



Multiple Criteria Decision Analysis (MCDA) for evaluating new medicines in Health Technology Assessment and beyond: The Advance Value Framework



Aris Angelis^{*}, Panos Kanavos

Department of Health Policy and Medical Technology Research Group, LSE Health, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom

ARTICLE INFO

Article history:

Received 20 October 2016

Received in revised form

12 May 2017

Accepted 19 June 2017

Available online 20 June 2017

Keywords:

European health policy

Multiple Criteria Decision Analysis

New medicines

Pharmaceuticals

Health Technology Assessment

Value Based Assessment

Value framework

Decision theory

ABSTRACT

Escalating drug prices have catalysed the generation of numerous “value frameworks” with the aim of informing payers, clinicians and patients on the assessment and appraisal process of new medicines for the purpose of coverage and treatment selection decisions. Although this is an important step towards a more inclusive Value Based Assessment (VBA) approach, aspects of these frameworks are based on weak methodologies and could potentially result in misleading recommendations or decisions.

In this paper, a Multiple Criteria Decision Analysis (MCDA) methodological process, based on Multi Attribute Value Theory (MAVT), is adopted for building a multi-criteria evaluation model. A five-stage model-building process is followed, using a top-down “value-focused thinking” approach, involving literature reviews and expert consultations. A generic value tree is structured capturing decision-makers’ concerns for assessing the value of new medicines in the context of Health Technology Assessment (HTA) and in alignment with decision theory.

The resulting value tree (Advance Value Tree) consists of three levels of criteria (top level criteria clusters, mid-level criteria, bottom level sub-criteria or attributes) relating to five key domains that can be explicitly measured and assessed: (a) burden of disease, (b) therapeutic impact, (c) safety profile (d) innovation level and (e) socioeconomic impact. A number of MAVT modelling techniques are introduced for operationalising (i.e. estimating) the model, for scoring the alternative treatment options, assigning relative weights of importance to the criteria, and combining scores and weights.

Overall, the combination of these MCDA modelling techniques for the elicitation and construction of value preferences across the generic value tree provides a new value framework (Advance Value Framework) enabling the comprehensive measurement of value in a structured and transparent way. Given its flexibility to meet diverse requirements and become readily adaptable across different settings, the Advance Value Framework could be offered as a decision-support tool for evaluators and payers to aid coverage and reimbursement of new medicines.

© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Background

Scarce resources, rising demand for health services, ageing populations and technological advances threaten the financial sustainability of many health care systems and render efficient and fair resource allocation a cumbersome task (Beauchamp, 2003; Eddy, 1991; Emanuel, 2000; Fleck, 2001; Rawls, 1999). Decision-making in health care is inherently complex as numerous

objectives need to be balanced, usually through the involvement of many stakeholders. One set of tools used widely to improve efficiency in resource allocation is Health Technology Assessment (HTA). The use of HTA has expanded significantly over the past 20 years and it is now used to assess and appraise the value of new medical technologies as well as inform coverage decisions.

Evidence-based medicine (Guyatt, 1992; Sackett et al., 1996), economic evaluations (Drummond and McGuire, 2001; Drummond et al., 2015), burden of disease estimates (Murray and Lopez, 1996), and budget impact analysis (Mauskopf et al., 2007) can be used to inform decisions on resource allocation. Nevertheless, they offer

^{*} Corresponding author.

E-mail address: a.n.angelis@lse.ac.uk (A. Angelis).

limited guidance to decision-makers, as their results cannot be integrated and judged simultaneously and neither can the associated value trade-offs (Baltussen and Niessen, 2006). The use of economic evaluation techniques such as cost effectiveness analysis (CEA) has become the preferred analytical method adopted by many HTA bodies. However, all value-related concerns of decision-makers are not adequately reflected in a cost effectiveness model (Angelis et al., 2017a). For example, the use of cost utility analysis (CUA) and the incremental cost per quality adjusted life year (QALY) has become the metric of choice for many HTA bodies when assessing and appraising value, although by definition the latter only considers length of life in tandem with health related quality of life, and does not adequately capture social value such as the wider innovation and socioeconomic impact (Brouwer et al., 2008; Wouters et al., 2015).

Due to the complexity of these multidimensional problems, decision-makers often tend to adopt intuitive or heuristic approaches for simplification purposes, but, as a consequence, important information may be either under-utilised or altogether excluded leading to choices based on an ad-hoc priority setting process (Baltussen and Niessen, 2006). Eventually the decision making process tends to be explicitly informed chiefly by evidence from economic evaluations, with social value concerns being considered on an implicit and ad-hoc basis. As a consequence, when faced with multiple trade-offs across a range of societal values, decision-makers seem to not be adequately equipped to make well-informed and “rational” decisions (Bazerman, 1998; McDaniels et al., 1999); this may diminish the reasonableness and credibility of decision outcomes. It is probably the case that a more “rational” approach is needed that can simultaneously take into account the multiplicity of criteria as well as aggregate the performance of alternative interventions across the criteria of interest while accounting for differences between their relative importance, therefore enabling the overall construction and analysis of decision-makers' preferences in a simple and transparent way (Angelis et al., 2017b; Baltussen and Niessen, 2006). The multiple value framework initiatives that have emerged over the past few years aiming to aid reimbursement agencies, health care professionals and patients understand the value of new therapies and make better choices about their use serve as a testament to this particular gap (Neumann and Cohen, 2015).

Some of the most prominent and well known value frameworks that have attracted high levels of attention include those proposed by the American College of Cardiology and the American Heart Association (ACC/AHA) (Anderson et al., 2014), the American Society of Clinical Oncology (ASCO) (Schnipper et al., 2015), the European Society of Medical Oncology (ESMO) (Cherny et al., 2015), the Institute for Clinical and Economic Review (ICER) (Pearson, 2015), the Memorial Sloan Kettering Cancer Centre (MSKCC) (Bach, 2015), the National Comprehensive Cancer Network (NCCN) (NCCN, 2015), and the Working Group on Mechanisms of Coordinated Access to Orphan Medicinal Products (MoCA-OMP) (Working Group on Mechanism of Coordinated Access to Orphan Medicinal Products, 2013), among others. These value frameworks adopt multiple criteria approaches in an attempt to decompose complex problems into slightly simpler ones and address these sequentially. As such, they are an important step towards a more inclusive Value Based Assessment (VBA) process despite their methodological limitations, which may limit their value for policy or clinical decision-making (Angelis and Kanavos, 2016a; Malone et al., 2016). Despite the proliferation of value frameworks, “value” remains an elusive target and a wider consensus about what dimensions of value to include may still be some way off, with the potential exception of some frameworks that are applied in clinical practice (as potential value dimensions are more restricted in nature).

The use of multiple criteria decision analysis (MCDA) methods has been proposed as an alternative methodological approach for assessing the value of health care interventions in different contexts, ranging from licensing decisions at marketing authorization stage (Mussen et al., 2007a; Phillips et al., 2011), to coverage decisions at HTA stage (Goetghebeur et al., 2012; Thokala and Duenas, 2012), to treatment selection decisions at prescribing level (Felli et al., 2009; Tervonen et al., 2015). MCDA methods can be used for quantifying benefits, risks and uncertainties in order to aid the decision-making process, by considering an explicit set of criteria and their relative importance under a fully transparent process, while incorporating a wide range of stakeholder views to express a more societal perspective.

A methodological process towards the development of a robust MCDA framework following key principles to ensure methodological robustness, consisting of different phases and stages for implementation in the context of HTA, has already been proposed (Angelis and Kanavos, 2016b). Possibly the most fundamental phase of the MCDA process with the highest impact on the overall decision outcomes relates to model building and the criteria selection phase, which influences the choice of the value concerns which the alternative options will be assessed against.

In this paper we outline a new value framework focusing on model-building and criteria selection; we develop a generic value based model taking the form of a value tree for the purpose of assessing the value of new medicines in the context of HTA; and we do so by adopting the perspective of payers as decision-makers. Although in theory such a value tree can be generic for any type of health technology, including medicines, medical devices and other health interventions, for ease of illustration, we focus on new medicines. In Section 2 we discuss the theoretical foundations of MCDA, starting with the theoretical axioms of decision analysis that essentially address the question of “what is the basis of MCDA”. In section 3 we outline the methods for model building and the selection of different evaluation criteria. In section 4 we present the results involving the assembly of decision-makers' concerns into a comprehensive generic value tree, which addresses the question of “what is value in the context of HTA”. In section 5 we propose ways to operationalise the value tree through a precise combination of MCDA techniques addressing the question of “how to apply MCDA in HTA” through the new value framework, we discuss a number of behavioural biases that MCDA practitioners should be aware of, we link the use of MCDA with opportunity costs and resource allocation prioritisation, and compare the new value framework with existing frameworks. Finally, in section 6 we draw the main conclusions.

2. Theoretical foundations

Decision analysis was originally defined by Howard as “a logical procedure for the balancing of the factors that influence a decision” incorporating “uncertainties, values, and preferences in a basic structure that models the decision” (Howard, 1966). The logic behind decision analysis was described as “divide and conquer” whereby a complex problem is decomposed into simpler problems and each individual problem is analysed separately before all analyses are connected together, resulting in a program of action for the complex problem (Raiffa, 1968).

The methodology of decision analysis is simplified to four main steps, notably (a) structuring the decision problem; (b) assessing possible impacts of each alternative; (c) determining preferences (values) of decision-makers; and (d) evaluating and comparing alternatives (Keeney, 1982; Raiffa, 1968).

The main outcome of decision analysis theoretical axioms is that, first, the utility of an alternative is the indication of its

desirability, and second, that an alternative A with higher utility $U(A)$ should be preferred to an alternative B with lower utility $U(B)$, in other words indicating utility maximisation, i.e.

$$A > B \Leftrightarrow U(A) > U(B) \quad (1)$$

Decision analysis seems to be particularly useful for coping with the complexities arising from uncertainty and multiple conflicting objectives (Raiffa, 1968). Uncertainties can be traded-off against some value aspects of the outcomes, formally through the incorporation of probabilities. Similarly, trade-offs can be made among different objectives and their associated values. These trade-offs are judgements, relating to the relative importance of the respective value dimensions. However, given that trade-offs are personal, there are no universal rules for making them, and they are therefore subjective in nature (von Winterfeldt and Edwards, 1986). By definition, decision-making is inherently subjective as it depends on individual utility and preferences. Therefore, the notion of “rationality” is used with the goal of making logical inferences and decisions. Possibly the most prominent of such rationality rules would be the maximisation of (expected) utility or value (Savage, 1954).

In decision analysis there is a clear conceptual difference between value functions, which assess the incremental benefit of an option, and utility functions, which incorporate preferences, therefore assessing both marginal benefit and risk attitude. The former are employed for riskless choices and the latter for choices under uncertainty. Expectation mainly relates to the concept of probability theory, with *expected* utility or value being a weighted average of the pay-off for an outcome (utility or value) and its respective probability of actually occurring (von Winterfeldt and Edwards, 1986). Expected utility and expected value are, therefore, calculated in a similar way but the pay-offs in the former case correspond to subjective utilities rather than objective quantities, with their formal terminology being *subjectively expected utilities* (SEUs): both the utilities and the probabilities incorporated are numbers, but are subjective in nature giving rise to *numerical subjectivity*, the notion of subjective judgments expressed as numbers (von Winterfeldt and Fasolo, 2009). In the context of HTA, the evaluation of new medical technologies predominantly relates to the evidence-based assessment of their value by measuring their marginal benefits. Therefore we choose to use the term *value* rather than *utility*, which reduces the heterogeneity of the value judgements based on expected utility.

The expected value (EV) of an event could then be written as

$$EV = p_1 v_1 + p_1 v_2 + p_3 v_3 + \dots + p_n v_n \quad (2)$$

Or alternatively as

$$EV = \sum_{i=1}^n p_i v_i \quad (3)$$

where p_i is the probability that event i will take place and v_i is the value or pay-off associated with the event.

Given that the evaluation of health care interventions as part of HTA predominantly relates to the assessment of their value using existing evidence and not expected evidence, for simplicity we detach expectation and the probability concept while assuming the existence of evidence for the measurement of benefit. However, it could also be argued that in various cases absence of (satisfactory or adequate) evidence introduces an expectation variable given the attached probabilities of the respective outcomes taking place (e.g. clinical outcome), especially as part of early-HTA settings, where a new medicine might still be in clinical development and evidence generation phase.

The same axiomatic rule arising from (expected) value maximisation could be adopted in the context of HTA: a medical technology A with higher value would be preferred over another medical technology B with lower value. The term “preference” would relate to a coverage decision (i.e. reimbursement) for the health care technology with the higher value, the aim being to maximise efficiency in resource allocation.

The starting point for this discussion is that the value of new medical technologies is multidimensional and not strictly limited to their clinical effect or benefit. Besides the traditional dual clinical consideration of health benefits and risks, taking the form of efficacy and safety considerations respectively, or the emerging dual economic consideration of health benefits and costs, taking the form of health outcomes and costs per unit of outcome respectively, other factors may also be important in determining the value of a new medicine. The severity and unmet need of the disease (i.e. burden of illness), the clinical novelty and convenience to patients (i.e. innovative nature of technology), or the wider benefits to society (i.e. non-health objectives to the NHS), have often been perceived as important considerations of value to decision-makers and can contribute to the debate on efficient resource allocation (DH, 2010; NICE, 2014a).

Therefore, the value of new medical technologies could be illustrated as a function of different evaluation parameters, namely:

$$Value = f(a_{ij}, b_{ij}, c_{ij}, d_{ij}, \dots, n_{ij}) \quad (4)$$

where a, b, c, d, \dots, n denote the different parameters of interest, i denotes the medical technology's partial value in regards to a particular parameter and j denotes the relative importance weight of the same parameter based on decision-maker or other stakeholder views. Additionally, questions remain on how to incorporate the views of all relevant stakeholders and how to derive relative weights for the different parameters. Generally, a linear additive (i.e. weighted average) model is used to combine scores and weights together, whose specification requires the use of explicit lower and higher reference levels (typically taking the values of 0 and 100) and positive criteria weights the sum of which should be equal to one (or 100%).

Given that the value of new medical technologies is based on a multitude of value dimensions and the limitations of current approaches for their assessment in the context of resource allocation decisions, there is a need for a methodological approach that encompasses multiple value domains explicitly, therefore a decision analysis method that addresses multiple attributes of benefit is required (Angelis et al., 2017b).

Among methodological tools for assessing value quantitatively as part of a decision-making process, MCDA could be indicated as an ideal method, ordering a set of alternative options based on the degree to which a number of different objectives are achieved (Department for Communities and Local Government, 2009; Keeney and Raiffa, 1976). One of the main aims of MCDA methods is to enable decision-makers to reach a decision by facilitating an understanding of the problem, objectives, and associated values, through organising and synthesising information of complex and conflicting nature (Belton and Stewart, 2002). MCDA can facilitate decision-making by explicitly integrating objective measurement with value judgement while managing subjectivity in a transparent way, however it cannot act as a substitute for decision-making.

