Social Science & Medicine 246 (2020) 112595

Contents lists available at ScienceDirect

Social Science & Medicine

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

Multiple Criteria Decision Analysis for HTA across four EU Member States: Piloting the Advance Value Framework

A. Angelis[a,*], M. Linch[b], G. Montibeller[c], T. Molina-Lopez[d], A. Zawada[e], K. Orzel[e], F. Arickx[f], J. Espin[g,h], P. Kanavos[a,*]

[a] LSE Health and Department of Health Policy, London School of Economics, London, UK
[b] University College London Cancer Institute and University College London Hospital, London, UK
[c] School of Business and Economics, Loughborough University, Loughborough, UK
[d] Andalusian Agency for Health Technology Assessment, Sevilla, Spain
[e] Agency for Health Technology Assessment and Tariff System, Warsaw, Poland
[f] National Institute for Health and Disability Insurance, Brussels, Belgium
[g] Andalusian School of Public Health, Instituto de Investigación Biosanitaria (ibs.GRANADA), Granada, Spain
[h] CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

ARTICLE INFO

Keywords:
Multiple Criteria Decision Analysis (MCDA)
Health Technology Assessment (HTA)
Advance Value Framework (AVF)
Decision conference
Value assessment
Decision making
Pharmaceuticals
Oncology

ABSTRACT

Multiple Criteria Decision Analysis (MCDA) has emerged as a methodology for Health Technology Assessment (HTA). However, limited empirical evidence is available on its use by decision-makers; where available, it only comes from single-setting exercises, while cross-country comparative studies are unavailable. This study applies the Advance Value Framework (AVF), an MCDA methodology for HTA based on multi-attribute value theory, through a series of case studies with decision-makers in four countries, to explore its feasibility and compare decision-makers' value preferences and results.

The AVF was applied in the evaluation of three drugs for metastatic, castrate resistant, prostate cancer (abiraterone, cabazitaxel and enzalutamide) in the post-chemotherapy indication. Decision conferences were organised in four European countries in collaboration with their HTA or health insurance organisations by involving relevant assessors and experts: Sweden (TLV), Andalusia/Spain (AETSA), Poland (AOTMiT) and Belgium (INAMI-RIZIV). Participants' value preferences, including performance scoring and criteria weighting, were elicited through a facilitated decision-analysis modelling approach using the MACBETH technique.

Between 6 and 11 criteria were included in each jurisdiction's value model, allocated across four criteria domains; Therapeutic Benefit criteria consistently ranked first in relative importance across all countries. Consistent drug rankings were observed in all settings, with enzalutamide generating the highest overall weighted preference value (WPV) score, followed by abiraterone and cabazitaxel. Dividing drugs' overall WPV scores by their costs produced the lowest "cost per unit of value" for enzalutamide, followed by abiraterone and cabazitaxel. These results come in contrast with the actual country HTA recommendations and pricing decisions.

Overall, although some differences in value preferences were observed between countries, drug rankings remained the same. The MCDA methodology employed could act as a decision support tool in HTA, due to the transparency in the construction of value preferences in a collaborative manner.

1. Introduction

In recent years, the introduction of new and costly health technologies, particularly in oncology, combined with moderate health gains, has sparked extensive debate on their value for patients and health care systems, how this value should be assessed and what should be the evaluation criteria informing coverage decisions (Cohen, 2017; Linley and Hughes, 2013). The debate has been fuelled by diverging coverage recommendations across settings for several medicines, often related to diseases associated with high morbidity and mortality (Clement et al., 2009; Faden et al., 2009; Nicod and Kanavos, 2012). Difference in opinion often arises in resource allocation decisions amongst different stakeholders, attributable, at least in part, to current evaluation methodologies not adequately capturing different notions of value

* Corresponding authors. LSE Health and Department of Health Policy, London School of Economics, Houghton Street, London, WC2A 2AE, UK.
E-mail addresses: a.n.angelis@lse.ac.uk (A. Angelis), p.g.kanavos@lse.ac.uk (P. Kanavos).

https://doi.org/10.1016/j.socscimed.2019.112595
Received 26 March 2019; Received in revised form 29 September 2019; Accepted 10 October 2019
Available online 15 October 2019
0277-9536/

(Drummond et al., 2013); this includes, for example, the Quality Adjusted Life Year (QALY), the use of which in economic evaluations can at times be regarded as blunt and insufficient, among other reasons because it may not adequately reflect important value aspects in a variety of disease areas (Devlin and Lorgelly, 2017; Efthymiadou et al., 2019; Wouters et al., 2015). Given the limited consideration of overall value in traditional economic evaluations, additional parameters have been included in value assessments; however, this is often done in a non-systematic or ad-hoc manner, which may impact the transparency of decision-making processes (Angelis et al., 2018) and lead to inconsistencies in drug coverage decisions.

A growing body of literature is increasingly debating the use of highly expensive new drugs, which are perceived to bring marginal added clinical benefit on the grounds of poor value-for-money and high budget impact (Nadler et al., 2006; Shih et al., 2013; Sulmasy and Moy, 2014). High and rising drug prices and the need to understand the importance of different evaluation criteria have catalysed the generation of numerous "value frameworks" aiming to inform payers, clinicians and patients on the assessment of new medicines, required for making coverage and treatment selection decisions (Anderson et al., 2014; Bach, 2015; Cherny et al., 2015; Schnipper et al., 2015). Although this is an important step towards a more inclusive value-based assessment approach (Malone et al., 2016), aspects of these frameworks may be based on weak methodologies or ad hoc considerations, which could potentially result in misleading recommendations or decisions (Angelis and Kanavos, 2016).

In response to some of the concerns raised above, multiple criteria decision analysis (MCDA) has emerged as a likely alternative or supplementary approach to traditional economic evaluation techniques with the prospects of addressing some of their limitations in Health Technology Assessment (HTA) (Angelis et al., 2017; Devlin and Sussex, 2011; Goetghebeur et al., 2008; Kanavos and Angelis, 2013; Marsh et al., 2014; Radaelli et al., 2014; Thokala, 2011), but also for eliciting stakeholder preferences and facilitating treatment selection (Danner et al., 2011; Ijzerman et al., 2008; Tervonen et al., 2015). A number of MCDA empirical studies have explored the question of value in a number of therapeutic areas, often simulating hypothetical HTA settings (Angelis et al., 2017; Goetghebeur et al., 2010; Sussex et al., 2013; Wagner et al., 2017). However, very few studies have explored the same issue by eliciting the preferences of HTA agencies and sitting decision makers and only in single-case exercises (Angelis, 2018; Jaramillo et al., 2016; Tony et al., 2011). To the best of our knowledge, no study has ever elicited and compared the value preferences of sitting decision-makers across multiple settings using a full MCDA methodology, while considering identical sets of evidence.

By involving HTA agencies and health insurance organisations in four EU Member States, we applied the Advance Value Framework (AVF), a recently developed multi-criteria value framework applicable to HTA (Angelis and Kanavos, 2016, 2017), to assess the value of treatment options indicated for metastatic castrate resistant prostate cancer (mCRPC) following first line chemotherapy. This indication was selected because of its high disease burden and the availability of several new and expensive biologic drugs, making it a highly relevant appraisal topic for several HTA agencies.

The main research questions of the study relate, first, to testing the feasibility of this MCDA methodology for HTA decision-makers and, second, to observing any differences in their value perceptions as reflected through the consistency of drugs' value rankings, including value trade-offs.

2. Methods

2.1. Methodological Framework

An MCDA approach based on Multi-Attribute Value Theory (MAVT) was adopted (Keeney and Raiffa, 1993; von Winterfeldt and Edwards, 1986), involving the phases of problem structuring, model building, model assessment, model appraisal, and development of action plans (Angelis and Kanavos, 2016b). A series of facilitated workshops were organised taking the form of decision conferences (Phillips, 2007), adopting a facilitated decision analysis modelling approach (Franco and Montibeller, 2010b; Phillips and Phillips, 1993), in collaboration with decision-makers from four HTA agencies and health insurance bodies: the Dental and Pharmaceutical Benefits Agency (TLV, Sweden), the Andalusian Health Technology Assessment Agency (AETSA, Spain), the Agency for Health Technology Assessment and Tariff System (AOTMiT, Poland), and the National Health Insurance Agency (INAMI-RIZIV, Belgium). The agencies in these countries were selected in order to represent a set of organisations with different governance structure (arms' length HTA agency, e.g. AOTMiT, TLV and AETSA, vs integrated HTA function, e.g. INAMI-RIZIV) and responsibilities (regulatory, e.g. TLV, vs advisory, e.g. AOTMiT and AETSA). The research was undertaken in the context of Advance-HTA, an EU-funded project focusing on HTA methodological advancements (London School of Economics, 2019), and all four HTA organisations were contacted to participate under the auspices of the project.

