Managed Care 2008;14:581–8.
[12] Morgan S, McMahon M, Mitton C, Roughead E, Kirk R, Kanavos P, et al. Centralized drug review processes in Australia, Canada, New Zealand, and the United Kingdom. Health Affairs 2006;25:337–47.
[13] Nicod E, Kanavos P. Commonalities and differences in HTA outcomes: a comparative analysis of five countries and implications for coverage decisions. Health Policy 2012;108:167–77.
[14] Shah K, Mestre-Ferrandiz J, Towse A, Smyth E. A review of health technology appraisals: case studies in oncology. International Journal of Technology Assessment in Health Care 2013;29:101–9.
[15] Kanavos P, Nicod E, Van den Aardweg S, Pomedli S. The impact of health technology assessments: an international comparison. EuroObserver 2010;12:1–6.
[16] Pomedli S. HTA and access to cancer medicines. EuroObserver 2010;12:7–8.
[17] Nicod E. Harmonizing the outcomes of HTA: an application to Central Nervous System treatments. EuroObserver 2010;12:9–13.
[18] Van den Aardweg S. HTA of orphan drugs across six countries: fair, flawed or failing? Euro Observer 2010;12:14–9.
[19] European Medicines Agency. Human medicines, rare disease (orphan) designations. London, UK: European Medicines Agency; 2013, http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/landing/orphan_search.jsp&mid=WC0b01ac058001d12b.
[20] Le Cam Y. Inventory of access and prices of orphan drugs across Europe: a collaborative work between national alliances on rare diseases & Eurordis. Eurordis Rare Diseases Europe; 2010.
[21] Dupont A, Van Wilder P. Access to orphan drugs despite poor quality of clinical evidence. British Journal of Clinical Pharmacology 2011;71:488–96.
[22] Creswell J. The selection of a research approach. In: Oaks T, editor. Research design: qualitative, quantitative and mixed methods approaches. Los Angeles, CA: Sage; 2003. p. 1–24.
[23] Creswell J. Plano Clark V. Designing and conducting mixed methods research. Los Angeles, CA: Sage Publications; 2011.
[24] Teddlie C, Tashakkori A. Overview of contemporary issues in mixed methods research. In: Tashakkori A, Teddlie C, editors. Sage handbook of mixed methods in social & behavioral research. Los Angeles, CA: Sage Publications; 2010. p. 1–42.
[25] Bryman A. Social research methods. Oxford: Oxford University Press; 2004.
[27] Creswell J, Plano Clark V. Collecting data in mixed methods research. In: Creswell J, Plano Clark V, editors. Designing and conducting mixed methods research. Los Angeles, CA: Sage Publications; 2011. p. 110–27.
[29] Gerring J ,. What is a case study good for? Case study versus large-n cross-case analysis. In: Gerring J, editor. Case study research: Principles and practices. New York, NY: Cambridge University Press; 2007. p. 37–63.
[30] Onwuegbuzie A, Combs J. Emergent data analysis techniques in mixed methods research: a synthesis. In: Tashakkori A, Teddlie C, editors. Sage handbook of mixed methods in social & behavioral research. Los Angeles, CA: Sage Publications; 2010. p. 397–430.
[31] Boyatzis RE. Developing themes and codes. In: Boyatzis RE, editor. Transforming qualitative information: thematic analysis and code development, vol xvi. Thousand Oaks, CA: Sage Publications; 1998. p. 184.
[32] Ltd QIP. NVivo qualitative data software; 2012.
[33] StataCorp. In: Station C, editor. Stata statistical software: release 13. TX: StataCorp LP; 2013.
[35] Dickenson W. Visual displays for mixed methods findings. In: Tashakkori A, Teddlie C, editors. Sage handbook of mixed methods in social & behavioral research. Los Angeles, CA: Sage Publications; 2010. p. 469–504.
[36] Bartholomew D, Steele F, Moustaki I, Galbraith J. Analysis of multivariate social sciences data, 2nd ed. Florida: Taylor & Francis; 2008.
[37] Cohen J. Coefficient of agreement for nominal scales. Educational and Psychological Measurement 1960;20:37–46.
[38] Altman D. Practical statistics for medical research. London, England: Chapman and Hall; 1991.
[39] Fischer K. A systematic review of coverage decision-making on health technologies-Evidence from the real world. Health Policy 2012;107:218–30.
[40] Carroll C, Kaltenthaler E, FitzGerald P, Boland A, Dickson R. A thematic analysis of the strengths and weaknesses of manufacturers' submissions to the NICE Single Technology Assessment (STA) process. Health Policy 2011;102:136–44.
[41] Creswell J, Tashakkori A. Developing publishable mixed methods manuscripts. Journal of Mixed Methods Research 2007;1:107–11.

Please cite this article in press as: Nicod E, Kanavos P. Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: A mixed methods study. Health Policy (2015), http://dx.doi.org/10.1016/j.healthpol.2015.11.007