Having introduced the theoretical foundations of decision analysis and the overall MCDA process, the next step is to focus on the model-building phase by applying MCDA principles. As part of the model-building phase, criteria selection is crucial. This involves the identification and assembly of criteria into a hierarchical structure taking the form of a tree (Berkeley and Humphreys, 1982;

von Winterfeldt and Fasolo, 2009). The aim is to arrive at a generic value model for new medicines that can be adapted to capture all relevant dimensions of value across different decision-making contexts and therapeutic indications. Criteria represent the key concerns and considerations influencing a particular decision. Structuring all criteria in the form of a tree is known as a *value tree* and provides an organized schematic representation of the various concerns under consideration by the decision-maker. The criteria-based evaluation of options is operationalised through the use of performance descriptors, either of a qualitative or quantitative nature, known as *attributes*, which measure the fulfillment of the criteria (Box 1).

3. Methods

3.1. Model building approaches

The three main steps in building a multi-criteria model are to (a) structure a value tree that identifies and represents the objectives or key concerns of decision-makers, (b) define attributes at the bottom level of the value tree to measure the achievement of these objectives, and (c) select decision alternatives (Franco and Montibeller, 2010).

Structuring a value tree can generally be performed using two approaches: either a top-down approach (known as value-focused thinking) (Keeney, 1992), driven by the overall objective or value concern, which is decomposed into lower levels of sub-objectives or sub-concerns; or a bottom-up approach (known as alternative-focused thinking), driven by the alternative options under consideration and based on particular attributes used to distinguish between them, which are grouped into higher levels of objectives and concerns (von Winterfeldt and Edwards, 1986).

Given the aim of the study is to build a generic model that can subsequently be adapted and applied across different decision-making contexts, we used the top-down, value-focused thinking approach (Keeney, 1992), so that the model can reflect the overall value concerns of decision-makers while being adaptable to different decision problems. With regards to the completion of the remaining two tasks (attribute definition and selection of decision alternatives), we adopt the bottom-up, alternative-focused thinking approach, following the definition of the decision problem (Buede, 1986), so that the model can become decision-specific to precisely assess the performance of alternative treatments as needed. As such, we suggest the completion of the two tasks and the model building process as part of particular

Box 1

Definitions of decision analysis terminology

Criterion: an “individual measurable indicator” of a key value dimension (Department for Communities and Local Government, 2009) or more specifically, a “particular perspective” according to which alternative technologies may be compared (Belton and Stewart, 2002).

Attribute: a “quantitative or qualitative measure of performance associated with a particular criterion” (Belton and Stewart, 2002), i.e., a descriptor of performance or impact requiring ordering of preference (Keeney, 1992).

Value tree: an organized schematic representation of the various objectives, criteria and attributes under consideration.

applications, where the comparison of actual decision options feeds into the selection of precise attributes for distinguishing their value.

Overall, in order to complete the three main steps, we have proposed a “value-alternative hybrid thinking” approach (Angelis and Kanavos, 2016b), in which the core structure of the value tree takes place as part of a top-down approach, and the definition of attributes is ultimately completed following the selection of the decision alternatives as part of a bottom-up approach.

Importantly, both criteria and attributes of the value tree must possess a number of properties for the results to be robust (Angelis and Kanavos, 2016b). These include being essential, non-overlapping and concise (for the criteria); unambiguous, comprehensive and direct (for the attributes); understandable, operational, and preference-independent (for both criteria and attributes) (Belton and Stewart, 2002; Franco and Montibeller, 2010; Keeney and Raiffa, 1993; Keeney and Gregory, 2005). A discussion of these properties together with specific examples as well as the significance of adhering to them to ensure a robust process is provided later in the text and also elsewhere (Marsh et al., 2016).

3.2. HTA adaptation process and staging

In the context of HTA, it would be reasonable to assume that the value concerns (i.e. the evaluation criteria) of decision-makers would encompass all key factors that are necessary for a comprehensive value appraisal of a new medicine or health care intervention, with their relevant characteristics and impacts acting as descriptors of performance (i.e. attributes) for measuring the extent of satisfying these criteria. However, a universal or common set of decision-makers' value concerns and their measures are neither clearly defined, nor well-established, both within and across health systems. Therefore, a definition of *value* in the context of HTA is currently absent (Angelis et al., 2017a; Antoñanzas et al., 2016).

In order to capture different decision-makers' concerns in a comprehensive manner that ultimately would lead to the structuring of a generic value tree, a five-stage iterative model-building process was adopted, involving both secondary and primary data collection and using a top-down “value-focused thinking” approach. The five steps, shown in Fig. 1 were as follows:

First, we conducted a systematic review of the peer-reviewed literature to identify what value dimensions are considered in the evaluation processes of HTA bodies in eight EU countries (Angelis et al., 2017a). The Centre for Review and Dissemination guidance for undertaking systematic reviews in health care was followed (Centre for Reviews and Dissemination, 2009). A flow diagram of the literature review process is shown in Fig. 2. The eight study countries (and their HTA bodies) were France (Haute Autorité de Santé, HAS), Germany (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, IQWiG), Sweden (Tandvårds-och läkemedelsförmånsverket, TLV), England (National Institute for Health and Care Excellence, NICE), Italy (Agenzia Italiana del Farmaco, AIFA), the Netherlands (Zorginstituut Nederland, ZIN (formerly College voor zorgverzekeringen, CVZ)), Poland (The Agency for Health Technology Assessment and Tariff System, AOTMiT) and Spain (Red de Agencias de Evaluación de Tecnologías Sanitarias y Prestaciones del Sistema Nacional de Salud (RedETS) and the Inter-ministerial Committee for Pricing (ICP)). The rationale for their selection was the variation in their health system financing (tax-based vs. social insurance-based), the organisation of the health care system (central vs. regional organisation), the type of HTA in place (predominantly economic evaluation vs. predominantly clinical benefit assessment) and the perspective used in HTA (health system vs. societal), so that the sample is representative of the main types of health systems and

HTA approaches across Europe. Inclusion criteria for the review were the following: (i) English language, (ii) evidence from the eight study countries of interest (and their respective agencies), (iii) HTA context from a national coverage perspective, and (iv) a publication date between January 2000 and January 2014; product-specific technology appraisals or evaluation studies and conference proceedings or records with no abstract available were excluded. The electronic databases of Medline and Social Science Citation Index (SSCI) were searched for English peer review literature using the keywords “health technology assessment” OR “value assessment” AND “pharmaceuticals” OR “methodologies”. Reference lists of the selected studies were screened and the HTA bodies' websites were searched for any published guidelines. Material on the “type of evidence and evaluation criteria considered in HTAs” was collected and analysed, along with material on the “responsibilities and structure of national HTA models and processes”, “methods and techniques applied in HTA”, and “outcomes and implementation of HTAs”. The identified value dimensions sourced from the “evidence and evaluation criteria” component of the review would form the fundamental domains of the model, shaping into top-level criteria groups (i.e. clusters) and informing their decomposition into lower level criteria, which comprised the core structure of the value tree.

Second, literature findings were supplemented with expert consultation, where national HTA experts from the countries of interest were invited to review and validate the results. This took place because, upon early communication of the preliminary results with the partners of the Advance-HTA project consortium (Advance-HTA, 2013), it became evident that in a few cases the evidence from the peer review literature may have been outdated and, in some cases contradictory, not reflecting actual practices. As a result, the findings of the systematic literature review were

validated with national experts from the countries in question.

Third, we incorporated findings from other relevant literature on health care evaluation, including grey literature, to identify value concerns of decision-makers that might not be reflected as part of the current or formal HTA evaluation criteria in place. We considered studies on the benefit-risk assessment of new drugs from a licensing perspective (EMA, 2007, 2008, 2011, 2012; FDA, 2005; Mussen et al., 2007a; Mussen et al., 2009; Mussen et al., 2007b; Phillips et al., 2011; Walker et al., 2009), value based pricing and assessment from a payer perspective (DH, 2010, 2011; Miners et al., 2013a, 2013b; NICE, 2011, 2014a,b; OHE, 2011), and patient access to orphan drugs from a social responsibility perspective Working Group on Mechanism of Coordinated Access to Orphan Medicinal Products, 2013; Hughes-Wilson et al., 2012). The findings from this step supplemented the lower level criteria and their decomposition into bottom-level sub-criteria or attributes.

In the fourth stage, the emerging structure of the value tree and its criteria were subjected to a detailed consultation with 28 HTA experts who provided feedback on the comprehensiveness of the model, but also on its perceived usefulness and practical limitations, as part of the Advance-HTA project (Advance-HTA, 2013). These experts were selected because they acted as partners, Scientific Advisory Board members or affiliated scientists in the Advance-HTA project and included health care professionals (e.g. clinicians, nurses, pharmacologists), methodology experts (e.g. health economists, HTA experts, statisticians) patient representatives and policy-makers/regulators. They were affiliated with a range of international organisations with focus on health economics and/or HTA, including nine academic-research institutions, four HTA bodies, one HTA research network, one coordinating patient and health care professional organisation and one

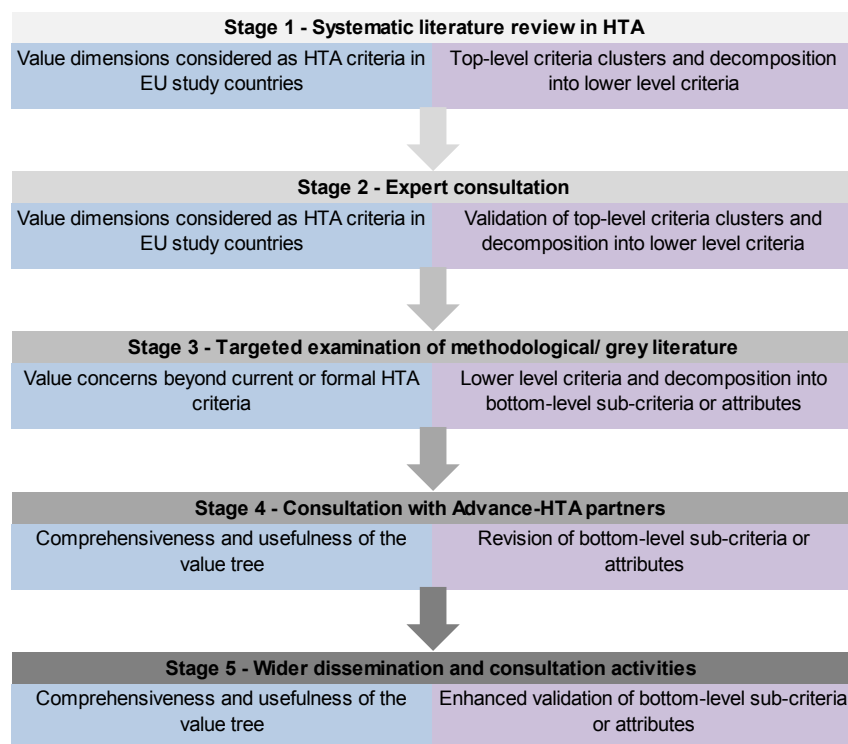


Fig. 1. Flowchart of the five-stage model building process for structuring the generic value tree.

Caption: The left hand side of the flow chart (blue coloured) indicates the type of evidence collected and tasks involved as part of each stage; the right hand side of the flow chart (purple coloured) indicates the respective model input resulting from each stage. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

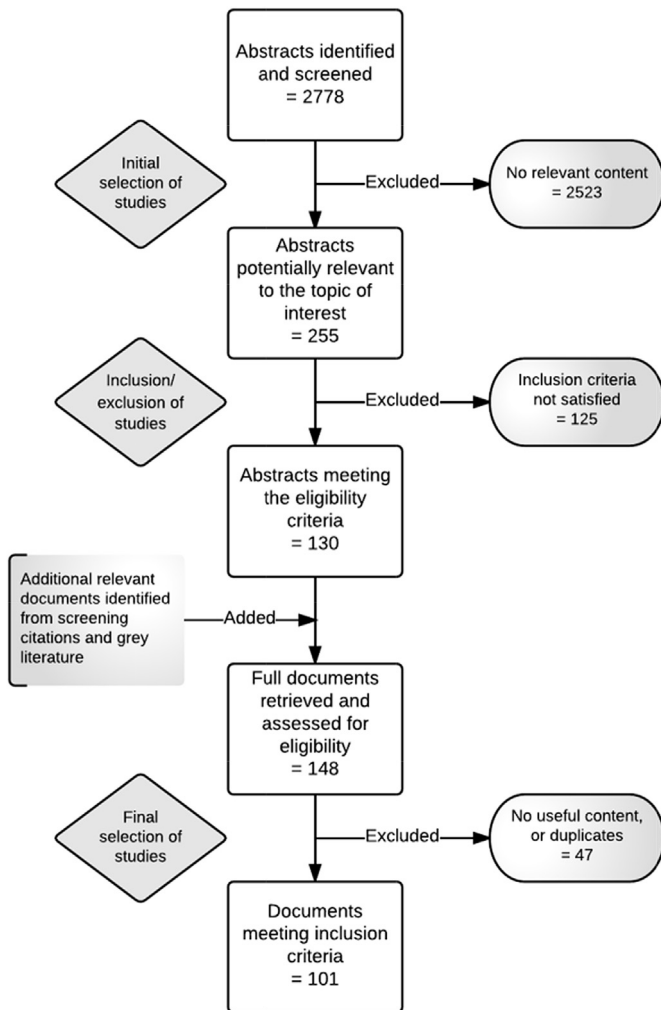


Fig. 2. Flow diagram of the systematic literature review stage.

international public health organisation (see Appendix 1). Based on the feedback received from these experts, the structure of the value tree was revised in an iterative manner, mainly informing the bottom-level sub-criteria or attributes.

Finally, following a series of dissemination activities involving presentations and seminars as part of capacity building workshops, feedback was collected from a wide range of stakeholders (mainly decision-makers from ministries of health, health insurance organisations and HTA bodies) and key opinion leaders across settings beyond the eight countries identified originally, where the aim was to validate results and further enhance the encompassing nature of the model by capturing additional expert and wider geographical perspectives (Angelis and Kanavos, 2014, 2015; Kanavos and Angelis, 2014, 2015). Specifically, the value tree was disseminated at four workshops attended by a total of 230 participants from the Latin American (November 2014 and September 2015) and Eastern European regions (September 2014 and 2015), in order to capture perspectives from low- and middle-income countries via smaller focus groups of 10–15 participants.