The methodological process used in terms of the design, implementation and analysis, is aligned with the ISPOR good practice guidelines on the use of MCDA for health care decisions (Marsh et al., 2016).

2.2. Problem Structuring: clinical practice and scope of the exercise

Prostate cancer is the second most commonly diagnosed cancer in men globally, the most frequently diagnosed cancer among men in developed countries and the fifth leading cause of cancer death globally (Torre, 2015). Death rates have been decreasing in most developed countries and this has been attributed mainly to improved treatment and/or early detection (Center et al., 2012).

The decision context relates to the assessment of value of second line treatments for mCRPC based on the approved European Medicines Agency (EMA) indication (EMA, 2016a, b, c), the subsequently defined scope of Technology Appraisals (TAs) by a number of HTA agencies and the ESMO guidelines (Horwich et al., 2013; NICE, 2012a, b, 2014; TLV, 2014, 2015a).

The first treatment to demonstrate a survival benefit for mCRPC patients was docetaxel chemotherapy in combination with prednisolone when compared to mitoxantrone in combination with prednisolone (Berthold et al., 2008; Tannock et al., 2004). Subsequently, new therapeutic agents have been tested in the post-chemotherapy setting with considerable success. Abiraterone, a steroid synthesis inhibitor, in combination with prednisolone showed a 3.9-month improvement in survival compared to prednisolone alone in patients pre-treated with docetaxel (14.8 vs 10.9 months, HR 0.65, $p < 0.001$) (de Bono et al., 2011). Similarly, enzalutamide, an androgen receptor antagonist, showed a 4.8-month improvement in survival (18.4 vs 13.6 months, HR 0.63, $p < 0.001$) compared to placebo alone in the same patient group (Scher et al., 2012). Cross-resistance appears to exist between abiraterone and enzalutamide meaning that patients are unlikely to derive clinical benefit by switching from one agent to the other (Bianchini et al., 2014; Loriot et al., 2013). The third agent that is widely used following progression on docetaxel is cabazitaxel, a taxane chemotherapy. Cabazitaxel led to an overall survival (OS) benefit of 2.4 months (15.1 vs 12.7 months, HR 0.70, $p < 0.0001$) compared to mitoxantrone (de Bono et al., 2010). Given this therapeutic landscape for patients with mCRPC who have progressed on first line docetaxel chemotherapy, characterised by an availability of different treatments and the apparent cross-resistance between some of them, we adopt post-chemotherapy mCRPC as the decision context for the application of the AVF methodology.

2.3. Model Building: Advance Value Tree adaptation, treatments compared and reference levels

The model building phase comprised a number of tasks, notably the Advance Value Tree adaptation for mCRPC, the consideration of alternative drug treatments together with the respective evidence, and the definition of criteria, attributes and their associated ranges, all of which are discussed below. Detailed discussion on the rationale of each non disease-specific criterion and their value scales can be found elsewhere (Angelis and Kanavos, 2017; Angelis et al., 2017).

2.3.1. Adaptation of the Advance Value Tree for metastatic prostate cancer

At the core of AVF lies the Advance Value Tree, a hierarchical structure of evaluation criteria taking the form of a generic value tree reflecting value concerns of HTA experts and decision-makers for new medicines (Angelis and Kanavos, 2017). The Advance Value Tree consists of five criteria domains, aiming to capture the essential value attributes of new medicines in the HTA context under a prescriptive decision-aid approach. These are divided into (a) Burden of Disease (BoD); (b) Therapeutic Benefit (THE); (c) Safety Profile (SAF); (d) Innovation Level (INN); and (e) Socioeconomic Impact (SOC), summarised by the following value function:

$$Value = \boldsymbol{f}\,(\boldsymbol{BoD},\ \boldsymbol{THE},\ \boldsymbol{SAF},\ \boldsymbol{INN},\ \boldsymbol{SOC}) \tag{1}$$

The Advance Value Tree was adapted into a disease-specific mCRPC value model using a bottom-up approach by comparing the characteristics of the drugs under consideration (Franco and Montibeller, 2010a). In consultation with a specialist medical oncologist (co-author of the paper), the generic evaluation criteria were converted into disease-specific criteria, while adhering to required criteria properties such as non-redundancy and preferential-independence (Keeney, 1992), to ensure methodological robustness and an adequate value model rooted in decision theory. Based on the above, a preliminary mCRPC-specific value tree was produced with four criteria domains and a total of 18 criteria as shown in Fig. 1, each operationalised by an attribute, i.e. performance indicator. The BoD domain was not considered in the adaptation process on the grounds of conciseness, as all drugs were indicated for the same indication which would have identical BoD.

Criteria definitions (together with their consideration in each jurisdiction and their rankings) are provided in Table 1. The preliminary version of the mCRPC value tree was subsequently validated by decision conference participants, in line with a "socio-technical" approach, a constructive decision-aid process allowing groups of participants to interact with and learn from each other (Bana e Costa and Beinat, 2005).

2.3.2. Alternative treatments compared and evidence considered

The alternative drug options assessed in the exercise were cabazitaxel in combination with prednisolone, abiraterone in combination with prednisolone and enzalutamide monotherapy. The key evidence sources used to assess their performance included (a) the peer review publications concerning their pivotal clinical trials that were considered for their licensing by the EMA (de Bono et al., 2010, 2011; Fizazi et al., 2012; Scher et al., 2012); (b) the Product Information sections of EMA's European Public Assessment Reports (EPAR) (Annex I and III) (EMA, 2016a, b, c); (c) the Anatomical Therapeutic Chemical (ATC) classification system indexes available through the portal of the WHO Collaborating Centre for Drug Statistics Methodology (World Health Organisation Collaborating Centre, 2016); and (d) the US National Library of Medicine clinical trials database (NIH, 2016). Additional sources of evidence included national sources (BNF, 2015; Connock et al., 2011; NICE, 2012a, b, 2014; Riemsa et al., 2013) and other peer review literature (Burström et al., 2001; Collins et al., 2007; Kearns et al., 2013; Sullivan et al., 2007), which was relevant to the study indication. Sources of evidence used relating to the performance of drugs across evaluation criteria are shown in Appendix Table A1, alongside additional information on the evidence considered.

2.3.3. Options performance and references levels

By considering the performance of the alternative drug options across the value scales, "lower" (x_l) and "higher" (x_h) reference levels were defined to serve as benchmarks for the value scores of 0 and 100 respectively, acting as value anchors for constructing value functions and eliciting their relative weights (Bana e Costa and Vansnick, 1999; Keeney, 1992). The "lower" reference levels denoted a less preferred state reflecting a "satisfactory" performance level, whereas the "higher" reference levels denoted a more preferred state reflecting an "ideal" performance level.

The reference levels for the clinical attributes informing the Therapeutic and Safety criteria domains, were defined in consultation with the clinical oncologist (co-author of the paper). In principle, the rationale involved adopting the Best Supportive Care (BSC) performance as a "satisfactory" reference level, with a hypothetical 20% improvement of the best available performance acting as the "ideal" reference level (e.g. 'overall survival'), or alternatively the best possible limit of the performance scale acting as an "ideal" level in cases where this was naturally restricted (e.g. 'treatment discontinuation'). The 20% hypothetical performance improvement was selected because it was perceived to be a realistically plausible scenario for future treatment options. By considering the performance of best available option(s) among the treatments evaluated and accounting for plausible performance improvement in the near future, the value scale essentially reflected characteristics of a "global" scale to account for the performance of future options not captured in the exercise, i.e. *what is best plausible* (Belton and Stewart, 2002). Where a BSC performance was not meaningful to act as a "lower" reference level, then the lowest (i.e. worst) possible limit of the performance scale was adopted (e.g. 'Phase 3'), or, alternatively, 20% lower than the lowest performing option was used (e.g. 'medical costs impact'). An exception to the above was the 'health related quality of life' (HRQoL) attribute for which the stable disease state's utility score was adopted as the "lower" level and the general population utility score was used as the "higher" level.

The emerging partial value function scores of the drugs for each criterion can take negative values or values higher than 100 where $v(x_{lower}) = 0$ and $v(x_{higher}) = 100$. "Lower" and "higher" reference levels for all attributes at the pre-decision conference stage and the basis of their selection are outlined in Appendix Table A2. A matrix listing the performance of drug options across the final criteria that were considered in the decision conferences, together with their reference levels, is shown in Table 2.

2.4. Model Assessment and Appraisal: decision conferences, MCDA technique and cost calculation

The model assessment and appraisal phases comprised the tasks of conducting the decision conferences, the application of the MCDA technique for the elicitation of value preferences and treatments' cost calculation. These are discussed below.