Throughout the five stages of the model-building process, the various pieces of evidence collected to inform criteria selection and their structuring into a value tree took place in alignment with decision theory principles, aiming to satisfy the required criteria properties and ensure that the model is rigorous and the analysis outcome robust (Franco and Montibeller, 2010). For example, an

additional criterion would be added if a value concern was not captured in the initial set of criteria so that all the essential value concerns of decision-makers could be addressed. Further, if a particular criterion was perceived to reflect the same value concern as another criterion, one of them would be removed in order to avoid double counting. Similarly, two individual criteria would be replaced if their underlying concern could be reflected from a single criterion on the basis of conciseness, as only the smallest number of criteria required for evaluation should be included for simplicity. Individual criteria would be replaced with other criteria if their meaning was not clear in order to improve comprehension; similarly, individual criteria would be replaced if their measurement was not possible, in order to ensure high levels of functionality. Finally, if by assessing the value of one criterion it became evident that knowledge on the performance of another criterion would be required, then the two criteria would be aggregated into a single one in order to avoid preference-dependence.

The value tree aims to capture a comprehensive generic set of value concerns that can be adapted to different decision-making contexts, problems, indications or treatments. As a result, following the completion of the model-building phase involving the precise attribute definition and the selection of alternative treatment options for the case of specific decision problems, some of these criteria might not satisfy the required properties, in which case the underlying issues should be addressed with caution. For example, following the operationalisation of two criteria with specific attributes it might become evident that there is possible double-counting between them, in which case one of the two would have to be excluded. A few cases in which the theoretical properties of some criteria could be put into question have emerged following consultation with experts; these are discussed in the respective sections of the criteria below.

The five-stage process took place between February 2013 and June 2016.

4. Results - The Advance Value Tree

We briefly present the findings from the systematic literature review and expert consultation in HTA, which acted as the first and second stages of the model-building process, mainly because it informed the primary identification of value dimensions which established the core structure of the value tree. Then we outline the completed value tree and discuss the logic behind its various components, including the overall criteria grouping and decomposition from top-level criteria clusters into lower level criteria and bottom level sub-criteria or attributes. Finally, we briefly suggest how to deal with quality of evidence, and, more specifically, with clinical validity concerns.

4.1. Primary identification of value dimensions: findings from the systematic literature review and expert consultation in HTA

In total, 2778 article abstracts were screened. Out of these, 255 articles were selected to be read in full due to their potential relevance and 101 articles were ultimately used. The main groups of value dimensions that were found to be considered as evaluation criteria among the study group of European countries, identified through the first and second stages of the model-building process included: (a) burden of disease, (b) therapeutic impact, (c) safety profile, (d) innovation level, (e) socioeconomic impact, (f) efficiency considerations, and (g) others (Angelis et al., 2017a). Individual value dimensions falling under the first five groups of evidence (a – e), together with their intensity of use by each country are shown in Table 1; efficiency considerations and other types of concerns were excluded because they would contradict the necessary theoretical

properties of criteria as explained below. Based on the available evidence, these five clusters of value dimensions were perceived to comprise the critical aspects of value concerns to decision-makers for assessing the value of new medicines in the context of HTA, providing the core foundation of the value tree.

4.2. Incorporating the value dimensions into a generic model: The Advance Value Tree

Ultimately, the resulting generic value tree spans three levels of evaluation criteria (top-, middle- and bottom-level), where top-level groups of criteria (i.e. criteria clusters) are decomposed into middle-level criteria and bottom-level sub-criteria or attributes, relating to the five key value domains outlined above that can be explicitly measured and assessed: (a) burden of disease (BoD), (b) therapeutic impact (THE), (c) safety profile (SAF) (d) innovation level (INN), and (e) socioeconomic impact (SOC). With the exception of the BoD cluster, which relates to the disease or indication of interest, the remaining four clusters relate to the impact or characteristics of the medicine. The hierarchical representation of the three levels of evaluation criteria forms the different components of the value tree, which we called “Advance Value Tree” (Fig. 3). It should be noted that in certain evaluation contexts, the hierarchical organisation of these value concerns could be perceived to represent a combination of objectives and indicators rather than a value tree in a strict sense.

Although additional types of value concerns might exist falling under other categories (Guindo et al., 2012; Mirelman et al., 2012; Tanios et al., 2013), such as efficiency (e.g. cost-effectiveness), equity (e.g. priorities, fairness, ethics) and implementation complexities (e.g. organisational, skill and legislative requirements), they are not included as criteria because they would contradict

with the desired criteria properties and the adopted scope of “value”. Strictly speaking, these dimensions do not represent intrinsic disease-technology characteristics or impacts on patient outcomes and health system resources but, instead, capture extrinsic characteristics and impacts that generally depend on the value tree’s core variables, or other features of health care systems reflecting health system’s goals and building blocks (Tromp and Baltussen, 2012).

For example, the concept of efficiency comprises two components, i.e. cost and benefit, with the latter already reflected in the model and thus its inclusion would violate the principle of non-overlap and would result in double-counting. Concerns relating to equity, implementation complexities, and other characteristics of the overall health systems’ context (e.g. stakeholder pressure and political power), are usually of subjective nature and not easily quantifiable, therefore becoming difficult to operationalise. All these extrinsic value dimensions do not relate to the “value” of a new medicine *per se*, but instead depend on the particular settings of the health system under consideration. They could therefore be considered or incorporated on an optional basis as needed for the particular decision context and problem in question, possibly through the application of other analytical frameworks in parallel (EUnetHTA; Goetghebeur et al., 2008).

Examples of iterations in the value tree that have resulted from consultations with experts include the aggregation of “safety” and “tolerability” criteria into a single “safety & tolerability” criterion when a potential overlap between their measures became evident following discussions with clinicians (described in the Safety Profile section below), or the addition of a “Carer” sub-criterion under the “Indirect costs” criterion to capture the wider socioeconomic impact of a treatment following discussion with patients.

Table 1

Value dimensions considered as evaluation criteria among the study group of European countries and their intensity of use.

	France	Germany	Sweden	England	Italy	Netherlands	Poland	Spain
Burden of disease								
Severity	***	**	**	**	*	**	**	**
Availability	***	*	*	***	*	**	*	**
Prevalence	*	**	*	*	**	**	**	**
Therapeutic								
Direct endpoints	***	***	***	***	***	***	***	***
Surrogate endpoints	**	**	**	**	**	**	**	**
Safety								
Adverse events	***	***	***	***	***	***	***	***
Tolerability	**	**	**	**	**	**	**	**
Contraindications	**	**	**	**	**	**	**	**
Innovation								
Clinical novelty	***	*	*	*	**	**	***	**
Nature of treatment	***	*	*	**	X	*	***	**
Ease of use & comfort	*	*	**	*	X	*	X	*
Socioeconomic								
Public health	**	**	*	**	*	***	***	*
Budget impact	*	***	**	***	**	**	***	**
Social productivity	*	**	***	**	*	**	*	**

Caption: Differences in the intensity of use of the different value dimensions as evaluation criteria: Three stars (***) denote the highest intensity of use in place, i.e. “mandatory/formal/explicit/planned/directly measured/grading system available”. Two stars (**) denote a medium intensity of use in place, i.e. “recommended, informal/implicit but planned, formal/explicit but ad-hoc/indirectly measured” etc. One star (*) denotes the lowest intensity of use in place, i.e. “optional/informal/implicit/ad-hoc/indirectly measured/no grading system available”. No stars (x) denote that the value dimension is not considered in any way as an evaluation criterion.

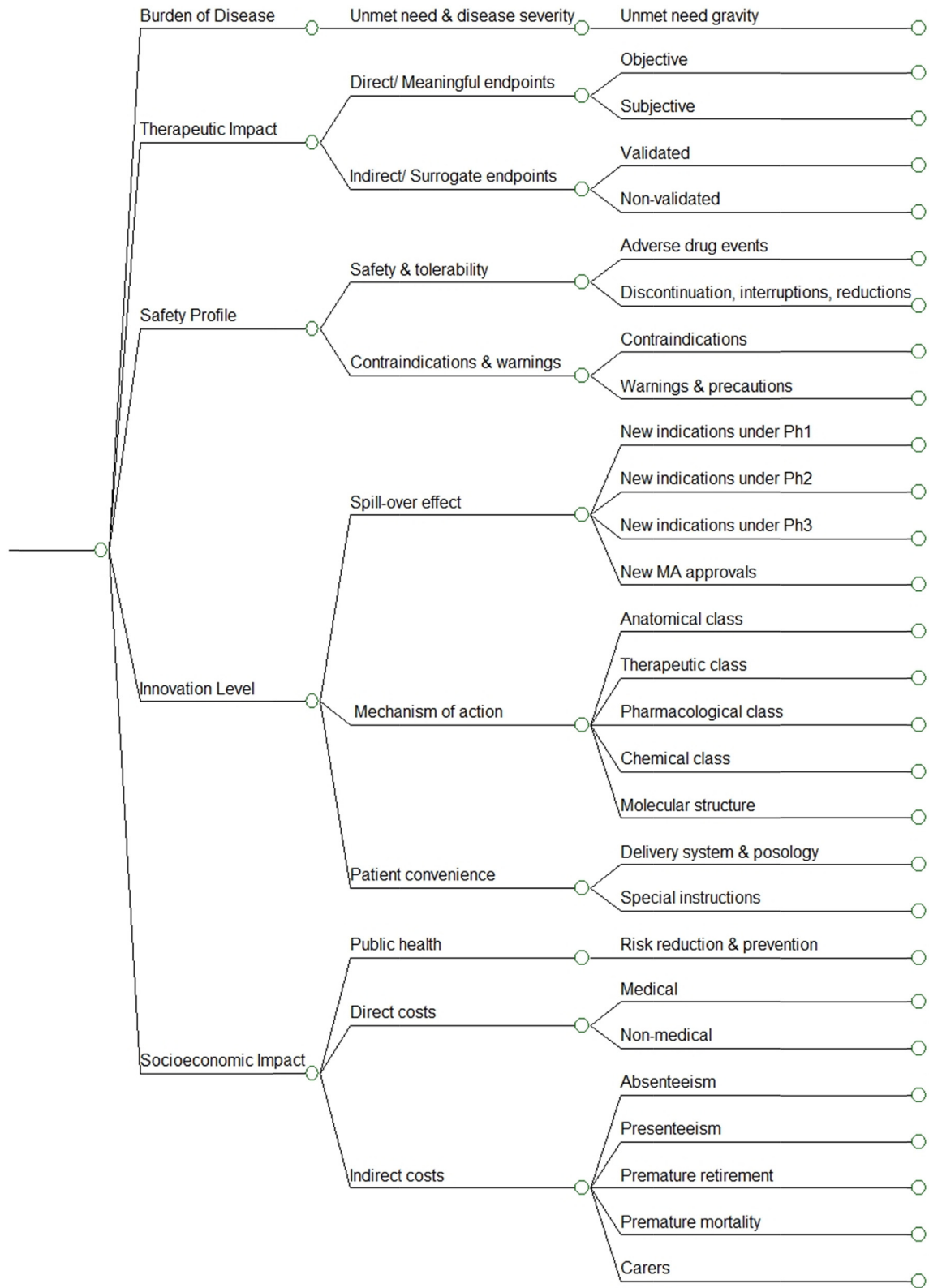


Fig. 3. The Advance Value Tree for new medicines evaluation.
Caption: Hierarchical decomposition of top level criteria clusters, to middle level criteria and bottom level sub-criteria and attributes (from left to right), across five generic value domains (from top to bottom).

4.2.1. Burden of disease

Burden of Disease (BoD) forms a special set of value dimensions as they do not relate to the medical technology itself, but to the disease it is indicated for and, as such, could encompass the severity and unmet need of the disease the treatment addresses (Linley and Hughes, 2013; Miners et al., 2013a). Severity of the disease relates to the condition's degree of seriousness in respect to mortality and morbidity-derived disability, which could be defined on the basis of disability-adjusted life years (DALYs) lost (Murray and Lopez, 1996), or the expected remaining life years adjusted for their quality of life (Hansson et al., 1994). Unmet need reflects the availability of treatments, essentially the degree to which there are existing treatments (DH, 2010) and could relate to methods of diagnosis, prevention or treatment (European Commission, 2006). Because these dimensions are preference-dependent (i.e. in order to assess the impact of unmet need one needs to consider the severity of the disease) they are operationalised through the use of a single aggregated attribute which could be defined as the gap between the health status that patients with a particular medical condition can attain using existing medical interventions and the health status they could expect if they did not have that medical condition, or in other words “the number of QALYs lost by a patient because of their condition” (Miners et al., 2013a), i.e. the unmet need gravity attribute. Therefore, the BoD attribute reflects the difference in the years of life remaining (adjusted) with the respective health-related quality of life (HRQoL) for patients receiving existing technologies, versus the years of life remaining (adjusted) with the respective HRQoL for healthy individuals (of the same age). The larger the difference or the gap between the two (i.e. diseased vs. healthy states), the higher the disease burden. It should be noted that ideally the “existing” medical intervention or treatment should not be one of the options being assessed so that there is no double counting with the “Therapeutic impact” cluster.

The size of the population being affected by the disease, and the possible differentiation of a disease on the grounds of high prevalence or low prevalence, is not taken into account because it could be perceived as unethical. The justification of a special status of a disease based on its prevalence would be questionable, as it entails valuing one disease differently to another because they are a more common vs. a less common disorder (McCabe et al., 2005). The only justification for providing special status to rare diseases would be on equity grounds, using the rationale that “patients suffering from rare conditions should be entitled to the same quality of treatments as other patients” (European Parliament, 2000). However, this concern is already addressed through the unmet need criterion and the associated (un-) availability of effective treatments; consequently, the inclusion of a criterion for capturing “population size” would essentially lead to double-counting and, therefore, is not used. Valuing a condition *more* or *less* as a result of prevalence would be incompatible with equity principles and theories of justice (McCabe et al., 2005).

4.2.2. Therapeutic impact

For most medical conditions there is a set of multidimensional health outcomes that need to be jointly used to capture overall patient benefit including survival, functional status, sustainability of recovery and others, including complications (Porter, 2010). Outcomes are distinct from biologic indicators, the former usually relating directly to health status in contrast to the latter which act as predictors of results.

The “Therapeutic impact” cluster captures clinical benefit by measuring both direct outcomes and indirect indicators relating to the efficacy and/or the effectiveness of an intervention with the view to reflecting the health status but also disease recovery, progression or prevention (complication outcomes are considered

separately in the “Safety profile” cluster discussed in the next section).

In order to distinguish between direct outcomes and indirect indicators, based both on the literature and subsequent expert consultation, “Therapeutic impact” criteria have been divided into (a) direct, clinically meaningful endpoints and (b) indirect, surrogate endpoints respectively; the latter are used as substitutes for direct endpoints, and are usually disease-specific in nature (e.g. HbA1c for complications in diabetes mellitus, prostate-specific antigen (PSA) for prostate cancer, blood pressure for cardiovascular disease). Specifically, a direct outcome endpoint is defined as a “characteristic or variable that reflects how a patient feels, functions, or survives” while an indirect indicator endpoint is defined as a “biomarker intended to substitute for a clinical endpoint” which is “expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence” (Biomarkers Definitions Study Group, 2001) (page 91); in turn, the definition of a biomarker states that it is a “characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” (Biomarkers Definitions Study Group, 2001) (page 91). However, both types of endpoints are useful and, for specific indications, a biomarker (i.e. indirect indicator) might be more directly related to real health status than self-reported outcomes (i.e. direct outcome). Below we propose a further decomposition of direct meaningful endpoints and indirect surrogate endpoints, once again guided by the five-stage evidence collection process.

4.2.2.1. Direct meaningful endpoints. Direct clinically meaningful endpoints can be divided into objective and subjective endpoints (Sullivan, 2013) and, possibly, other health related outcomes (Spahn, 2003). Objective direct endpoints mainly refer to survival, disease exacerbation/alleviation and clinical events (Pazdur, 2008). The availability of evidence on different endpoints depends on the study designs adopted by the respective clinical trials for evaluating the clinical benefit of interventions and, therefore, would depend on their type or nature.

Subjective direct endpoints mainly relate to HRQoL and disease symptoms. HRQoL embraces the broader concept of *health* that includes physical, emotional and social wellbeing (World Health Organisation, 1948), by including both personal health status and social wellbeing (Guyatt et al., 1993) which are usually “subjectively” assessed through patient reported outcome measures (PROMs). HRQoL is multidimensional (Chen et al., 2005; Cherepanov et al., 2010) and can generally be measured through the use of generic PROMs that provide a summary of HRQoL attributes through the production of health utilities (Guyatt et al., 1993), such as SF-36 and EQ-5D instruments. Generic PROM instruments do not target specific population groups and for the case of particular disease states they might not be sensitive enough to adequately capture the impact across all HRQoL dimensions in which case disease specific instruments might be needed. From a societal perspective all of these dimensions should be considered both for the cases of patients and carers.

Importantly, following consultation with experts, it became apparent that the assessment of particular combinations of objective and subjective endpoints as, for example, overall survival (OS) and HRQoL (e.g. through EQ-5D) might be preference-dependent. For example, in a metastatic cancer setting, stakeholders highlighted that their preferences for the performance of different therapies in terms of OS would be meaningful only if the HRQoL performance was also known for the same therapies. Where preference dependence is evident (and assuming a linear additive model is used), the dependent attributes should be combined into a

single attribute, as for example QALYs, essentially aggregating them into a common attribute capturing both OS and HRQoL dimensions.

4.2.2.2. Indirect surrogate endpoints. Indirect surrogate endpoints can be divided into validated and non-validated. Validation of a surrogate endpoint is the process of retrospectively linking it to the actual clinical endpoint (or outcome), i.e. demonstrating a relationship between the two by evaluating how well the surrogate endpoint predicts the clinical outcome of interest ([Biomarkers Definitions Study Group, 2001](#)). Over and above the existence of a strong statistical correlation between the surrogate and the clinical endpoint that is required in order to attain an accurate prognosis of the clinical outcome, i.e. “individual-level surrogacy”, there is also a prerequisite for clinical correlation through the existence of biologic plausibility, i.e. scientific evidence on the causality between the disease, surrogate and outcome; importantly, there is also a need for demonstrating “trial-level surrogacy” which refers to the correlation between a change in the surrogate and a change in outcome due to a therapeutic intervention ([Buyse et al., 2000](#)). Although trial-level data are usually coming from multiple trials or units for the same type of intervention, for a surrogate to be validated it means that its “validity is generalizable to include other interventions that affect the surrogate endpoint” ([Biomarkers Definitions Study Group, 2001](#)) (page 93). As a result of these requirements, surrogate endpoints are rarely validated. Generally, most only manage to predict the clinical endpoint or outcome of interest whilst others fail to predict the clinical endpoint and are used simply as a measure of biological activity ([Fleming, 2003](#)).

4.2.3. Safety profile

The safety profile of an intervention comprises information relating to the degree of its safety and toxicity for the indicated patient population of interest. It is usually measured through safety and tolerability but can also be reflected through any contraindications for its use and special warnings (and precautions) for particular sub-populations.

4.2.3.1. Safety and tolerability. Safety has been traditionally reflected through the incidence of adverse events. An adverse event (AE) refers to an adverse outcome occurring when a drug is administered to a patient or at some time afterwards, in which case the drug might or might not be the cause of the AE ([Aronson and Ferner, 2005](#)). If such outcomes can be attributed with some degree of probability and through a causative link to an action of the drug, then they are known as adverse drug events (ADEs).

The magnitude of both AEs and ADEs is measured through the combination of their seriousness and their frequency (or probability of taking place). Given that seriousness and probability of adverse events seem to be preference dependent, an aggregated attribute reflecting “seriousness and frequency” would be needed. Seriousness could be operationalised through the use of the Common Terminology Criteria for Adverse Events (CTCAE) classification, which contains five grades of adverse events ranging from mild (Grade 1) to death (Grade 5), with three grades in between (moderate, severe, life-threatening) ([National Cancer Institute, 2009](#)). However, for cases that OS is already incorporated in the value tree under the therapeutic impact cluster, then considering Grade 5 adverse events which relate to deaths would constitute double counting and should be excluded. In turn, given that the distinction of adverse events between mild versus moderate and between severe versus life-threatening is often subjective, they could be aggregated and categorised as “non-serious” and “serious”, the former comprising Grades 1 and 2, and the latter comprising Grades 3 and 4 adverse events. Frequency could be operationalised through the use of absolute incidence (in percentage terms), or,

alternatively, using a frequency classification system such as the Naranjo scale which expresses frequency in terms of definite, probable, possible or doubtful ([Naranjo et al., 1981](#)).

Tolerability refers to the overall ability of the patient to tolerate the intervention, mainly in regards to bearing and enduring any adverse events. It is usually reflected through the variables of treatment discontinuation and treatment interruption or reduction, measured either as the proportion of patients discontinuing the treatment (or interrupting/reducing its dosing) or the time to treatment discontinuation (TTTD) from the treatment.

Importantly however, as it became evident following consultation with clinical experts, in the case of discontinuation (or interruption or reduction) of a treatment being due to the incidence of known ADEs, then incorporation of both measure types could lead to double-counting and therefore caution would be needed to ensure the most appropriate of the two criteria is chosen.

4.2.3.2. Contraindications, special warnings and precautions.

Contraindications usually refer to risk and safety factors that act as a reason for an intervention not to be used by a patient, therefore having an impact on the number of potential patients using and benefiting from the intervention. Strictly speaking, contraindications can be categorised as absolute and relative. In the former categorization there are no circumstances under which the patient might use the intervention, compared to the latter where the risk might be outweighed in favour of other considerations (e.g. x-rays for a pregnant woman with a risk of having a child with birth defects). Special warnings and precautions for use are designed with the view to notifying potential risks associated with the use of the intervention in regards to specific patient sub-populations with particular characteristics. These characteristics mainly relate to the administration of concomitant medication, the coexistence of other accompanying diseases and the presence of idiosyncratic patient pathological features that, as a result, might influence the expected action of the drug; as a consequence, caution in the form of careful monitoring is usually recommended.

In reality, some patients with known contraindications (or special warnings and precautions) for a treatment may receive it, in which case incorporation of any evident ADEs or tolerability consequences that can be related to them in the value tree, would lead to double counting between the two and should therefore be addressed with caution, similar to the case between ADEs and discontinuation.

4.2.4. Innovation level

“Innovation” in the context of medical technologies, ranging from biopharmaceuticals and diagnostics to medical devices, is a complex concept lacking a universal consensus ([Kesselheim et al., 2013](#)). From a patient perspective, “innovation” mainly relates to “therapeutic innovation” requiring *novelty of effectiveness*: value needs to be created by generating improved health outcomes that were previously unattainable, the degree of which could be assessed through the combination of the significance of the unmet medical need the intervention addresses and the extent to which it improves the health outcomes for that need ([Morgan et al., 2008](#)).

Below we discuss secondary innovation dimensions over and above “therapeutic innovation”, given that the significance of unmet medical need and the extent of health outcomes improvement are captured under the “Burden of Disease” and “Therapeutic impact” clusters respectively. These dimensions include (a) medicine’s mechanism of action, (b) spill-over effects, and (c) patient usefulness (i.e. convenience).

4.2.4.1. Mechanism of action. The innovativeness of new medicines can be differentiated according to their type or nature, based

on whether they offer a novel or target-specific mechanism of action, or whether they act through more typical or general mechanisms of action. A relatively practical and objective way of describing this performance would be through the use of WHO's Anatomical Therapeutic Chemical (ATC) Classification System. In the ATC classification system, the active substances are divided into different groups according to the organ or system on which they act and their therapeutic, pharmacological and chemical properties. Moving from general to specific, medicines are classified in groups using five different levels relating to anatomical, therapeutic, pharmacological, chemical and molecular dimensions which could act as the respective criteria. Medicines are classified into fourteen main groups (1st level), with therapeutic subgroups (2nd level). The 3rd and 4th levels are pharmacological/chemical subgroups and the 5th level relates to the molecular substance.

Based on this classification, a medicine characterised by a novel therapeutic action, could be perceived as more innovative than a medicine with a novel pharmacological action. In other words, the broader the level at which the medicine differentiates as “original” compared to current existing alternatives, the more innovative the nature of that medicine would be. As a result, the medicine's relative market entrance in regards to the different innovation subgroups (i.e. ATC levels) could act as the respective attributes, in order to reflect, for example, whether it is first-in-chemical-class (i.e. first entrance of a technology at level 4, chemical subgroup level), second-in-pharmacological-class (i.e. second entrance of a technology at level 3, pharmacological subgroup), and so on.

4.2.4.2. Spill-over (dynamic efficiency). Any type of innovation can have R&D spill-over effects that can lead to the development of subsequent innovation(s), entailing a certain degree of diffusion of scientific knowledge and/or technical know-how. Innovation “spill-over effects” could be defined as “the R&D positive externalities that can lead to the development of subsequent innovation(s)”, thus essentially relating to dynamic efficiency in regards to long-term product innovation at future market conditions. These effects could take the form of internal (within the innovator) or external (outside the innovator) effects (OHE, 2005). As Lipsey and Carlow have argued, “major radical innovations never bring new technologies into the world in a fully developed form”, but “appear in a crude and embryonic state with only a few specific uses” (Lipsey and Carlow, 2000). Instead, successive improvements are accumulated through the processes of “learning by doing” and “learning by using” (Mestre-Ferrandiz et al., 2012), the former referring to the improvement of workers' skills at the manufacturing level (Arrow, 1962), while the latter relating to enhancements of knowledge at the level of utilisation by the final user (Rosenberg, 1982). Subsequent to the market entry of a new medicine, new uses (i.e. new indications) for the same medicine could be uncovered following its investigation in patients with other diseases, or next generation medicines could be successively developed either for the same disease indication or a different one. New scientific knowledge and/or technical know-how could be diffused into other contexts/sectors, leading to innovations therein. Because the impact on future innovation is uncertain, it would be practically challenging if not impossible to be predicted and assessed at a single point in time (Rosenberg, 2001), as for example the rate at which performance improvements take place and the speed at which new uses are discovered. In reality, a significant portion of any such spill-over effects will need time to materialize requiring the original innovation to be in the market for a period of time.

Although these effects can relate to new uses for the same technology, new technologies for the same use, or new

technologies for new uses, only the first ones are considered here because the latter two are not operational. We choose to exclude the latter two because it would be very difficult, if not impossible, to identify whether a new technology has been developed following the R&D process of another technology, hence establishing the causal link required would be challenging. For example, in the case of a new technology developed by the same manufacturer of the parent technology, the follow-on technology could be the one driving the development of the first technology, but regulatory delays caused one to be marketed before the other. This criterion would refer to the extent to which the medicine has a spill-over effect in the context of expansion into new indications and could be operationalised by examining the number of new indications for which the medicine is investigated at each stage of clinical development (e.g. Phase I, Phase II, Phase III, Marketing Authorization phase).

4.2.4.3. Patient usefulness (ease and convenience). Aspects relating to patient usefulness would form another group of innovation-related dimensions. The satisfaction of patients with medical care has been shown to correlate with compliance, i.e. the willingness and ability to follow health-related advice including adherence to prescribed medication, while also acting as a predictor of future compliance (Becker and Maiman, 1980; Nagy and Wolfe, 1984; Smith et al., 1987). Satisfaction can be categorised into a range of different domains, including symptom relief/efficacy, side effects, ease and convenience, impact on HRQOL, general satisfaction, and other domains specific to the given question (Shikhar and Rentz, 2004). Given that most of these domains are associated with the health outcomes of the treatment, only ease and convenience are considered here, essentially relating to mode of administration, dosing schedule, medication restrictions, and product-specific designs. Poor compliance has been reported as the most common cause of non-response to medication, with evidence supporting that better health outcomes are produced when patients adhere to treatment recommendations compared to those who do not (Murphy and Coster, 1997). Therefore, any improvements on the above ease and convenience aspects could be translated into greater patient satisfaction and possibly better treatment compliance, which could lead to improved health outcomes, either through an increase in clinical effectiveness or due to a decrease in adverse events. In turn, individualising treatment and minimising its complexity has been proposed in order to encourage adherent behavior (O'Brien et al., 1992); for example, once-daily dosing has been suggested as an important part of enhancing compliance, patient convenience and regimen simplification in hypertension (Rudd, 1995), among others.

Given that patient health outcomes are already incorporated in the value tree, it would be expected that it is necessary to assume an explicit disconnect between better patient satisfaction and improved health outcomes in order to avoid double-counting. This theoretical rationality would support the notion that greater satisfaction through improved ease and convenience is a critical value dimension of new medicines in its own right, contributing to product novelty. Nevertheless, such a strict dissociation between satisfaction and health outcomes might not be required in the first place because of the gap between experimental efficacy data and real world effectiveness (see Appendix 2).

Specific aspects of improved ease and convenience can include less invasive delivery systems, improved posology due to reduced dosing frequency or duration of administration and abolition of special administration instructions. An altered delivery system could involve a change in dosage form (e.g. tablet, liquid, spray, gel) or a change in route of administration (RoA) (e.g. oral, subcutaneous, transdermal), but because these two variables are almost

always correlated (e.g. a tablet will most probably administered orally, and a gel will most probably administered transdermally), both aspects could be captured through a common “delivery system” attribute reflecting the combination of dosage form and RoA. Differences in posology could be captured by an attribute reflecting the combination of dosing frequency in a given time period and treatment duration of each dose. However, following consultation with patients it became apparent that in order to assess a treatment’s “posology”, one might need to consider the treatment’s “delivery system”, therefore suggesting for preference dependence between the two. As a result, both types of value concern were aggregated into a common “delivery system and posology” criterion. Special instructions could be assessed by an attribute reflecting the existence of any guidelines accompanying the administration of the treatment (e.g. with vs. without food, crushing tablets, among others). Alternatively, the medication regimen complexity index (MRCI) could be used as an all-inclusive proxy attribute to operationalise the assessment of ease and convenience as it incorporates a number of medication regimen aspects, including dosage forms, dosage frequency and administration instructions (George et al., 2004).

4.2.5. Socioeconomic impact

Socioeconomic impact dimensions are used to incorporate any other concerns or benefits in the wider context, which mainly relate to (a) public health and (b) economic considerations.