2.4.1. Decision conferences

Model assessment and model appraisal took place through a series of decision conferences (Phillips, 2007), taking the form of facilitated workshops with the participation of decision-makers. These included assessors and national experts, all of whom were affiliated with the four study HTA organisations, either as members of staff or visiting external experts (their difference being in full-time employment versus part-time or visiting capacity employment). For the purposes of this study, they were both regarded as "decision-makers", given their influence on methodological development within the agencies and on the appraisals' decision outcomes. Across the four countries, between four (for the case of TLV) and 13 (for the case of AOTMiT) participants were involved, typically comprising health care professionals (clinicians, pharmacists),

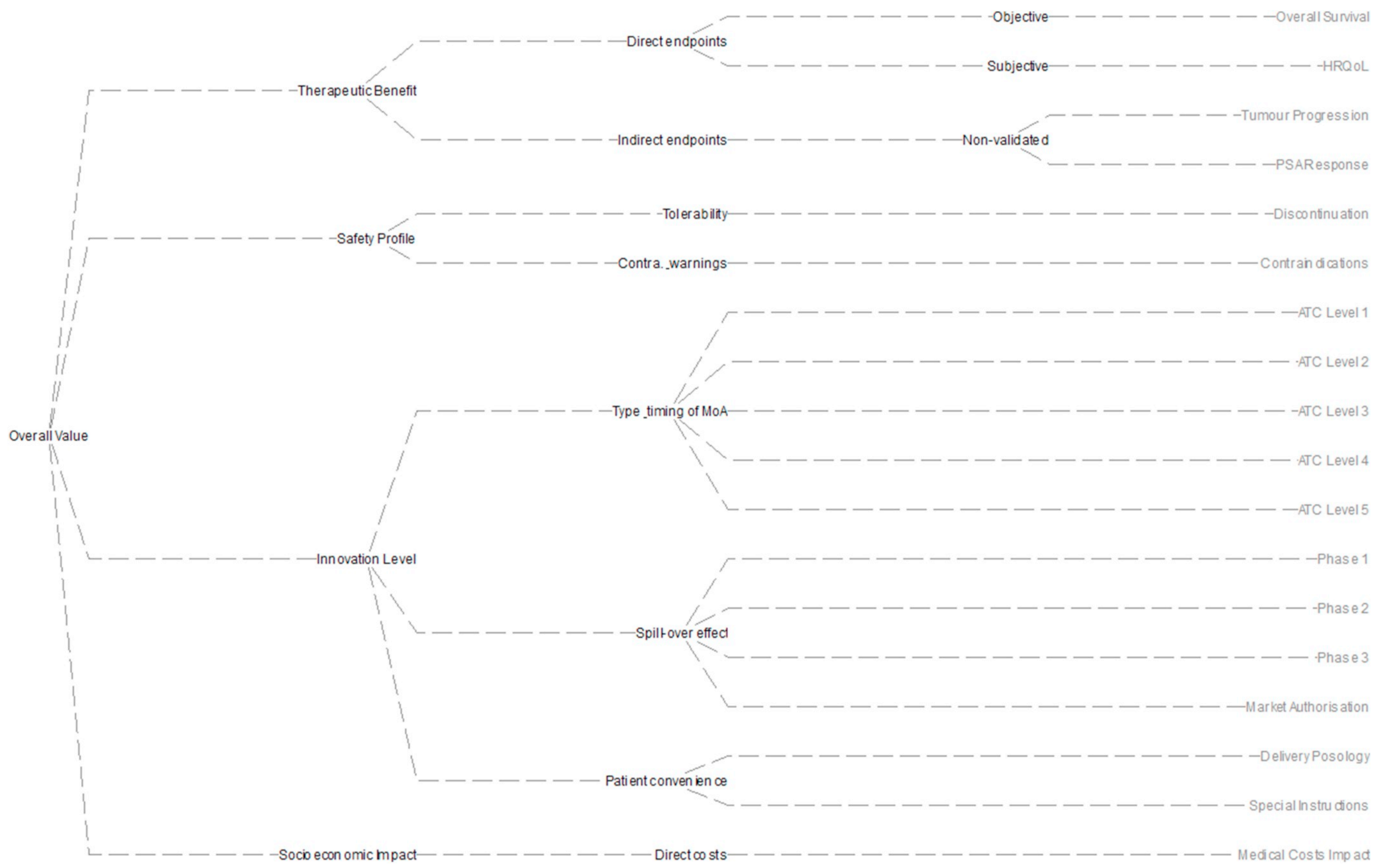

Fig. 1. Preliminary value tree for metastatic prostate cancer (pre-workshop).
Notes: Contra. = Contraindications; MoA = Mechanism of action; HRQoL = Health related quality of life; PSA = Prostate-specific antigen; ATC = Anatomical therapeutic chemical (classification). Image produced using Hiview3 3.2.0.4.

HTA methodology experts (health economists, statisticians, HTA agency directors) and institutional executives (members of HTA appraisal committees, representatives from insurance funds and national medicines agencies). Background material introducing the scope of the exercise in greater detail was sent to participants one week before each decision conference. Decision conferences were hosted at the head offices of the different HTA organisations between June 2015 and April 2016: Stockholm (TLV), Seville (AETSA), Warsaw (AOTMiT), and Brussels (INAMI-RIZIV).

The lead author acted as an impartial facilitator, assisted the groups' deliberative interactions and guided participants through the decision problem using the preliminary version of the mCRPC-specific value tree (Fig. 1) and the relevant data. This acted as the model's starting point that was subsequently validated, based on which value judgements and preferences were elicited in each decision conference while seeking group interaction and agreement (Franco and Montibeller, 2010b; Phillips, 1984; Phillips and Bana e Costa, 2007; Schein, 1999). The *Appendix* provides more information on the decision conferences.

2.4.2. MCDA technique

AVF adopts a value measurement MCDA methodology making use of a simple additive (i.e. linear, weighted average) value model for the aggregation of scores and weights (Angelis and Kanavos, 2017). This assumes preference independence between the different criteria, with overall value V(.) of an option *a* defined by the equation below (Keeney, 1992; von Winterfeldt and Edwards, 1986):

$$V(a) = \sum_{i=1}^{m} w_i \, v_i(a) \tag{2}$$

where m is the number of evaluation criteria, $w_i \, v_i(a)$ is the weighted partial value function of evaluation criterion i for treatment a, and V(a) is the overall value of the treatment a. V(.) is therefore an overall value function based on multi-attribute value theory (Keeney and Raiffa, 1993).

A value function associated with each attribute, converting the treatment performance on the attribute range to a value scale, was elicited from the participants during each decision conference using the Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH) questioning protocol and the M-MACBETH software (Bana e Costa and Vansnick, 1999). This protocol requires pairwise comparisons where qualitative judgements about the difference in value between different pairs of attribute levels (i.e. difference in value between x and y units of a criterion's performance indicator) are expressed using seven qualitative categories (i.e. no difference, very weak difference, weak difference, moderate difference, strong difference, very strong difference, or extreme difference) (Bana E Costa et al., 2012; Bana e Costa and Vansnick, 1994). MACBETH provides a constructive and user-friendly approach to generate a cardinal (interval) value scale based on the input of these qualitative pair-wise judgements, which are then converted into value scores via an optimization algorithm (Bana e Costa et al., 2016b); this approach has been widely used as a decision support tool (Bana e Costa et al., 2014; Bana e Costa et al., 2002; Bana e Costa and Oliveira, 2012; Bana e Costa and Vansnick, 1997).

Weights for a multi-attribute value function should be elicited using indirect questioning protocols that consider a performance range for each attribute (Keeney, 2002), as for example the value of a "swing" between two reference levels. The weights are scaling constants that convert partial value scores into overall value scores that must reflect value trade-offs and, therefore, should not be interpreted as measurements of 'direct importance'. An indirect (qualitative) swing weighting

Table 1
Criteria definitions, their consideration in each jurisdiction and their ranking.