4.2.5.1. Public health impact. Public health impact is primarily associated with any risk reduction in transmitting and developing the disease under consideration or any other disease within the broader population, thus reflecting a societal dimension of prevention. The levels of risk reduction could range from no risk reduction, to reduction of prevalence risk factors, to reduction in transmission, to prevention and prophylaxis from the disease (Goetghebeur et al., 2008), and as a result such a variable would be more applicable to the case of infectious diseases and diseases with known risk factors. The emergence and dissemination of resistance among pathogenic bacteria to the available antibiotics used in medical practice, leading to the latter’s drop in effectiveness, would be a good example. A hypothetical new class of antibiotics that can effectively treat a bacterial strain known to have developed antimicrobial resistance, would offer an important risk reduction of the disease within the broader population by “dealing with resistance” and inhibiting its transmission. The added value of such a medicine would materialize through the combination of improved health outcomes (as reflected via the “Therapeutic impact” cluster) and through a risk reduction in disease transmission as reflected via the “Public health impact” dimension in the “Socioeconomic impact” cluster.

4.2.5.2. Economic impact. Economic impact dimensions reflect the economic burden of disease and can be divided into direct costs and indirect costs. Cost of illness studies are used to identify and measure all costs related to a particular disease and although different studies can employ different methodologies and designs they usually adopt a common classification of cost types (Hodgson and Meiners, 1982). Direct costs can be either medical or non-medical; medical costs result directly from the disease treatment and include diagnostic tests, prescription medicines, inpatient care (hospital or physician), outpatient care (physician or ER), nursing home care, rehabilitation care and home health care including any disposable or replacing items (e.g. prosthetic limbs). Direct non-medical costs refer to non-healthcare costs and include costs related to transportation, relocation, household, comfort/rehabilitation items, property alterations and counselling services.

However, including “cost” as a value dimension is prone to

criticism, mainly because criteria should be conceived as attributes of benefit (Claxton, 2013). For that reason “impact on costs” rather than absolute costs are considered, looking at the incremental difference versus an alternative treatment option that could act as a benchmark, while excluding purchasing costs. This benchmark comparator could fall outside the scope of the analysis; for example, in an oncology setting the chosen comparator in question could be palliative care or best supportive care, rather than an active treatment being evaluated as part of the analysis. The purchasing costs of the options could then be incorporated at a later stage of the analysis in order to establish their efficiency by considering their total value in relation to their total cost similarly to a cost-benefit approach.

Indirect costs reflect productivity losses arising from patient absenteeism, presenteeism, premature retirement, and premature mortality (Zhang et al., 2011). In addition, as it became evident following consultation with patients, indirect costs for carers (i.e. caregivers) are an important dimension of productivity losses that should be considered as part of a societal context and can relate either to financial costs (in case they are not already included in the medical and non-medical costs) or time-off work.

4.3. Dealing with quality of evidence

In the context of health care decision-making, several attributes relating to evidence quality could be identified such as adherence to the requirements of the regulator or decision-maker, completeness of reporting according to the regulator’s guidelines, consistency of reporting with the sources cited, relevance of evidence to the context in question and validity of evidence with regards to scientific standards or research methodological guidelines (Goetghebeur et al., 2008).

Focusing on the validity of clinical evidence, which would mainly be relevant for evidence feeding the performance measurement of alternative options under the “Therapeutic impact” cluster, it could be decomposed into internal and external validity characteristics. Internal validity refers to the extent to which an observed effect can be attributed to the intervention under investigation, or, in other words, the extent to which it is free from bias, in contrast to external validity which relates to the generalisability of clinical effect from experimental settings to real world settings (Centre for Reviews and Dissemination, 2009; Eldridge et al., 2008; Rothwell, 2006). Rather than operationalising these evidence quality (i.e. validity) concerns through their incorporation as criteria and attributes in the value tree, the use of “penalty” functions could instead be explored when the clinical evidence used for the alternative options is of different quality. This is because it would not be methodologically robust to incorporate quality of evidence concerns as criteria; doing so would entail that the quality of evidence used to assess the performance of the options across the respective criteria could be compensated by the options’ performance across the same criteria.

A penalty factor could potentially be applied through a function that multiplies it with the performance of the alternative options across the relevant “Therapeutic impact” criteria or their respective value scores, in order to adjust them for their validity. This penalty factor would be tailor-made for the particular decision-making context and defined based on the relative importance of evidence validity as a source of concern to decision-makers. For example, assuming that a weak internal validity is associated with the comparative treatment effect of a new medicine due to inadequate allocation concealment and lack of double-blinding, a “strict” decision-maker might be willing to “discount” the clinical performance of the medicine on some of its clinical endpoints up to x% of their original level. The lower the validity of the evidence, the higher

the impact the penalty functions would have on the performance of the treatments or their respective value scores across the relevant “Therapeutic impact” criteria. However, it should be noted that penalty functions should generally be used with caution as they may be incompatible with the use of an additive value model (unless they are used within the attributes) in which case a multiplicative or multi-linear model should be used for aggregation.

A useful categorization for the quality of clinical evidence is provided by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system, a framework developed for the purpose of producing consistent clinical guidelines in terms of rating quality of evidence and grading strength of recommendations (Guyatt et al., 2008). According to GRADE, *high quality* evidence could be defined as evidence for which “further research is very unlikely to change our confidence in the estimate of effect”, in contrast to *very low quality* evidence for which “any estimate of effect is very uncertain”. In turn, the Cochrane Collaboration tool for risk of bias and the RE-AIM framework could be applied for the assessment of internal and external validity respectively, the findings of which could then be used for the estimation of the penalty factor.

Based on a systematic classification of internal validity sources, the Cochrane Collaboration has developed a tool for assessing the risk of bias in RCTs, for a series of items covering different bias domains and their sources (Higgins et al., 2011). These bias domains are divided into seven different types based on which RCTs can be assessed and rated (as “low”, “unclear” or “high” risk of bias), namely *selection bias* (with sources of bias including random sequence generation, and allocation concealment), *performance bias* (blinding of participants and personnel), *detection bias* (blinding of outcome assessment), *attrition bias* (incomplete outcome data), *reporting bias* (selective reporting) and *other bias* (i.e. anything else).

On the other hand, the RE-AIM framework which was originally developed by Glasgow and colleagues to evaluate the public impact of health interventions, could be partially applied in order to assess the external validity of clinical trials, mainly through the dimensions of reach, adoption, implementation and maintenance (Glasgow et al., 1999).

5. Discussion

5.1. Robustness of the Advance Value Tree

In this paper we developed and proposed a generic model taking the form of a value tree for assessing the value of new medicines in the context of HTA as part of a methodological approach which moves away from traditional economic evaluation, uses decision theory and adopts multi-criteria evaluation methods. Our model consists of a value tree (Advance Value Tree) with three levels (top-level: 5 clusters; middle-level: 11 criteria; bottom-level: 28 sub-criteria or attributes), as shown in Fig. 3.

Conceptual and theoretical advantages emerge from the methodological and empirical process that was adopted to build the structure of the Advance Value Tree. A five-stage iterative model-building process was followed involving extensive literature review and expert consultation and feedback, with the aim of making it as comprehensive as possible, whilst maintaining its flexibility and adaptability to serve the needs of decision-makers for different decision problems. A key advantage of the Advance Value Tree is its strong alignment with decision theory principles, adopting a top-down value-focused thinking approach while paying attention to the required criteria properties.

For the case of specific decision problems and following the selection of particular treatment options, the value tree could be adapted using a bottom-up approach in order to capture the

particular value dimensions of the actual treatment-indication pairs. This would involve the definition of precise attributes, which should again take place in alignment with decision theory principles and would complete the model-building phase of the process.

Following the model-building phase, model-assessment and model-appraisal phases need to be completed. As part of this, the model can be estimated by applying different types of MCDA modelling techniques for the formation of value judgements and elicitation of preferences across the options.

5.2. MCDA methods and selection of modelling techniques to operationalise the Advance Value Tree

Multiple MCDA methods exist of varying complexity that could be used for the estimation of the model. Specifically, an MCDA modelling technique is required for expressing preferences on the performance of each option against each criterion (scoring), while equating the preference units across all criteria (weighting), and combining preferences of individual criteria together into a combined overall preference (aggregating). Therefore, “modelling” in this context acts as an instrument for enabling decision-makers to understand their own preferences, by helping them to construct their perceptions given a set of assumptions, as part of the overall process of identifying the best decision alternative (Belton and Stewart, 2002).

Among the plurality of existing MCDA approaches, they could be categorised into three main groups of methods governed by different “schools of thought”, notably (a) value measurement methods, including multi-attribute value theory (MAVT) and multi-attribute utility theory (MAUT) methods, (b) outranking methods, and (c) satisficing and aspiration level methods as suggested by others (Belton and Stewart, 2002; Department for Communities and Local Government, 2009; Hammond et al., 1999; von Winterfeldt and Edwards, 1986). These three main groups of methods essentially relate to different classifications of preference modelling. Other classifications have also been proposed as there is no consensus on a universal categorization (Diaby and Goeree, 2014; Dolan, 2010; Goodwin and Wright, 1997; Thokala and Duenas, 2012).

The methodological process we have proposed pertains to the value measurement methods category (Angelis and Kanavos, 2016b), mainly because of the multiple decision contexts that they can be applied to and the simplicity of the value judgements required (in addition to limited restrictions imposed by the axioms employed), features which would probably influence their adoption as the most widely used MCDA methods in health care (Marsh et al., 2014). Specifically, we argue in favour of MAVT methods because of their comprehensiveness and robustness (von Winterfeldt and Edwards, 1986), as well as their ability to reduce ambiguity and motivational bias. We would therefore suggest the operationalisation (i.e. estimation) of the Advance Value Tree through a MAVT approach. This is consistent with recent work on good practices for MCDA in Health Care Decisions (Marsh et al., 2016). Yet, the value tree and the incorporated criteria clusters could also be operationalised through other MCDA methods that go beyond the value measurement category (Guitouni and Martel, 1998).

With MAVT methods, value functions for the scoring of the options can be elicited in different ways, mainly by using direct or indirect rating techniques (Belton and Stewart, 2002; von Winterfeldt and Edwards, 1986). In contrast to direct rating techniques, indirect techniques aim to uncover decision-makers' preferences indirectly through a series of questions that involve differences in the attribute scale and their relation to a value scale, producing a transparent valuation relationship across the complete

attribute range, rather than being limited to the performance of the options under consideration. Techniques that explore the magnitude of increments in the attribute scale that yield equal units in the value scale are known as indifference techniques, whereas techniques that estimate points on the attribute scale that act as mid-points on the value scale are known as bisection techniques; both could be regarded as sub-types of indirect techniques. An alternative indirect approach would be “Measuring Attractiveness by a Categorical Based Evaluation Technique” (MACBETH) which considers pairwise attribute comparisons (Bana e Costa et al., 2012; Bana e Costa and Vansnick, 1994). In MACBETH value functions are elicited through comparison of two attribute levels at a time and the use of seven non-numeric (i.e. semantic) categories to express differences in their value (ranging from “no difference” to “extreme difference”). It is perceived as a convenient way to express value judgements by lowering cognitive load, with numerous real world applications (Bana e Costa and Oliveira, 2012; Bana e Costa and Vansnick, 1997).

In the case of decision contexts requiring repeat decisions, indirect MAVT techniques making use of value functions are recommended in order to ensure the efficient and consistent scoring of alternatives as they become available for evaluation (Marsh et al., 2016); however indirect MAVT techniques could also benefit one-off decisions as they would ensure transparency between the performance of a criterion and the respective value preferences.

As part of weighting, trade-offs between criteria are elicited taking the form of quantitative weights to convert criteria value scores into a common value scale (Marsh et al., 2016). Using direct rating techniques of criteria importance, as for example, requiring the distribution of 100 points over attributes to reflect their relative importance, is associated with two, potentially serious, problems: they could produce flatter importance weight distributions instead of ratio estimation (von Winterfeldt and Edwards, 1986), which could lead to an underestimation of trade-offs between attributes (Peacock et al., 2007) and might be insensitive to the attribute ranges used for the performance assessment (i.e. measurement scales) (Gabrielli and von Winterfeldt, 1978).

Attempting to assign “importance” weights without taking into account the attribute ranges (i.e. measurement scales) is known to be one of the most common mistakes in making value trade-offs (Keeney, 1992; Keeney, 2002). Instead, the use of an indirect swing weighting technique for eliciting relative criteria weights is recommended as common practice through the decision analysis literature (Bana e Costa et al., 2012; Keeney and Raiffa, 1993; Phillips and Bana e Costa, 2007; von Winterfeldt and Edwards, 1986). This technique involves making judgements of relative value between changes (i.e. “swings”) from a lower reference level to a higher reference level on each attribute, with each swing (i.e. relative change from a lower to a higher attribute level), being valued between 0 and 100 and the most valuable swing being anchored at 100 (von Winterfeldt and Edwards, 1986). Alternatively, a Discrete Choice Experiment (DCE), which is based on random utility theory (Amaya-Amaya et al., 2008) could be conducted to incorporate a randomness element on responder choices and reflect preference heterogeneity (Marsh et al., 2016). However, the number of criteria would have to be relative small and most DCEs in the field of health care include up to five attributes (de Bekker-Grob et al., 2012). Methodological guidance on the sound design and implementation of DCEs taking the form of good research practice is provided elsewhere (Bridges et al., 2011; Johnson et al., 2013). Both of these techniques better meet the conditions needed in order to treat weights as scaling constants with no algebraic meaning (Marsh et al., 2016), in alignment with decision theory.

Finally, in terms of aggregating, a technique is needed for

combining criteria scores and weights together and, more specifically, for selecting a function that allows the combination of the attributes in consistency with responder (e.g. stakeholder) preferences (Peacock et al., 2007). The application of a simple additive (i.e. linear weighted average) model is the most commonly applied function in health care applications (Marsh et al., 2016), mainly because of its simplicity and comprehensible nature making it easily explained and understood by decision-makers (Belton and Stewart, 2002). However, its use is associated with a number of properties, the most restrictive of which is the existence of preference independence across criteria and attributes, which means that no two or more criteria, or their attributes, can independently have a large impact on the overall benefit of the options, in which case a multiplicative or multi-linear model should be used (von Winterfeldt and Edwards, 1986). Such models are less commonly used as they are perceived to be more complex to populate (Keeney and Raiffa, 1993); additionally, the incorrect aggregation of preferences through these more advanced models can result in considerably greater errors than additive models as evident from empirical evidence of simulation studies (Stewart, 1995).

The precise choice of scoring, weighting and aggregating techniques will ultimately depend on a number of characteristics of the decision-making problem under consideration, in relation to theoretical relevance, level of precision required in the evaluation of the options and cognitive burden posed to stakeholders and decision-makers (Marsh et al., 2016). Deciding on the optimal combination of modelling techniques represents an important topic that requires further research to better understand the impact of different technique combinations on the above issues and the results of the analysis.