Criteria Sub-Domain	Evaluation criteria	Definition	Country (competent HTA organisation)			
			Belgium (INAMI/ RIZIV)	Poland (AOTMiT)	Andalusia (AETSA)	Sweden (TLV)
Criteria Domain 1: Therapeutic Benefit						
Direct endpoints	Overall survival x Health related quality of life*	The median time from treatment randomisation to death adjusted for the mean health related quality of life using the EQ-5D utility score	✓(1st)	✓(1st)	✓(1st)	✓(1st)
Indirect endpoints	Radiographic tumour progression	The median survival time on which patients have not experienced disease progression (using RECIST criteria)	✓(5th)	✓(5th)	✓(4th)	✓(5th)
	PSA response	The proportion of patients having a $\geq$50% reduction in PSA			✓(8th)	
Criteria Domain 2: Safety Profile						
Tolerability	Treatment discontinuation	The proportion of patients discontinuing treatment due to adverse events	✓(3rd)	✓(3rd)	✓(2nd)	✓(2nd)
Contra-indications & warnings	Contra-indications	The existence of any type of contra-indication accompanying the treatment	✓(4th)		✓(3rd)	✓(4th)
Criteria Domain 3: Innovation Level						
Type and timing of MoA	ATC Level 1	The technology's relative market entrance in regards to its ATC Level 1 (Anatomical)				
	ATC Level 2	The technology's relative market entrance in regards to its ATC Level 2 (Therapeutic)				
	ATC Level 3	The technology's relative market entrance in regards to its ATC Level 3 (Pharmacological)				
	ATC Level 4	The technology's relative market entrance in regards to its ATC Level 4 (Chemical)	✓(6th)		✓(10th)	
	ATC Level 5	The technology's relative market entrance in regards to its ATC Level 5 (Molecular)				
Spill-over effect	Phase 1	The number of new indications for which the technology is investigated in Phase 1 clinical trials				
	Phase 2	The number of new indications for which the technology is investigated in Phase 2 clinical trials	✓(8th)			
	Phase 3	The number of new indications for which the technology is investigated in Phase 3 clinical trials	✓(9th)		✓(9th)	
	Marketing authorisation	The number of new indications that the technology has gained an approval for at the stage of marketing authorisation	✓(10th)		✓(7th)	
Patient convenience	Delivery posology	The combination of the delivery system (RoA and dosage form) with the posology (frequency of dosing and duration of administration) of the treatment	✓(7th)	✓(4th)	✓(6th)	✓(7th)
	Special instructions	The existence of any special instructions accompanying the administration of the treatment	✓(11th)	✓(6th)	✓(11th)	✓(6th)
Criteria Domain 4: Socioeconomic Impact						
Direct costs	Medical costs impact	The impact of the technology on direct medical costs excluding the purchasing costs of the technology	✓(2nd)	✓(2nd)	✓(5th)	✓(3rd)

Notes: *: Aggregation between OS and HRQoL criteria took place due to preference-dependence leading to a combined criterion; PSA = prostate-specific antigen; MoA = Mechanism of Action; ATC = Anatomical Therapeutic Chemical classification system; RoA = Route of Administration.
Source: The authors, based on input from decision conferences.

Table 2
Performance matrix and reference levels considered across the final criteria attributes.

Criterion	Attribute	Lower level	Abiraterone	Cabazitaxel	Enzalutamide	Higher level
Overall survival (OS)*	Months	13.6	15.8	15.1	18.4	22.1
Health Related Quality of Life (HRQoL), stable disease*	Utility (EQ-5D)	0.72	0.76	0.76***	0.76	0.82
Health Related Quality of Life (HRQoL), progressive disease*	Utility (EQ-5D)	0.64	0.64	0.64	0.64	0.82
OS X HRQoL**	Quality adjusted life months (QALMs)	9.2	11	10.5	12.8	18.1
Radiographic tumour progression, i.e. progression free survival (PFS)	Months	2.9	5.6	8.8	8.3	10.6
PSA response	% of patients	1.5	29.5	39.2	54	64.8
Treatment discontinuation	% of patients	10	19	18	8	0
Contra-indication(s)	Type of contra-indication	hyp + hep imp + low neut	hyp + hep imp	hyp + hep imp + low neut	hyp	None
ATC Level 4, i.e. chemical mechanism of action	Relative market entrance	5th	2nd	2nd	1st	1st
Phase 2	Number of new indications	0	1	13	4	16
Phase 3	Number of new indications	0	1	2	0	2
Marketing authorisation	Number of new indications	0	0	0	0	1
Delivery posology	Type of delivery system & posology combinations	Oral, daily - one off + IV, every 3 weeks - 1 h	Oral, daily - one off	Oral, daily - one off + IV, every 3 weeks - 1 h	Oral, daily - one off	Oral, daily - one off
Special instructions	Type(s) of special instructions	Concomitant and/or pre-med + no food	Concomitant and/or pre-med + no food	Concomitant and/or pre- med	None	None
Medical costs impact	GBP	10,000	5,750	7,992	567	0

Notes: * Used for the calculation of the quality adjusted life months (QALMs) attribute of the aggregated OS x HRQoL criterion; ** Calculated assuming an equal 50% split in time duration between the stable disease and progressive disease states in HRQoL; *** Used the same score of the other two options as data not available; hyp = hypersensitivity; hep imp = hepatic impairment; low neut = low neutrophil count.
Source: The authors based on the literature.

Table 3
Number of criteria per cluster, relative weights per criteria cluster and their ranking across the four HTA settings.

HTA Agency/ Criteria Clusters	Sweden (TLV)			Andalusia (AETSA)			Poland (AOTMiT)			Belgium (INAMI-RIZIV)		
	Criteria numbers	Criteria weights	Criteria ranking	Criteria numbers	Criteria weights	Criteria ranking	Criteria numbers	Criteria weights	Criteria ranking	Criteria numbers	Criteria weights	Criteria ranking
Therapeutic Benefit	2	44.5	1st	3	54.3	1st	2	40.0	1st	2	40.0	1st
Safety Profile	2	33.3	2nd	2	26.0	2nd	1	20.0	3rd	2	26.7	2nd
Innovation Level	2	7.4	4th	5	11.8	3rd	2	10.0	4th	6	13.3	4th
Socioeconomic Impact	1	14.8	3rd	1	7.9	4th	1	30.0	2nd	1	20.0	3rd
Total	7	100		11	100		6	100		11	100	

Source: The authors based on input from decision conferences.

technique was applied to elicit relative criteria weights by first ordering the swings of each attribute and then valuing their differences using the MACBETH qualitative categories (Bana E Costa et al., 2012).

The above MACBETH-based scoring and weighting techniques were operationalised using the software M-MACBETH (Bana e Costa and Vansnick, 1999). The software automatically checks the consistency between qualitative judgement inputs and derives value scores and weights using linear programming rules. It also automates the additive aggregation of value scores and weights in order to derive overall weighted preference value (WPV) scores and allows for sensitivity analysis on the criteria weights. Additionally, the software enables the use of visual graphics to build a model of values, acting as a facilitation tool to inform both the design and the evaluation phases of the methodological framework (Bana e Costa et al., 2016a; Bana e Costa and Vansnick, 1999; Bana e Costa et al., 1999). As part of a broader deliberative process, manual consistency checks were also performed by the facilitator, both to confirm the cardinality of the value scales and the understanding of the swing weights, followed by discussion and interpretation of the overall results. More information regarding the technical details of MACBETH is available in the *Appendix*.

2.4.3. Cost calculation

UK list prices at ex-factory level were used as found in the BNF (BNF, 2015). This acted as a neutral benchmark in order to allow the measurement of cost(s) in a common unit across all study settings, so that overall WPV scores could then be viewed against the same cost denominator to produce comparable cost-value ratios. Access to confidential prices through risk sharing agreements was not possible. Information on the recommended dosages and treatment durations were sourced from the pivotal trial peer review publications and the respective EMA EPARs (de Bono et al., 2011; de Bono et al., 2010; EMA, 2016a, b, c; Scher et al., 2012). Drug administration costs for cabazitaxel were kept consistent with the respective NICE TA (NICE, 2012b), whereas for abiraterone and enzaluatmide these costs were not applicable as they are orally administered.

3. Results

3.1. Final value trees, options performance, criteria weights and value functions

Across the four countries, decision conferences were characterised by increased interaction and extensive debate between participants, especially in cases where there was disagreement about certain values. Because the majority of participants had a shared understanding of the decision problem but also a sense of common purpose and commitment to a way forward, all of which are conditions for good practice in decision conferencing, the deliberative process of each decision conference instigated a fruitful discussion and exchange of views around different criteria values and their relative importance.

General consensus was reached among participants in terms of criteria consideration and model validation with no major value aspects deemed to be missing. All criteria included in each country's final mCRPC value tree, as emerged following open interaction with decision conference participants and their rankings, are shown in Table 1 (schematic illustrations of the individual value trees are shown in Appendix Fig. A1). The main reason for not including a criterion in the value tree was because participants would consider it to be non-fundamental to the evaluation, in all cases of which a zero weight was assigned. Most of the criteria that were assigned a zero weight belonged to the Innovation Level (INN) domain, which comprised the highest number of criteria.

The performance of the drug options across the different criteria attributes that were considered to be fundamental in the model (i.e. weight greater than zero) together with the “lower” and “higher” reference levels are shown in Table 2.

Between 6 (AOTMiT) and 11 (AETSA/INAMI) criteria were included in each country's final value tree, as shown in Table 3. In terms of the different criteria domains composition, the Therapeutic Benefit contained between two (TLV/AOTMiT/INAMI) and three (AETSA) criteria, the Safety Profile between one (AOTMiT) and two (TLV/AETSA/INAMI) criteria, the Innovation Level between two (TLV/AOTMiT) and six (INAMI) criteria, and the Socioeconomic Impact always one criterion.

During the elicitation of the ‘overall survival’ (OS) and/or ‘HRQoL’ criteria value functions, it became evident that these criteria might be preference dependent. When participants were asked to judge the difference in value between different increments in their performance (either in ‘OS’ or ‘HRQoL’), a request for clarification was raised by some participants relating to what level of performance this change was associated with on the other criterion. In order to address the plausible preference-dependence observed, we combined together the two criteria in an aggregated form. The two criteria were combined by multiplying the number of months in ‘OS’ and their EQ-5D utility scores in ‘HRQoL’ attributes respectively, assuming an equal (i.e. 50%) distribution of stable and progressive disease states, essentially deriving quality adjusted life months (QALMs). An example of a MACBETH value judgements matrix and its conversion into a value function for the case of the ‘radiographic tumour progression' (i.e. progression free survival (PFS)) criterion in months is shown in Appendix Fig. A2.

There was a common set of six criteria that were considered as fundamental in all countries: (a) ‘OS x HRQoL’; (b) ‘radiographic tumour progression’ (PFS); (c) ‘treatment discontinuation’; (d) ‘delivery posology’; (e) ‘special instructions’; and (f) ‘medical costs impact’. This common set of criteria comprised the complete set of TLV's value tree (n = 6), whereas AOTMIT's value tree also considered ‘contra-indications’ (n = 7). Further to these, AETSA's value tree considered ‘Prostate-specific antigen (PSA) response’, ‘ATCL4’, ‘Phase 3’ and

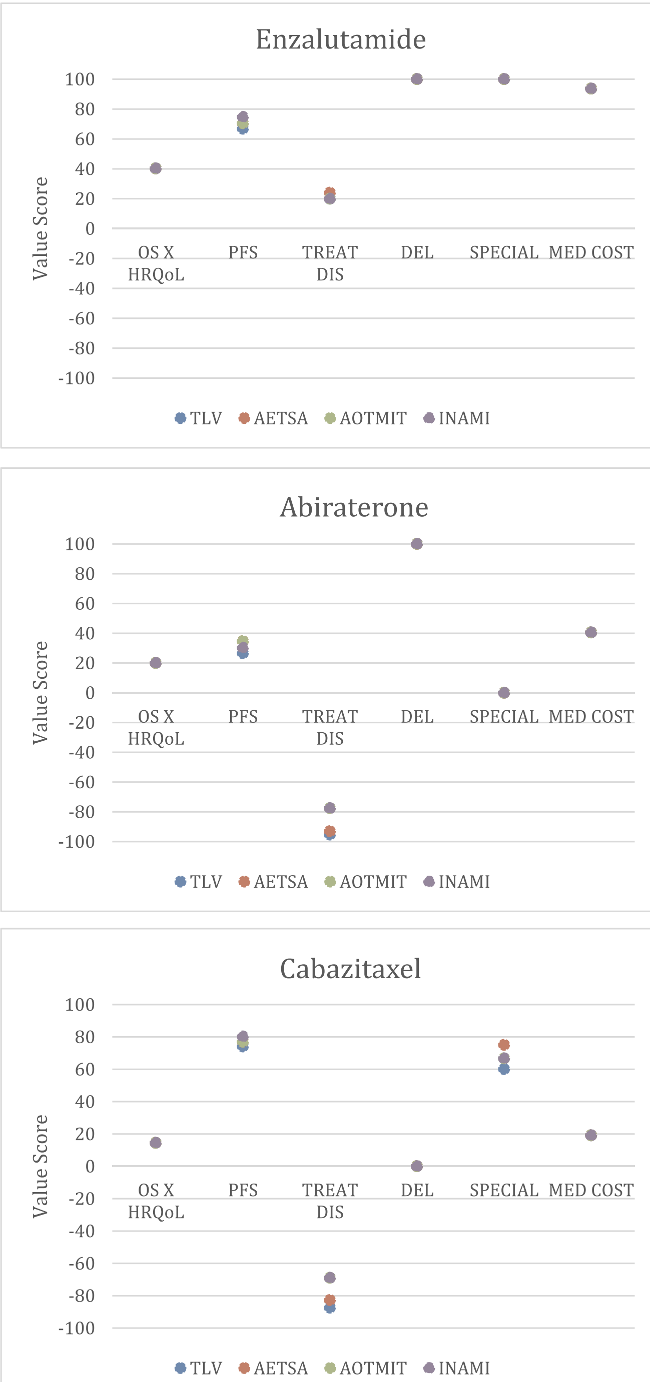

Fig. 2. Criteria valuation drug profiles across the six common criteria.

'marketing authorisation' (n = 11), whereas INAMI's value tree considered the same additional criteria but with 'Phase 2' instead of 'PSA response' (n = 11).

Overall, the different groups of decision conference participants agreed in the valuation of performance for the six common criteria that were considered across all four countries, as revealed through the elicitation of their value functions. Fig. 2 plots the value scores of each drug across the six common criteria showing very similar valuations between countries.

The weights of relative importance assigned to the different criteria across the four jurisdictions are shown in Fig. 3. By taking into account the relative swings of the criteria attributes, i.e. the gap between the "lower" and "higher" reference levels, quantitative weights were derived for each criterion using M-MACBETH. The 'OS x HRQoL' aggregated criterion was always assigned the highest relative weight out of 100 ([31,44] for INAMI and AETSA, respectively), followed either by 'treatment discontinuation' ([17,21] for AETSA and TLV, respectively) or 'medical costs impact' ([20,30] for INAMI and AOTMiT, respectively). Depending on the country, the 3rd ranked criterion was then either 'treatment discontinuation' (AOTMiT, INAMI), 'medical costs impact' (TLV), or 'contraindications' (AETSA) and 'PFS' was ranked 4th or 5th. 'Special instructions', although a fundamental criterion across settings, was ranked last in three out of four settings with the 'delivery posology' usually at a higher position, with the exception of TLV where that order was reversed.

In terms of the total weights assigned across the different criteria domains, the Therapeutic Benefit weight ranged from 40% to 54% (for AOTMiT/INAMI and AETSA, respectively), the Safety Profile weight ranged from 20% to 33% (for AOTMiT and TLV, respectively), the Innovation Level weight ranged from 7% to 13% (for TLV and INAMI, respectively) and the Socioeconomic Impact weight ranged from 8% to 30% (for AETSA and AOTMiT, respectively) (Table 3). The above differences in relative weights reflect the different priorities of decision-makers, including the number of fundamental objectives being considered.

3.2. Overall drug rankings and value-for-money analysis

With regards to the overall WPV scores shown in Table 4, enzalutamide consistently yielded the highest score across all four countries, always followed by abiraterone and cabazitaxel. The overall scores of abiraterone and cabazitaxel were in part influenced by a "negative" performance in the 'treatment discontinuation' attribute (19% and 18% respectively) which lay below the lower reference level of the scale (i.e. 10%), affecting negatively their overall value scores.

A stacked bar plot of the drugs' overall WPV scores across all settings is shown in Fig. 4. By using rounded up cost figures for enzalutamide (£24,600), abiraterone (£21,900) and cabazitaxel (£23,900, of which £22,190 related to drug cost and the remainder £1,710 to administration cost) and dividing them with overall WPV scores, their costs per MCDA value unit ranged as follows: (a) enzalutamide: £410 - £501 (for AOTMiT and AETSA, respectively); (b) abiraterone: £1,366 - £9,221 (for INAMI and TLV, respectively); and (c) cabazitaxel: £2,196 - £6,816 (for INAMI and AOTMiT, respectively) (Table 4). The overall value score of each option was driven by the fundamental objectives considered (i.e. criteria influencing the model), the criteria weights which were anchored on reference levels, and the shape of value functions which would influence the value scores.

In terms of value-for-money, cabazitaxel was shown to be dominated by abiraterone, and was very close to being dominated by enzalutamide (i.e. a difference of £500 based on the prices used). Enzalutamide on the other hand was associated with a higher cost (a difference of £2,500 based on the prices used) and a higher overall WPV score compared to abiraterone, with a difference in score ranging between 40.4 and 52.7 value units (for AETSA and TLV, respectively). Cost benefit plots of the different options, using their overall WPV scores versus their purchasing (plus any administration) costs across the four HTA organisations is shown in Fig. 5.