Ultimately, results should be examined and sensitivity analysis be conducted, possibly in combination with robustness analysis, in order to validate the model and findings. Deterministic sensitivity analysis can be used to explore the impact of baseline weight (or score) changes on option rankings and address parameter uncertainty. Robustness analysis can be used as part of an n-way sensitivity analysis to test how simultaneous changes in the criteria reference levels would impact the ranking of the alternative treatments. As a result, differences in viewpoints and any disagreements between stakeholders can be resolved, as for example by testing whether the ranking of the treatments might be sensitive or not to a criterion's relative weight.

5.3. Cognitive and motivational biases

Research on behavioural decision analysis has indicated that construction of judgements as part of decision-making is prone to a number of biases, relating to faulty cognitive processes or due to motivations for preferred analysis outcomes (Montibeller and Winterfeldt, 2015). The presence of these biases could be present across the different phases of the methodological process on which the framework is based, and especially in the model-building, model-assessment, and model-appraisal phases, involving the participation of experts and decision-makers. Such biases could potentially reduce the quality of the model and the results of the analysis. For this reason, a number of de-biasing techniques should ideally be applied to overcome these limitations (Montibeller and Winterfeldt, 2015).

Although the precise choice of these techniques would depend on the particular decision problem under consideration and the evidence available, our framework could support several of these. For example, in terms of cognitive biases, *equalizing bias* relating to the allocation of similar weights to all value concerns could be addressed through the hierarchic elicitation of weights as part of swing weighting. *Gain-loss bias* which occurs as alternative

descriptions of a choice and its outcomes (which may lead to different answers) could be addressed by constructing value judgements in relevance to (marginal change from) a best supportive care option, which could be used as a reference level and act as a *status quo*. *Myopic problem representation bias* which occurs when an oversimplified problem is based on an incomplete mental model for the decision problem could be addressed by explicitly encouraging experts and decision-makers to think more about their value concerns in the wider socioeconomic context, as through the value domains captured by the Advance Value Tree. *Omission of important variables bias* could be addressed by using group elicitation techniques so that no important variable is overlooked, for example by using MAVT indirect elicitation techniques as part of a group meeting. *Range insensitivity bias*, occurring when objectives are not properly adjusted to changes in the range of attributes could be addressed by making attribute ranges explicit and using a swing weighting technique. *Splitting bias* which occurs when the structuring of the criteria affects their weights, could be addressed using MAVT indirect elicitation techniques instead of direct rating techniques, together with hierarchical estimation of relative weights such as swing weighting.

In terms of motivational biases, a number of biases could exist because of peoples' emotions, desires and motives (Montibeller and Winterfeldt, 2015): these include *affect influenced bias* relating to the emotional predisposition for or against a specific outcome, *confirmation bias* relating to the desire to confirm one's belief leading to unconscious selectivity, *desirability of a positive event* or *consequence bias* occurring when the desirability of an outcome leads to an increase in the extent to which it is expected to take place (i.e. *wishful thinking* or *optimism bias*), *under-desirability of a negative event* or *consequence bias* occurring when there is a desire to be cautious or conservative in estimates that may be related to harmful consequences, and *desirability of options* or *choice bias* leading to over- or underestimating values or weights in a direction that favours a desired alternative. All of these biases could be addressed by engaging multiple experts with alternative points of view, collecting views from a range of different experts as stakeholders or decision-makers to provide different value perspectives, and using indirect MAVT techniques for scoring the options and weighting the criteria (Montibeller and Winterfeldt, 2015).

5.4. Budget constraints, opportunity cost and prioritisation for resource allocation

There are different perspectives on prioritisation and resource allocation most of which involve the ordering of alternative options as part of an evaluation or appraisal process, followed by choosing the best (or most attractive) combination of options given the available resources as part of a portfolio construction task (Phillips and Bana e Costa, 2007). In cost benefit analysis (CBA), which is based on the principles of social welfare economics, non-monetary benefits are translated into monetary values using willingness-to-pay, with the ratio of costs to benefits used to reflect value-for-money and risk often incorporated through discount rates (HM Treasury, 2003). In decision analysis (DA), decision trees are used to model the consequences and risks of alternative options with uncertainties about future events incorporated through probabilities, producing expected weighted consequences as the basis for ordering the options, which could be divided by their cost to give value-for-money indices (Clemen, 1996). Similarly, MCDA can be used to value the consequences of options, where risks can be even incorporated as criteria rather than probabilities (Keeney and Raiffa, 1976). Although the different perspectives differ on how they address benefits, costs and risks, all agree that the best value-for-money can be obtained through dividing benefits by costs,

essentially reflecting opportunity cost, which should act as the normative basis for prioritisation.

In MCDA, value scores obtained from the model can be used to derive cost-benefit ratios, prioritising the most efficient options with lower cost-benefit ratios until the resources or any budget constraint in place are exhausted (Peacock et al., 2007).

Specifically, as part of health economic evaluations it is usually assumed that the maximum amount to be spent per unit increase in the chosen measure of outcome is fixed at a certain level which is referred to as the societal willingness to pay (WTP) per unit of health gain (Postmus et al., 2014), according to which the optimal option is derived based on the grounds of the highest net monetary benefit (NMB) (Claxton and Posnett, 1996). However, this maximum societal level or threshold of costs to effects ratio depends on the level of the available budget (Weinstein and Zeckhauser, 1973), therefore a decision to reimburse one intervention should in theory be followed by a reallocation of the remaining budget as the coverage of the intervention will influence the maximum ratio depending, for example, on whether it replaces a more expensive intervention (ratio will increase) or a less expensive intervention (ratio will decrease) (Postmus et al., 2014). Given that in practice the relation between the maximum cost per outcome ratio and available budget is often neglected and WTP thresholds are imposed as exogenous parameters (Al et al., 2005), it could be argued that the use of the NMB framework is no longer congruent with the resource allocation perspective it originated from (Postmus et al., 2014). In turn, it has been demonstrated that the NMB framework in practice can still have a theoretical foundation when the NMB function is interpreted as a value function according to MAVT (Postmus et al., 2014). It has been shown that the NMB framework can be seen as a special case of a more general stochastic MCDA approach known as stochastic multi-criteria acceptability analysis (SMAA), which has already been applied for the benefit-risk evaluation of drugs (Tervonen et al., 2011).

5.5. The Advance Value Framework in perspective

The application of MCDA modelling techniques for the estimation of the Advance Value Tree through the construction of value judgements and elicitation of preferences based on the MAVT methodological process we have proposed (Angelis and Kanavos, 2016b), provides a new value framework based on MCDA principles for the purposes of HTA which can inform discussions and negotiations leading to coverage and reimbursement decisions. Overall, the Advance Value Framework (AVF) adopts an encompassing societal perspective, incorporating views from the wider stakeholder community while assuming that the payer is the ultimate decision-maker; however, it allows for the choice of perspective to be adapted to different circumstances and decision-making contexts. The AVF can be used for assessing the value of new medicines through the comprehensive set of evaluation criteria outlined in the Advance Value Tree; in addition to scientific value judgments relating to clinical considerations (i.e. therapeutic impact and safety profile), the value tree allows for the incorporation of social value judgements relating to burden of disease, innovation level and socioeconomic impact, all of which can be captured explicitly.

Finally, the AVF operationalises the Advance Value Tree through the implementation of a combination of MAVT techniques for the elicitation and aggregation of value preferences and can be used as a decision-making tool with the involvement of all stakeholders, being appropriately flexible to meet diverse requirements and readily adaptable across settings.

Table 2 provides a comparative breakdown of the main features of existing value frameworks and the AVF. The development of

Table 2
Comparison of the Advance Value Framework (AVF) with other value frameworks.

Framework	ACC/AHA	ASCO	ESMO	ICER	MSKCC	NCCN	MoCA	AVF
Decision context	Clinical practice	Shared decision making	Clinical practice	Coverage/ reimbursement	Pricing	Shared decision making	Pricing and reimbursement	Health Technology Assessment
Key actor	Physicians	Patients - Physicians	Physicians	Payer	Payer-Provider	Patients - Physicians	Payers - Manufacturers	All stakeholders
Value parameters	-Clinical benefit vs. risks -“Value” (cost- effectiveness)	-Clinical benefit (efficacy) -Toxicity (safety) -Cost (efficiency)	-Variability of estimated Hazard Ratio -Observed absolute difference in treatment outcomes:	-Clinical care value (comparative clinical effectiveness, additional benefits, contextual considerations, incremental cost per outcome) -Health system value (budget impact)	-Dollars per life year -Toxicity -Novelty -Cost of development -Rarity -Population burden of disease, Unmet need, Prognosis	-Efficacy of regimen -Safety of regimen -Quality of evidence -Consistency of evidence -Affordability of regimen	-Alternatives available/ unmet need -Relative effectiveness -Response rate -Degree of certainty	-Burden of disease -Therapeutic impact -Safety profile -Innovation level -Socio-economic impact
Conceptual basis	Stakeholder consultation (writing committee)	Stakeholder consultation (ASCO Value in Cancer Care Task Force)	Stakeholder consultation (ESMO Task Force with input from the ESMO faculty and a team of biostatisticians, followed by the ESMO- MCBS Task Force, the ESMO Guidelines Committee and a range of invited experts)	Stakeholder consultation (input from the ICER Policy Development Group, involving representatives from all major stakeholder groups)	Developed by a team of clinical experts	Stakeholder consultation (NCCN panel members)	Stakeholder consultation (the MoCA working group that was formed by volunteers from a number of EU countries)	Literature review; Stakeholder consultation (Advance- HTA partners and workshop participants, national experts); Decision theory
Strengths	Quality of evidence (LOE) explicitly ranked; class of recommendation given separately and not averaged together with the level/quality of evidence as a single metric	Net Health Benefit score and costs illustrated side by side to facilitate the decision making process of patients by making fully informed decisions	Both the variability of the estimated HR and the observed absolute difference in treatment outcomes are explicitly addressed	Integration of a technology's value-for- money with its budget impact	A range of domains incorporated, both relating to the drug and the disease	Easy and simple to comprehend visual output	Easy to comprehend and practical to use because of its simplicity	Multiplicity of value domains; Assignment of quantitative relative weights; Transparent; Extensive stakeholder engagement; Grounds on decision theory

these value frameworks is an important step towards a more inclusive Value Based Assessment (VBA) process as part of decision-making contexts in health care. An ideal framework should be comprehensive enough in terms of criteria inclusion in order to capture all important value dimensions, while at the same time providing users with flexibility for their selection in alignment with health system-specific needs. The individual value dimensions incorporated should possess a number of technical characteristics for overall value score rankings to be derived (Angelis and Kanavos, 2016b; Franco and Montibeller, 2010; Keeney and Gregory, 2005). Specifically, they should be operational so they can be measured and non-overlapping so that there is no double counting among them. Importantly, for some frameworks it is not clear how to operationalise performance measurement of the alternative options across the different value dimensions considered (so that options can be assessed) or how to mediate trade-offs among them, and in some cases such efforts seem to lack a theoretical basis, not least because they are derived in a non-scientific manner. Although most value frameworks focus mainly on the benefit component of the evaluation process relating to measuring the value of new medicines, issues relating to budget constraints or value for money considerations are crucial to consider in the prioritisation of resource allocation and would be the subject of further research.

6. Conclusion

By using a five-stage methodological iterative approach informed by secondary sources and extensive primary research and consultation, we developed a generic value tree (Advance Value Tree) which is embedded in decision theory. The tree incorporates a number of evaluation criteria that have traditionally been considered in the context of HTA, either explicitly in a systematic manner, or implicitly on an *ad hoc* basis. This work builds on the theoretical foundations of MAVT, influencing the relationship between the different criteria in accordance with a number of critical theoretical properties. We subsequently outlined the assembly of the evaluation criteria in the form of a generic value tree and finally we introduced a number of MCDA techniques for operationalising the value tree into a value framework (The Advance Value Framework) for use by decision-makers and stakeholders. In undertaking the above, we focused mainly on the benefit component of the evaluation process relating to measuring the value of new medicines while also accounting for how it could be used in practice given budget constraints in order to obtain best value-for-money. Future research could aim to test the value framework in practice, possibly through case studies involving specific health interventions, in order to better understand potential advantages and limitations.

Acknowledgements

This study has been prepared as part of Advance-HTA, a research project which has received funding from the European Commission, DG Research under the auspices of Framework Programme 7 (Grant No. 305983). The views represented in the study are those of the authors and do not necessarily reflect the views of the European Commission.

We would like to thank three anonymous reviewers, Joanna Coast, Monica Oliveira and Matteo Galizzi for providing constructive feedback and suggestions for improving the paper. We are grateful to Gilberto Montibeller for sharing his advanced decision sciences expertise and working with us on specific applications of the methodological framework which helped to shape its development. We would also like to thank Huseyin Naci and Olivier

Wouters for their feedback on earlier versions of this paper and Jennifer Gill for excellent editorial assistance.

We are grateful to the following experts for reviewing the systematic literature review results and providing us with useful feedback: Karl Arnberg (Sweden), Karen Brigham (France), Agnese Cangini (Italy), Guillaume Dedet (France & World Health Organisation), Jaime Espin (Spain), Mathilde Grande (France), Kinga Grzywacz (Poland), Patrick Jeurissen (The Netherlands), Douglas Lundin (Sweden), Armando Magrelli (Italy), Dimitra Panteli (Germany), Dominique Polton (France), Pierluigi Russo (Italy), Annalisa Sammarco (Italy), Pedro Serrano (Spain), Giovanni Tafuri (Italy & European Medicines Agency), Anna Zawada (Poland).

Feedback has been received from the participants at the HTAi conference in Oslo (June 2015), the YoungOR19 conference of the OR Society in Birmingham (September 2015) and the PPRI conference of the WHO Collaborating Centre for Pharmaceutical Pricing and Reimbursement Policies in Vienna (October 2015); we are grateful for all comments and suggestions received.

We would also like to thank the participants of the Advance-HTA dissemination and capacity building workshops in Mexico City (November 2014), Santiago (September 2015), and Warsaw (September 2014 and 2015), for providing constructive comments on the development of the value tree and the value framework.

Finally, we would like to thank the partners of Advance-HTA consortium for their contributions in Work Package 2, including their views and opinions during the meetings of the consortium but also for helpful comments and suggestions on earlier version of the paper.

All outstanding errors are our own.

Appendices

1. The partner organisations of the Advance-HTA consortium included, the following academic-research institutions: London School of Economics and Political Science (LSE), London School of Hygiene and Tropical Medicine (LSHTM), Instituto Superiore di Sanita (ISS), Universidad de Castilla – La Mancha (UCLM), Institute za Ekonomska Raziskovanja (IER), Technische Universitaet Berlin (TUB), Escuela Andaluza de Salud Publica (EASP), Universite Paris XII - Val de Marne (UPEC), and University College London (UCL); the following four HTA bodies: NICE International, Agencja Oceny Technologii Medycznych (AOTM), Tandvårds-och läkemedelsförmånsverket (TLV), and Haute Autorité de Santé (HAS); the following HTA research network: European Network for Health Technology Assessment (EUnetHTA); the following coordinating patient and health care professional organisation: European Brain Council (EBC); the following international public health organisation: Pan American Health Organisation (PAHO).
2. The available clinical evidence used for the assessment of therapeutic benefit will in most cases be sourced from explanatory trials that reflect ideal conditions, conducted on a carefully selected group of patients and under strictly controlled environments which could lead to relatively high regimen compliance. In contrast, pragmatic trials are conducted under real world settings and on patients representing the full population spectrum that may show varying compliance (Godwin et al., 2003). Indeed, evidence from electronic monitoring for a range of diseases including hypertension, glaucoma, seizure disorders and others, indicate that good adherence to prescribed regimens is only observed in between 50% and 60% of patients, with 5%–10% adhering poorly and 30%–45% adhering to an intermediate but widely variable degree (Cramer and Mattson, 1991; Gordon and Kass, 1991; Kass et al., 1986; Kruse and Weber, 1990; Rudd, 1995). Consequently, unless clinical

evidence comes from pragmatic trials resembling real world conditions, accounting for the ease and convenience of regimens could act as an adjustment towards more realistic expected clinical outcomes.

References

- Advance-HTA, 2013. Advance-HTA Webpage (www.advance-hita.eu).
- Al, M.J., Feenstra, T.L., Hout, B.A.V., 2005. Optimal allocation of resources over health care programmes: dealing with decreasing marginal utility and uncertainty. *Health Econ.* 14, 655–667.
- Amaya-Amaya, M., Gerard, K., Ryan, M., 2008. Discrete choice experiments in a nutshell. In: Ryan, M., Gerard, K., Amaya-Amaya, M. (Eds.), *Using Discrete Choice Experiments to Value Health and Health Care*. Springer, Dordrecht, The Netherlands.
- Anderson, J.L., Heidenreich, P.A., Barnett, P.G., Creager, M.A., Fonarow, G.C., Gibbons, R.J., et al., 2014. ACC/AHA statement on cost/value methodology in clinical practice guidelines and performance measures: a report of the American College of Cardiology/American Heart Association task force on performance measures and task force on practice guidelines. *J. Am. Coll. Cardiol.* 63, 2304–2322.
- Angelis, A., Kanavos, P., 2014. Value Assessment Methods and Application of Multiple Criteria Decision Analysis for HTA. 1st Advance-HTA Capacity Building Workshop in Eastern Europe, Warsaw, Poland.
- Angelis, A., Kanavos, P., 2015. Using Multiple Criteria Decision Analysis: Conceptual Framework and Model. 2nd Advance-HTA Capacity Building Workshop in Eastern Europe, Warsaw, Poland.
- Angelis, A., Kanavos, P., August 2016a. Critique of the American Society of Clinical Oncology Value Assessment Framework for cancer treatments: putting methodologic robustness first. *J. Clin. Oncol.* 34 (24), 2935–2936. <http://dx.doi.org/10.1200/JCO.2015.64.9673>.
- Angelis, A., Kanavos, P., 2016b. 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, 435–446.
- Angelis, A., Lange, A., Kanavos, P., 2017a Mar 16. Using health technology assessment to assess the value of new medicines: results of a systematic review and expert consultation across eight European countries. *Eur. J. Health Econ.* <http://dx.doi.org/10.1007/s10198-017-0871-0>.
- Angelis, A., Kanavos, P., Montibeller, G., March 2017b. Resource allocation and priority setting in health care: a multi-criteria decision analysis problem of value? *Global Pol.* 8 (Suppl 52) <http://dx.doi.org/10.1111/1758-5899.12387>.
- Antoñanzas, F., Terkola, R., Postma, M., 2016 Dec. The value of medicines: a crucial but vague concept. *Pharmacoeconomics* 34 (12), 1227–1239.
- Aronson, J.K., Ferner, R.E., 2005. Clarification of terminology in drug safety. *Drug Saf.* 28, 851–870.
- Arrow, K., 1962. The economic implications of learning by doing. *Rev. Econ. Stud.* 29.
- Bach, P., 2015. DrugAbacus App. Memorial Sloan Kettering Cancer Center (<http://app.drugabacus.org/abacus-mskcc>).
- Baltussen, R., Niessen, L., 2006. Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Eff. Resour. Alloc. C/E* 4, 14–14.
- Bana e Costa, C.A., De Corte, J.-M., Vansnick, J.-C., 2012. MACBETH. *Int. J. Inf. Technol. Decis. Mak.* 11, 359–387.
- Bana e Costa, C.A., Oliveira, M.D., 2012. A multicriteria decision analysis model for faculty evaluation. *Omega* 40, 424–436.
- Bana e Costa, C.A., Vansnick, J.-C., 1994. MACBETH — an interactive path towards the construction of cardinal value functions. *Int. Trans. Oper. Res.* 1, 489.
- Bana e Costa, C.A., Vansnick, J.-C., 1997. Applications of the MACBETH approach in the framework of an additive aggregation model. *J. Multi-Criteria Decis. Anal.* 6, 107–114.
- Bazerman, M., 1998. *Judgment in Managerial Decision Making*. John Wiley, New York.
- Beauchamp, T.L., 2003. Methods and principles in biomedical ethics. *J. Med. Ethics* 29, 269–274.
- Becker, M.H., Maiman, L.A., 1980. Strategies for enhancing patient compliance. *J. Community Health* 6, 113–135.
- Belton, V., Stewart, T., 2002. *Multiple Criteria Decision Analysis: an Integrated Approach*. Kluwer Academic Publishers, Dordrecht.
- Berkeley, D., Humphreys, P., 1982. Structuring decision problems and the 'bias heuristic'. *Acta Psychol.* 50, 201–252.
- Biomarkers Definitions Study Group, 2001. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* 69, 89–95.
- Bridges, et al., 2011 Jun. Conjoint analysis applications in health—a checklist: a report of the ISPOR good research practices for conjoint analysis task force. *Value Health* 14 (4), 403–413. <http://dx.doi.org/10.1016/j.jval.2010.11.013>.
- Brouwer, W., Van Exel, J., Baker, R., Donaldson, C., 2008. The new myth. *Pharmacoeconomics* 26, 1–4.
- Buede, D.M., 1986. Structuring value attributes. *Interfaces* 16, 52–62.
- Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D., Geys, H., 2000. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* 1, 49–67.
- Centre for Reviews and Dissemination, 2009. *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care*. University of York.
- Chen, T.-H., Li, L., Kochen, M.M., 2005. A systematic review: how to choose appropriate health-related quality of life (HRQOL) measures in routine general practice? *J. Zhejiang Univ. Sci. B* 6, 936–940.
- Cherepanov, D., Palta, M., Fryback, D.G., 2010. Underlying dimensions of the five health-related quality-of-life measures used in utility assessment: evidence from the National Health Measurement Study. *Med. Care* 48, 718–725.
- Cherny, N.I., Sullivan, R., Dafni, U., Kerst, J.M., Sobrero, A., Zielinski, C., et al., 2015. A standardised, generic, validated approach to stratify the magnitude of clinical benefit that can be anticipated from anti-cancer therapies: the European Society for Medical Oncology Magnitude of Clinical Benefit Scale (ESMO-MCBS). *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol./ESMO* 26, 1547–1573.
- Claxton, K., 2013. Should multi-criteria decision analysis (MCDA) replace cost effectiveness analysis (CEA) for evaluation of health care coverage decisions?. In: ISPOR 16th Annual European Congress Dublin.
- Claxton, K., Posnett, J., 1996. An economic approach to clinical trial design and research priority-setting. *Health Econ.* 5, 513–524.
- Clemen, R.T., 1996. *Making Hard Decisions; an Introduction to Decision Analysis*. Belmont, Duxbury.
- Cramer, J., Mattson, R., 1991. In: Cramer, J., Spilker, B. (Eds.), *Monitoring Compliance with Antiepileptic Drug Therapy*. Raven Press, New York, pp. 123–137.
- de Bekker-Grob, E.W., Ryan, M., Gerard, K., 2012. Discrete choice experiments in health economics: a review of the literature. *Health Econ.* 21, 145–172.
- Department for Communities and Local Government, 2009. *Multi-criteria Analysis: a Manual*.
- DH, 2010. In: Department of Health (Ed.), *A New Value-based Approach to the Pricing of Branded Medicines: a Consultation*. London.
- DH, 2011. In: Department of Health (Ed.), *A New Value-based Approach to the Pricing of Branded Medicines: Government Response to Consultation*. London.
- Diaby, V., Goeree, R., 2014. How to use multi-criteria decision analysis methods for reimbursement decision-making in healthcare: a step-by-step guide. *Expert Rev. Pharmacoeconomics Outcomes Res.* 14, 81–99.
- Dolan, J.G., 2010. Multi-criteria clinical decision support: a primer on the use of multiple criteria decision making methods to promote evidence-based, patient-centered healthcare. *The Patient* 3, 229–248.
- Drummond, M., McGuire, A., 2001. *Economic Evaluation in Health Care: Merging Theory with Practice*. Oxford University Press, USA.
- Drummond, M., Sculpher, M., Claxton, K., Stoddart, G., Torrance, G., 2015. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press.
- European Commission, 30.3.2006. Commission Regulation (EC) No. 507/2006 on the conditional marketing authorization for medicinal products for human use. *Eur. Comm. Off. J. Eur. Union. L* 92/6–L92/9. Available via: http://ec.europa.eu/health/sites/health/files/eudralex/vol-1/reg_2006_507/reg_2006_507_en.pdf.
- Eddy, D.M., 1991. Clinical decision making: from theory to practice. The individual vs society. Resolving the conflict. *JAMA* 265, 2399.
- Eldridge, S., Ashby, D., Bennett, C., Wakelin, M., Feder, G., 2008. Internal and external validity of cluster randomised trials: systematic review of recent trials. *BMJ* 336, 876–880.
- EMA, 2007. Report of the CHMP Working Group on Benefit–risk Assessment Models and Methods. European Medicines Agency.
- EMA, 2008. Reflection Paper on Benefit–risk Assessment Methods in the Context of the Evaluation of Marketing Authorisation Applications of Medicinal Products for Human Use. European Medicines Agency.
- EMA, 2011. Benefit-risk Methodology Project. Work Package 3 Report: Field Tests. European Medicines Agency.
- EMA, 2012. Benefit-risk Methodology Project. Work Package 4 Report: Benefit-risk Tools and Processes. European Medicines Agency.
- Emanuel, E.J., 2000. Justice and managed care. Four principles for the just allocation of health care resources. *Hastings Cent. Rep.* 30, 8–16.
- EUneHTA. The HTA Core Model®. Version 2.1 2015. European Network for Health Technology Assessment (<https://mekta.thl.fi/htacore/>).
- European Parliament, 22.1.2000. Regulation (EC) No 141/2000 of the European Parliament and of the Council of December 16, 1999 on orphan medicinal products. *Eur. Parliam. Off. J. Eur. Communities* 18, 1–9. Available via: http://ec.europa.eu/health/sites/health/files/eudralex/vol-1/reg_2000_141_cons-2009-07/reg_2000_141_cons-2009-07_en.pdf.
- Felli, et al., 2009. A multiattribute model for evaluating the benefit-risk profiles of treatment alternatives. *Med. Decis. Making* 29 (1), 104–115.
- FDA, 2005. *Guidance for Industry: Premarketing Risk Assessment*. US Food and Drug Administration, Center for Drug Evaluation and Research.
- Fleck, L., 2001. Healthcare justice and rational democratic deliberation. *Am. J. Bioeth.* AJOB 1, 20–21.
- Fleming, T., 2003. Design Issues in Colorectal Cancer Trials: Surrogate Endpoints and Non-inferiority Trials. Endpoints Workshop: Colorectal Cancer: Food and Drug Administration.
- Franco, L., Montibeller, G., 15 JUN 2010. Problem structuring for multicriteria decision analysis interventions analysis interventions. *Wiley Encycl. Oper. Res. Manag. Sci.* <http://dx.doi.org/10.1002/9780470400531.eorms0683>.
- Gabrielli, W., von Winterfeldt, D., 1978. Are Importance Weights Sensitive to the Range of Alternatives in Multiattribute Utility Measurement? Research report 78–6 University of Southern California, Social Science Research Institute.