3.3. Similarities and differences in value perceptions across settings

By looking at Table 3 (and Fig. 3) of the results, a number of similarities and differences in value preferences are observed across the four settings. The largest number of evaluation criteria were considered in Andalusia and Belgium (11 each), compared to Sweden and Poland

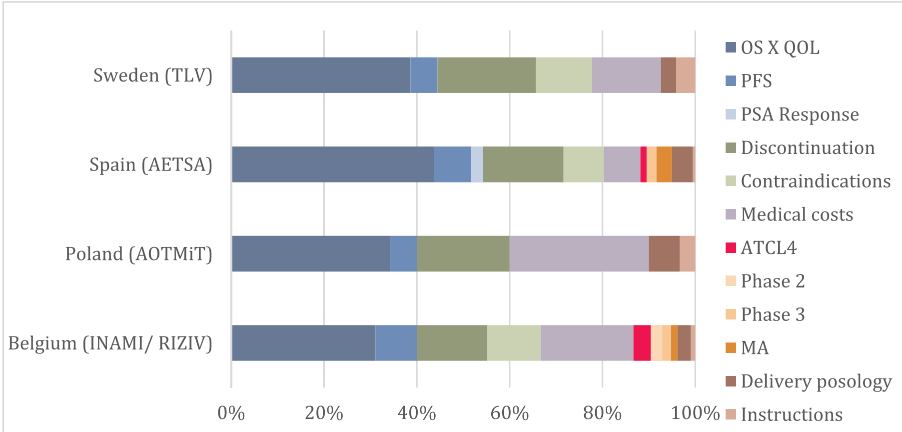

Fig. 3. Relative criteria weights stacked bars across the four HTA settings.

(7 and 6, respectively), partly due to a higher number of Innovation Level criteria (5 and 6, compared to 2 each, respectively). In terms of the relative importance of criteria domains, the Therapeutic Benefit cluster consistently ranked 1st across all settings. The Safety Profile cluster was ranked 2nd in three settings (except for Poland, where the Socioeconomic Impact cluster ranked higher (30% vs 20%)). The Socioeconomic Impact cluster ranked 3rd in Sweden and Belgium but 4th in Andalusia (8%). Finally, the Innovation Level cluster ranked 4th in three countries with the exception of Andalusia where it ranked 3rd (12%). The low relative importance of the Innovation Level cluster partly justifies why a hypothetical change in the final consideration of Innovation Level criteria across the different countries does not influence the ranking of the treatments, as described in the next section.

Despite the observed differences in evaluation criteria considered, the relative criteria weights assigned and the elicited value functions, the overall ranking of the treatments remained identical across countries (Table 4 and Fig. 4) with enzalutamide consistently having the highest score, followed by abiraterone and cabazitaxel in all four settings.

3.4. Sensitivity and robustness analysis

Following each decision conference, deterministic sensitivity analysis was conducted to address parameter uncertainty on criteria weights. Specifically, changes to baseline weights were explored to check their possible impact on treatments' overall value rankings. The results of the sensitivity analysis demonstrated that the ranking of treatments was robust to the relative criteria weights across the different settings.

The most sensitive criterion weight, which could change enzalutamide's ranking order from 1st to 2nd, was ‘PFS’ in the cases of INAMI and AETSA where a 10 and 11 times change (from 8.9% to 90.6% and from 8.0% to 88.5%) respectively, would be required for cabazitaxel to rank 1st and enzalutamide 2nd. In other words, a higher than 10-times difference on the ‘PFS’ weight would be required for cabazitaxel to outperform enzalutamide, with changes of higher order required in other criteria weights for either cabazitaxel or abiraterone to rank 1st, in any of the study settings. Criteria weights were more sensitive to the outperformance of abiraterone by cabazitaxel as the 2nd best treatment. Again, the most sensitive weight was for ‘PFS’ in the INAMI and AETSA cases, where a 2-times change (from 8.9% to 17.4% and from 8.0% to 16.7% respectively) would be needed for cabazitaxel to rank 2nd and abiraterone 3rd. This meant that the lowest change across criteria weights needed for an impact on treatment rankings to be observed was for the case of PFS with INAMI, where at least a 2-times difference was required for abiraterone to be outperformed. For the case of TLV and AOTMiT, the most sensitive criterion was treatment discontinuation in which a 3-times change would be needed (from

Table 4
Overall weighted preference value (WPV) scores, costs and costs per unit of value across the four HTA settings.

Treatments	**Enzalutamide**		**Abiraterone**		**Cabazitaxel**	
HTA agency	**Overall WPV score**	**Ranking per country**	**Overall WPV score**	**Ranking per country**	**Overall WPV score**	**Ranking per country**
Sweden (TLV)	55.1	1st	2.4	2nd	−3.4	3rd
Andalusia (AETSA)	49.1	1st	8.8	2nd	4.4	3rd
Poland (AOTMiT)	59.9	1st	12.1	2nd	3.5	3rd
Belgium (INAMI-RIZIV)	58.6	1st	16.0	2nd	10.9	3rd
Costs (£)	24,600		21,900		23,900	
	Cost per unit of value	**Ranking per country**	**Cost per unit of value**	**Ranking per country**	**Cost per unit of value**	**Ranking per country**
Sweden (TLV)	447	1st	9,221	2nd	N/A	3rd
Andalusia (AETSA)	501	1st	2,496	2nd	5,481	3rd
Poland (AOTMiT)	410	1st	1,805	2nd	6,816	3rd
Belgium (INAMI-RIZIV)	420	1st	1,366	2nd	2,196	3rd

Note: No cost per unit of value was calculated because of the negative overall WPV score (i.e. having a worst overall performance compared to the performance of the lower reference level), which would produce a negative cost per unit of value (£23,900/(-3.4) = −7,072) and would therefore faulty “improve” the median figure of the treatment.
Source: The authors based on input from decision conferences.

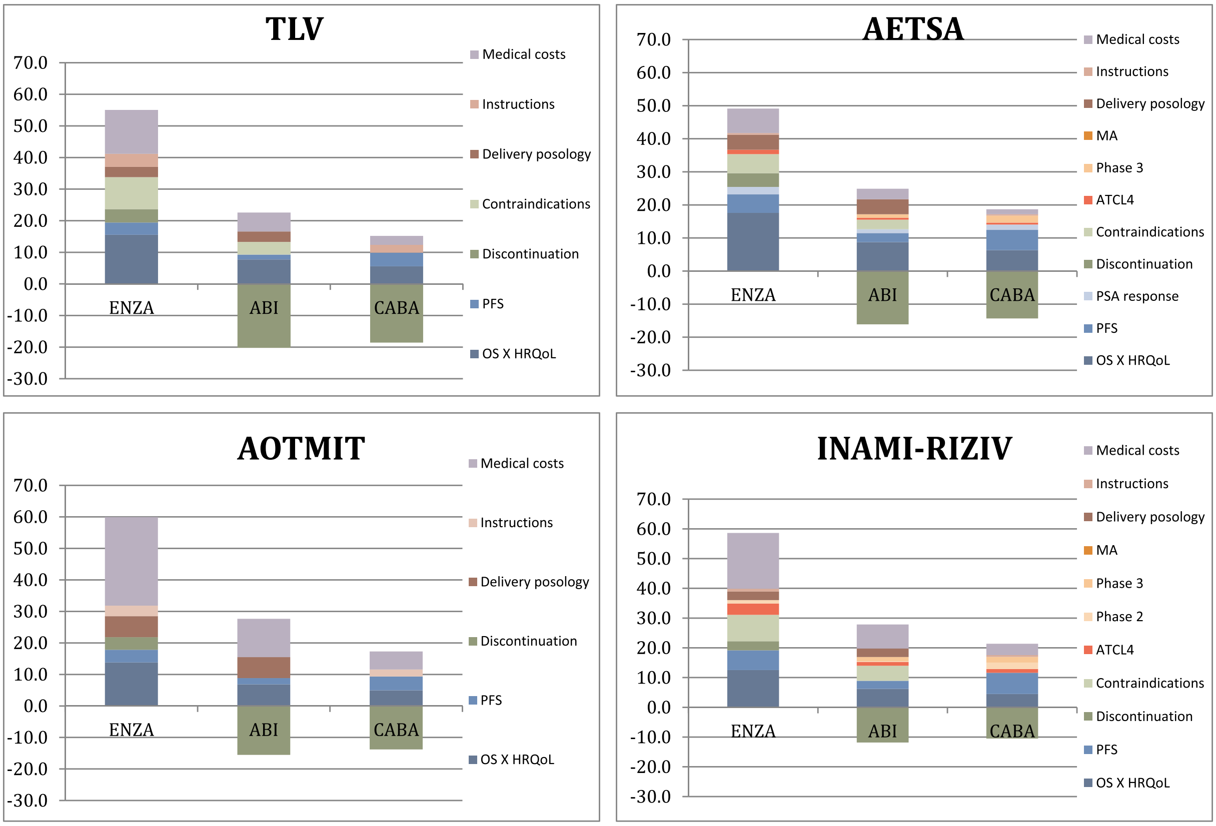

Fig. 4. Stacked bar plot of treatments' overall weighted preference value scores across the four HTA settings.

21.2% to 54.6% and from 20% to 60% respectively) for cabazitaxel to rank 2nd best.

The final consideration of the Innovation Level criteria cluster was explored in greater detail given that their relevance might be disputed. Removing the 'ATCL4' criterion and any spill-over effect criteria (i.e. 'Phase-2', 'Phase-3', 'MA') from the value tree of AETSA and INAMI, and any patient convenience criteria (i.e. 'delivery posology', 'special instructions') from all country value trees would not affect the treatment rankings.

4. Discussion

4.1. Policy implications

This study is the first comparative MCDA exercise utilising the Advance Value Framework and engaging sitting HTA decision-makers across four EU Member States to elicit and compare their preferences in the evaluation of three mCRPC treatments. In doing so, the objective was to test the feasibility of MCDA methods for HTA decision-makers and identify differences in value perceptions.

Based on the evidence used, our results showed that the most valuable therapy for second line mCRPC was enzalutamide, followed by abiraterone and cabazitaxel. Each treatment was assessed and ranked based on their overall WPV scores, reflecting the value of their performance against a set of evaluation criteria, weighted against their relative importance. These overall scores were based on the value preferences of decision-makers that were collected via a decision conference in each setting, yielding a comprehensive and transparent multi-dimensional benefit component. Subsequent consideration of drug costs (purchasing and administration) enabled the demonstration of value-for-money in the form of estimated "cost per unit of value" ratios which showed the second-ranked treatment (abiraterone) to dominate the third (cabazitaxel).

It should be noted that the constructed benefit metric used excludes the cost of the treatments, i.e. the WPV score considers the impact of the technology on medical costs other than the purchasing cost of the technology. Therefore, evaluation of the treatments based solely on their overall WPV scores might not be appropriately designed to inform an HTA decision context that considers the interventions' incremental cost per incremental benefit, but, rather, a value-based approach to reimbursement or pricing negotiation.

Attempting a comparison of the ranking achieved in this exercise with what has occurred in reality might prove challenging, partly because of how the clinical evidence was treated in the exercise, but also because it is not publicly known whether and how any of the additional value dimensions evaluated in the exercise were considered in the relevant HTA decision-making processes. In Sweden, although abiraterone's ICER vs BSC (manufacturer estimate of SEK820,000/QALY) (TLV, 2015a), was lower compared to enzalutamide's ICER vs BSC (TLV best estimate of SEK1,100,000/QALY) (TLV, 2014), or lower vs enzalutamide (SEK800,000/QALY) (TLV, 2015b), TLV assumed that both treatments had the same clinical effect; consequently, TLV focused on a cost-minimisation approach rather than a cost-utility analysis, leading to the implementation of a confidential risk sharing agreement (RSA) as part of which discounts could be provided based on treatment duration. A similar conclusion was reached in Spain, where the Ministry of Health in its Clinical Assessment Report (Informe de Posicionamento Terapeutico - IPT) recommended that there was no clinically relevant difference between the benefit-risk balance of enzalutamide and abiraterone, and therefore decisions should be guided based on drug costs (AEMPS, 2015). Pricing and reimbursement decisions were then taken by the Interministerial Committee for Pricing and Reimbursement, but

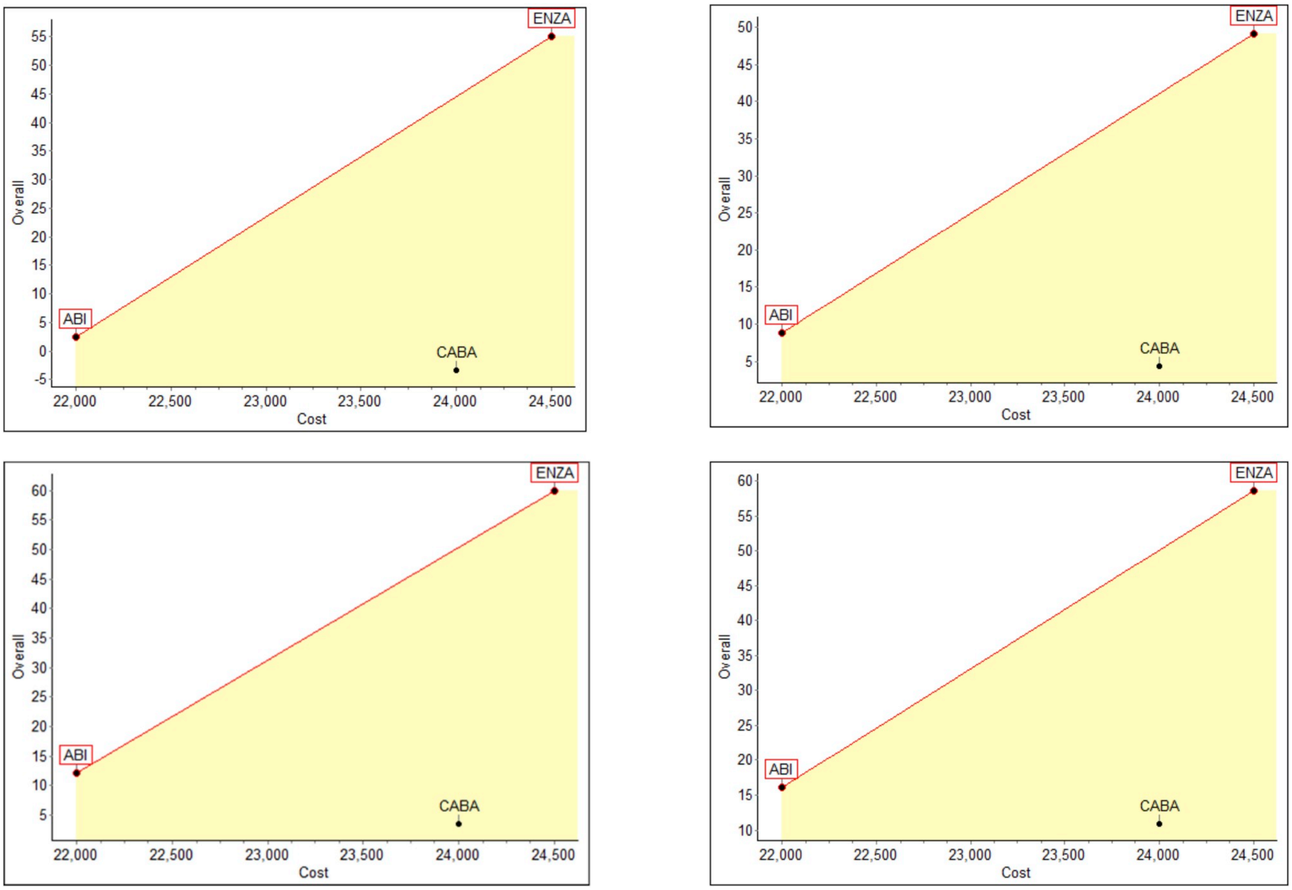

Fig. 5. Cost benefit plots of treatments overall weighted preference value scores versus their purchasing costs across the four HTA settings (TLV top left, AETSA top right, AOTMiT, bottom left, INAMI bottom right). Images produced using the M-MACBETH (beta) software version 3.0.0.

the final assessment is not publicly available. At regional/hospital level, a group of hospital pharmacists conducted a full health (clinical and economic) technology assessment, where enzalutamide and abiraterone were considered to be therapeutically equivalent (GHEMA, 2016). In Poland, although AOTMiT accepted that some additional clinical benefit existed for enzalutamide compared to abiraterone (mainly in secondary endpoints), enzalutamide was not found to be cost-effective compared to abiraterone; however, a confidential RSA enabled a final positive recommendation by AOTMiT (AOTMiT, 2017). The final decision implemented by the Ministry of Health was to reimburse enzalutamide, in a way similar to that of abiraterone (Obwieszczenie, 2017). In Belgium, following an indirect comparison, no clinically relevant differences were found in the treatment outcomes of abiraterone versus enzalutamide (INAMI, 2019); eventually, a managed entry agreement (MEA) enabled reimbursement.