- George, J., Phun, Y.-T., Bailey, M.J., Kong, D.C.M., Stewart, K., 2004. Development and validation of the medication regimen complexity index. *Ann. Pharmacother.* 38, 1369–1376.
- Glasgow, R., Vogt, T., Boles, S., 1999. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am. J. Public Health Aff.* 89, 1322–1327.
- Godwin, M., Ruhland, L., Casson, I., MacDonald, S., Delva, D., Birtwhistle, R., et al., 2003. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med. Res. Methodol.* 3, 28–28.
- Goetghebuer, M.M., Wagner, M., Khoury, H., Levitt, R.J., Erickson, L.J., Rindress, D., 2008. Evidence and value: impact on DEcisionMaking—the EVIDEM framework and potential applications. *BMC Health Serv. Res.* 8, 270.
- Goetghebuer, et al., 2012 Mar-Apr. Bridging health technology assessment (HTA) and efficient health care decision making with multicriteria decision analysis (MCDA): applying the EVIDEM framework to medicines appraisal. *Med. Decis. Making* 32 (2), 376–388.
- Goodwin, P., Wright, G., 1997. *Decision Analysis for Management Judgement*. John Wiley & Sons, Chichester.
- Gordon, M., Kass, M., 1991. Validity of standard compliance measures in glaucoma compared with an electronic eyedrop monitor. In: Cramer, J., Spilker, B. (Eds.), *Patient Compliance in Medical Practice and Clinical Trials*. Raven Press, New York, pp. 163–173.
- Guindo, L.A., Wagner, M., Baltussen, R., Rindress, D., van Til, J., Kind, P., et al., 2012. From efficacy to equity: literature review of decision criteria for resource allocation and healthcare decisionmaking. *Cost Eff. Resour. Alloc. C/E* 10, 9–9.
- Guitouni, A., Martel, J.-M., 1998. Tentative guidelines to help choosing an appropriate MCDA method. *Eur. J. Oper. Res.* 109, 501–521.
- Guyatt, G., 1992. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA* 268.
- Guyatt, G.H., Feeny, D.H., Patrick, D.L., 1993. Measuring health-related quality of life. *Ann. Intern. Med.* 118, 622–629.
- Guyatt, G.H., Oxman, A.D., Vist, G.E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., et al., 2008. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ (Clin. Res. Ed.)* 336, 924–926.
- Hammond, J., Keeney, R., Raiffa, H., 1999. *Smart Choices: A Practical Guide to Making Better Decisions*. Harvard University Press, Cambridge.
- Hansson, L.F., Norheim, O.F., Ruyter, K.W., 1994. Equality, explicitness, severity, and rigidity: the Oregon plan evaluated from a Scandinavian perspective. *J. Med. Philos.* 19, 343–366.
- Higgins, J.P.T., Altman, D.G., Gøtzsche, P.C., Jüni, P., Moher, D., Oxman, A.D., et al., 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ (Clin. Res. Ed.)* 343 d5928–d5928.
- HM, Treasury, 2003. *The Green Book: Appraisal and Evaluation in Central Government*: Treasury Guidance. TSO, London.
- Hodgson, T.A., Meiners, M.R., 1982. Cost-of-illness methodology: a guide to current practices and procedures. *Milbank Mem. Fund Q. Health Soc.* 60, 429–462.
- Howard, R.A., 1966. *Decision Analysis: Applied Decision Theory*. Wiley-Interscience.
- Hughes-Wilson, et al., 2012 Sep 26. Paying for the orphan drug system: break or bend? Is it time for a new evaluation system for payers in Europe to take account of new rare disease treatments? *Orphanet J. Rare Dis.* 7, 74. <http://dx.doi.org/10.1186/1750-1172-7-74>.
- Johnson, F.R., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D.A., et al., 2013. Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health J. Int. Soc. Pharmacoeconomics Outcomes Res.* 16, 3–13.
- Kanavos, P., Angelis, A., 2014. Value Assessment Methods and Application of Multiple Criteria Decision Analysis for HTA. 1st Advance-HTA Capacity Building Workshop in Latin America, Mexico City, Mexico.
- Kanavos, P., Angelis, A., 2015. Using Multiple Criteria Decision Analysis: Conceptual Framework and Model. 2nd Advance-HTA Capacity Building Workshop in Latin America, Santiago, Chile.
- Kass, M.A., Meltzer, D.W., Gordon, M., Cooper, D., Goldberg, J., 1986. Compliance with topical pilocarpine treatment. *Am. J. Ophthalmol.* 101, 515–523.
- Keeney, R., 1992. *Value Focused Thinking: A Path to Creative Decision Making*. Harvard University Press, Cambridge.
- Keeney, R., Raiffa, H., 1993. *Decisions with Multiple Objectives: Preferences and Value Trade-offs*. Cambridge University Press, Cambridge.
- Keeney, R.L., 1982. Decision analysis: an overview. *Oper. Res.* 30, 803–838.
- Keeney, R.L., 2002. Common mistakes in making value trade-offs. *Oper. Res.* 50, 935.
- Keeney, R.L., Gregory, R.S., 2005. Selecting attributes to measure the achievement of objectives. *Oper. Res.* 53, 1–11.
- Keeney, R.L., Raiffa, H., 1976. *Decisions with Multiple Objectives: Performances and Value Trade-offs*. Wiley, New York.
- Kesselheim, A.S., Wang, B., Avorn, J., 2013. Defining “innovativeness” in drug development: a systematic review. *Clin. Pharmacol. Ther.* 94, 336–348.
- Kruse, W., Weber, E., 1990. Dynamics of drug regimen compliance—its assessment by microprocessor-based monitoring. *Eur. J. Clin. Pharmacol.* 38, 561–565.
- Linley, W.G., Hughes, D.A., 2013. Societal views on NICE, cancer drugs fund and value-based pricing criteria for prioritising medicines: a cross-sectional survey of 4118 adults in Great Britain. *Health Econ.* 22, 948–964.
- Lipsey, R., Carlow, K., 2000. Technology policy: basic concepts. In: Edquist, C., McKelvey, M. (Eds.), *Systems of Innovation: International Competitiveness*. Edward Elgar, Cheltenham.
- Malone, et al., August 2016. International society for pharmacoeconomics and outcomes research comments on the American society of clinical oncology value framework. *J. Clin. Oncol.* 34 (24), 2936–2937.
- Marsh, K., Ijzerman, M., Thokala, P., Baltussen, R., Boysen, M., Kaló, Z., et al., 2016. Multiple criteria decision analysis for health care decision making—emerging good practices: report 2 of the ISPOR MCDA emerging good practices task force. *Value Health J. Int. Soc. Pharmacoeconomics Outcomes Res.* 19, 125–137.
- Marsh, K., Lanitis, T., Neasham, D., Orfanos, P., Caro, J., 2014. Assessing the value of healthcare interventions using multi-criteria decision analysis: a review of the literature. *Pharmacoeconomics* 32, 345–365.
- Mauskopf, J.A., Sullivan, S.D., Annemans, L., Caro, J., Mullins, C.D., Nuijten, M., et al., 2007. Principles of good practice for budget impact analysis: report of the ISPOR Task Force on good research practices—budget impact analysis. *Value Health J. Int. Soc. Pharmacoeconomics Outcomes Res.* 10, 336–347.
- McCabe, C., Claxton, K., Tsuchiya, A., 2005. Orphan drugs and the NHS: should we value rarity? *BMJ (Clin. Res. Ed.)* 331, 1016–1019.
- McDaniels, T.L., Gregory, R.S., Fields, D., 1999. Democratizing risk management: successful public involvement in local water management decisions. *Risk Anal. Int. J.* 19, 497–510.
- Mestre-Ferrandiz, J., Mordoh, A., Sussex, J., 2012. *The Many Faces of Innovation: a Report for the ABPI by the Office of Health Economics*. Office of Health Economics.
- Miners, A., Cairns, J., Wailoo, A., 2013a. Department of Health Proposals for Including Burden of Illness into Value Based Pricing: a Description and Critique. NICE Decision Support Unit.
- Miners, A., Cairns, J., Wailoo, A., 2013b. Department of Health Proposals for Including Wider Societal Benefits into Value Based Pricing: a Description and Critique. NICE Decision Support Unit.
- Mirelman, A., Mentzakis, E., Kinter, E., Paolucci, F., Fordham, R., Ozawa, S., et al., 2012. Decision-making criteria among national policymakers in five countries: a discrete choice experiment eliciting relative preferences for equity and efficiency. *Value Health J. Int. Soc. Pharmacoeconomics Outcomes Res.* 15, 534–539.
- Montibeller, G., Winterfeldt, D., 2015. Cognitive and motivational biases in decision and risk analysis. *Risk Anal. Int. J.* 35, 1230–1251.
- Morgan, S., Lopert, R., Greyson, D., 2008. Toward a definition of pharmaceutical innovation. *Open Med. A Peer-Reviewed, Indep. Open-Access J.* 2, e4–e7.
- Murphy, J., Coster, G., 1997. Issues in patient compliance. *Drugs* 54, 797–800.
- Murray, C., Lopez, A., 1996. *The Global Burden of Disease: Summary*. Harvard School of Public Health on behalf of the World Health Organisation and the World Bank.
- Mussen, F., Salek, S., Walker, S., 2007a. A quantitative approach to benefit-risk assessment of medicines - part 1: the development of a new model using multi-criteria decision analysis. *Pharmacoepidemiol. Drug Saf.* 16 (Suppl. 1), S2–S15. <http://dx.doi.org/10.1002/pds.1435>.
- Mussen, F., Salek, S., Walker, S., 2009. Benefit-risk Appraisal of New Medicines: a Systematic Approach to Decision Making. John Wiley, Surrey, UK.
- Mussen, F., Salek, S., Walker, S., Phillips, L., 2007b. A quantitative approach to benefit-risk assessment of medicines - part 2: the practical application of a new model. *Pharmacoepidemiol. Drug Saf.* 16 (Suppl. 1), S16–S41.
- National Cancer Institute, 2009. Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0. National Institutes of Health.
- Nagy, V.T., Wolfe, G.R., 1984. Cognitive predictors of compliance in chronic disease patients. *Med. Care* 22, 912–921.
- Naranjo, C.A., Busto, U., Sellers, E.M., Sandor, P., Ruiz, I., Roberts, E.A., et al., 1981. A method for estimating the probability of adverse drug reactions. *Clin. Pharmacol. Ther.* 30, 239–245.
- NCCN, 2015. *NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) with NCCN Evidence Blocks™*. National Comprehensive Cancer Network (<https://www.nccn.org/evidenceblocks/>).
- Neumann, P.J., Cohen, J.T., 2015. Measuring the value of prescription drugs. *N. Engl. J. Med.* 373, 2595–2597.
- NICE, 2011. *Value Based Pricing: Response to the Department of Health Consultation*. National Institute for Health and Clinical Excellence.
- NICE, 2014a. *Consultation Paper: Value Based Assessment of Health Technologies*. National Institute for Health and Care Excellence, London.
- NICE, 2014b. *Responses to TA Methods Addendum Public Consultation*. National Institute for Health and Care Excellence.
- O'Brien, M.K., Petrie, K., Raeburn, J., 1992. Adherence to medication regimens: updating a complex medical issue. *Med. Care Rev.* 49, 435–454.
- OHE, 2005. *The Many Faces of Innovation*. Office of Health Economics.
- OHE, 2011. *Response to the Department of Health's Consultation*. Office of Health Economics.
- Pazdur, 2008. Endpoints for assessing drug activity in clinical trials. *Oncologist* 13 (Suppl 2), 19–21. <http://dx.doi.org/10.1634/theoncologist.13-S2-19>.
- Peacock, S.J., Richardson, J.R.J., Carter, R., Edwards, D., 2007. Priority setting in health care using multi-attribute utility theory and programme budgeting and marginal analysis (PBMA). *Soc. Sci. Med.* 64, 897–910.
- Pearson, S., 2015. Framework for Payer Assessment of the Value of New Technologies: a US Approach. Seminar Briefing 16. Office of Health Economics.
- Phillips, L., Fasolo, B., Zafropoulos, N., Beyer, A., 2011. Is quantitative benefit-risk modelling of drugs desirable or possible? *Drug Discov. Today. Technol.* 8, e1–e42.
- Phillips, L.D., Bana e Costa, C.A., 2007. Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing. *Ann. Oper. Res.* 154, 51–68.

- Porter, M.E., 2010. What is value in health care? *N. Engl. J. Med.* 363, 2477–2481.
- Postmus, D., Tervonen, T., van Valkenhoef, G., Hillege, H.L., Buskens, E., 2014. A multi-criteria decision analysis perspective on the health economic evaluation of medical interventions. *Eur. J. Health Econ.* 15, 709–716.
- Raiffa, H., 1968. *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. Addison-Wesley, Reading, MA.
- Rawls, J., 1999. *A Theory of Justice*. Belknap Press of Harvard University.
- Rosenberg, N., 1982. *Inside the Black Box: Technology and Economics*. Cambridge University Press, Cambridge.
- Rosenberg, N., 2001. Keynote address: challenges for the social sciences in the new millennium. In: *OECD Social Sciences and Innovation*. OECD, Paris.
- Rothwell, P.M., 2006. Factors that can affect the external validity of randomised controlled trials. *PLoS Clin. Trials* 1, e9.
- Rudd, P., 1995. Clinicians and patients with hypertension: unsettled issues about compliance. *Am. Heart J.* 130, 572–579.
- Sackett, D.L., Rosenberg, W.M., Gray, J.A., Haynes, R.B., Richardson, W.S., 1996. Evidence based medicine: what it is and what it isn't. *BMJ (Clin. Res. Ed.)* 312, 71–72.
- Savage, L., 1954. *The Foundations of Statistics*. Wiley, New York.
- Schnipper, L.E., Davidson, N.E., Wollins, D.S., Tyne, C., Blayney, D.W., Blum, D., et al., 2015. American Society of Clinical Oncology statement: a conceptual framework to assess the value of cancer treatment options. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* 33, 2563–2577.
- Shikhar, R., Rentz, A.M., 2004. Satisfaction with medication: an overview of conceptual, methodologic, and regulatory issues. *Value Health J. Int. Soc. Pharmacoeconomics Outcomes Res.* 7, 204–215.
- Smith, N.A., Ley, P., Seale, J.P., Shaw, J., 1987. Health beliefs, satisfaction and compliance. *Patient Educ. Couns.* 10, 279–286.
- Spahn, J., 2003. Clinical trial efficacy: what does it really tell you? *J. Allergy Clin. Immunol.* 112, S102–S106.
- Stewart, T.J., 1995. Simplified approaches for multicriteria decision making under uncertainty. *J. Multi-Criteria Decis. Anal.* 4, 246–258.
- Sullivan, E. *Clinical Trial Endpoints: Food and Drug Administration presentation. Food and Drug Administration*. Accessed September 2013 (<https://www.fda.gov/downloads/Training/ClinicalInvestigatorTrainingCourse/UCM283378.pdf>).
- Tanios, N., Wagner, M., Tony, M., Baltussen, R., van Til, J., Rindress, D., et al., 2013. Which criteria are considered in healthcare decisions? Insights from an international survey of policy and clinical decision makers. *Int. J. Technol. Assess. Health Care* 29, 456–465.
- Tervonen, T., Naci, H., van Valkenhoef, G., Ades, A.E., Angelis, A., Hillege, H.L., et al., 2015. Applying multiple criteria decision analysis to comparative benefit-risk assessment: choosing among statins in primary prevention. *Med. Decis. Mak. Int. J. Soc. Med. Decis. Mak.* 35, 859–871.
- Tervonen, T., van Valkenhoef, G., Buskens, E., Hillege, H.L., Postmus, D., 2011. A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. *Stat. Med.* 30, 1419–1428.
- Thokala, P., Duenas, A., 2012. Multiple criteria decision analysis for health technology assessment. *Value Health J. Int. Soc. Pharmacoeconomics Outcomes Res.* 15, 1172–1181.
- Tromp, N., Baltussen, R., 2012. Mapping of multiple criteria for priority setting of health interventions: an aid for decision makers. *BMC Health Serv. Res.* 12, 454–454.
- von Winterfeldt, D., Edwards, W., 1986. *Decision Analysis and Behavioral Research*. Cambridge University Press, Cambridge.
- von Winterfeldt, D., Fasolo, B., 2009. Structuring decision problems: a case study and reflections for practitioners. *Eur. J. Oper. Res.* 199, 857–866.
- Walker, S., McAuslane, N., Liberti, L., Salek, S., 2009. Measuring benefit and balancing risk: strategies for the benefit-risk assessment of new medicines in a risk-averse environment. *Clin. Pharmacol. Ther.* 85, 241–246.
- Weinstein, M., Zeckhauser, R., 1973. Critical ratios and efficient allocation. *J. Public Econ.* 2, 147–157.
- Working Group on Mechanism of Coordinated Access to Orphan Medicinal Products, 2013. *Process on Corporate Social Responsibility in the Field of Pharmaceuticals Platform on Access to Medicines in Europe. Final Report*. European Commission, Brussels.
- World Health Organisation, 1948. *WHO Constitution*. Geneva.
- Wouters, O.J., Naci, H., Samani, N.J., 2015. QALYs in cost-effectiveness analysis: an overview for cardiologists. *Heart (Br. Heart Soc.)* 101, 1868–1873.
- Zhang, W., Bansback, N., Anis, A., 2011. Measuring and valuing productivity loss due to poor health: a critical review. *Soc. Sci. Med.* 72, 185–192.