Consequently, based on the evidence used to populate the MCDA model, the hypothetical coverage decisions emerging from the ranking of the treatments based on their overall WPV scores might have been different. Given the higher overall value of enzalutamide compared to abiraterone, a cost minimisation approach or price parity attained between the two, as inferred following the risk sharing agreements in place, might not have been justified.

One reason why our value models differentiate these treatments is because they have captured benefits that go beyond the current formal remits of HTA agencies, therefore, the results should be viewed as 'proof-of-concept' for the purposes of testing this methodology. Furthermore, the decision context addressed in the exercise was a one-off evaluation problem within the indication of mCRPC which might contradict the operational scope of some HTA agencies and health insurance bodies relating to repeated decisions around the reimbursement of drugs across different disease areas.

The extent to which HTA decision-makers can be relied upon, or not, to reflect societal preferences when constructing their value preferences is a very important topic for discussion but not aimed to be addressed in this study. Here, we simply elicited decision-makers' own preferences without considering whether these might be representative for society or not. In reality, evidence in Belgium suggests that health care coverage related preferences of decision-makers differ to those of the public (Cleemput et al., 2018), and, therefore, more research would be needed to reveal such discrepancies.

Overall, the HTA decision-makers that participated in the four decision conferences provided positive feedback on the potential usefulness of the AVF and the MCDA approach in general, raising the prospects of this value framework acting as a decision support tool in the evaluation of new medicines. According to participants, key advantages of the AVF included the feasibility to transparently assess the performance of the options across a number of explicit evaluation criteria, while allowing the elicitation of value trade-offs (i.e. their relative importance), and the framework's overall facilitative nature in the construction and analysis of group value preferences. Our results are in line with past empirical evidence on a different oncology indication (Angelis et al., 2017).

4.2. Challenges of MCDA applications in HTA

The assessment across four settings has offered a number of important insights relating to the application of MCDA in HTA and the challenges this represents. In order for any MCDA methodology to become a useful tool for HTA decision-makers and serve their needs, certain requirements must be met: first, sound methods should be used to ensure technical requirements are fulfilled (Keeney and Raiffa, 1993); second, social aspects of the process should be treated carefully to ensure various socio-technical requirements are fulfilled (Baltussen et al., 2017); and, third, tools and guidelines should be available and tailored for the appropriate audience ensuring that best practice requirements are fulfilled (Phillips, 2017).

Among the first group of technical requirements, one key challenge of MCDA studies in HTA relates to the theoretical properties that are required for the evaluation criteria. Due to the popularity of using a simple additive (i.e. weighted average) value model, the violation of preference-independence is of particular relevance as it might undermine the validity of such models and the insights offered by the results (Marsh et al., 2018; Morton, 2017). Evidence suggests that preference dependencies might exist between health gain and disease severity (Nord et al., 2009), or between OS and HRQoL (Angelis and Kanavos, 2017). The latter also featured strongly in this study, where such a preference dependence between OS and HRQoL was detected during the decision conferences and, as a result, the two criteria were combined together. Beyond combining the two criteria into a common aggregated one, other more technically complex solutions exist for addressing preference dependencies, such as using other functional forms of aggregation for combining scores and weights together, as for example multiplicative models (Chongtrakul et al., 2005). Furthermore, tests for identifying preference dependencies have existed for many years (Currim and Sarin, 1984; Keeney, 1992; Rodrigues et al., 2017).

Other technical challenges relate to the need for evaluation criteria to be non-overlapping so that there is no double counting, and that criteria weights are connected to the attribute ranges (Marsh et al., 2018). If one of these conditions is not satisfied, criteria weights could misrepresent decision makers' true value preferences. Furthermore, a number of cognitive biases may affect value judgments and thus appropriate elicitation protocols and de-biasing tools must be employed (Montibeller and Winterfeldt, 2015).

In order to avoid double-counting, a clear justification of their inclusion is needed, which should be on the grounds of addressing the fundamental objectives of the analysis, rather than be informed based on the existence of available evidence and data (Keeney, 1992; Keeney and Gregory, 2005). This process could be supported by the use of problem structuring tools aiming to distinguish between ‘fundamental objectives’ and ‘means objectives’ (Franco and Montibeller, 2010a), as we adopted in this exercise.

In terms of weighting, asking direct questions for the general importance of criteria is known to be one of the most common mistakes when eliciting value trade-offs (Keeney, 1992, 2002). Instead, sound weighting procedures for the assignment of relative weights should be deployed using explicit lower and higher reference levels of performance (Belton and Stewart, 2002; Keeney, 2002). Ideally, user-friendly indirect technique protocols should be adopted that can reduce bias, similar to what we aimed for in this exercise through the explicit definition of reference levels and the implementation of the qualitative (MACBETH) swing weighting technique.

A further challenge relates to the linking of MCDA results with coverage and resource allocation decisions, possibly through the use of specific value thresholds, that can reflect the efficiency and opportunity cost of funding decisions (Sculpher et al., 2017). In economic evaluation, incremental cost effectiveness ratio (ICER) thresholds are supposed to reflect the opportunity cost of the benefit foregone elsewhere in the health care system that would have resulted from the coverage of alternative technologies (Claxton et al., 2015). Assuming that a QALY-based ICER threshold is accurate, it could be used as a benchmark to create an MCDA value threshold by extrapolating the ICER threshold in proportion to how much of the MCDA model's weight is accounted for by non-QALY value components (Phelps and Madhavan, 2018). Alternatively, following the generation of a multi-dimensional benefit component, purchasing costs could be used to derive treatment cost-value ratios to inform resource allocation decisions within a fixed budget (Peacock et al., 2007), similar to our approach in this exercise with the calculation of the “cost per unit of value”.

4.3. Study limitations

The study has a number of limitations, related both to the clinical evidence used and the MCDA process followed, therefore results should be interpreted with caution. First, in terms of the clinical data used, there was a lack of relative treatment effects; in order to counteract that, absolute treatment effects from different clinical trials were used based on the assumption that they are directly comparable which might not be accurate even for similar patient populations in the relevant studies. As a result, differences in the performance of the options that have been valued might in reality not be statistical significant, e.g. in OS. Ideally, one would need indirect comparisons or a network meta-analysis (NMA) through a mixed treatment comparison (Jansen et al., 2011), therefore, an evidence synthesis step would be required as part of the model-building phase, as for example in the case of assessing the comparative benefit-risk of statins in primary prevention (Tervonen et al., 2015) or second-generation antidepressants (van Valkenhoef et al., 2012).

Second, another limitation relating to clinical evidence could be that only the treatments' impact on HRQoL for the stable disease state was assessed, because no treatment was assumed to have any effect during progression (NICE, 2014). This might not be true for other disease indications in which case the relevant HRQoL attribute would have to capture both the stable and progressive disease states.

Third, there are a number of limitations in terms of the MCDA process adopted. One of them relates to the relatively small number of participants in two of the decision conferences (TLV and AETSA), which could reflect a limited representation of perspectives for the purpose of informing policy-making. Group sizes of between 7 and 15 participants are known to be ideal as they are large enough to represent all major perspectives but small enough to work towards agreement, effectively allowing for efficient group processes to emerge while preserving individuality (Phillips and Phillips, 1993). However, in this instance capturing an all-round set of perspectives was not among the primary aims of the study.

Another issue relates to the value scale of the treatment discontinuation attribute in which the “lower” reference level of “10%” could be perceived as a limitation because it contributed towards the negative partial value scores of two treatments whose performance was worse. This took place in consultation with an oncologist, based on evidence from one of the clinical trials' placebo-controlled arms, because it was perceived to better resemble BSC used in practice; although others might have chosen a different performance level to define the “lower” reference level, the overall ranking of the treatments did not change after altering the lowest reference level to a much less preferred hypothetical performance (20% lower than the worst performing option), while keeping the weights constant.

One major advantage of MCDA is that it can be tailor-made to reflect decision-makers' needs, by taking into account different fundamental objectives through the consideration of various criteria, reflecting their priorities (by eliciting relative weights) and representing their preferences (by eliciting value functions). However, it should be recognised that the emerging differences described above prevent a generalised direct comparison of overall value scores for the alternative options; such a comparison would require identical value trees (i.e. the same set of criteria, weights and value functions across settings), in addition to the same evidence on options performance. The ranking comparisons that we have made in this study using ordinal scales reflect these limitations.

5. Conclusion

In this study we tested the application of AVF, a multi-criteria value framework, in collaboration with HTA decision-makers in order to deduce its feasibility and compare results across settings, in an effort to investigate its potential usefulness and limitations for the purposes of HTA. We found that the AVF methodology can act as a valuable decision support tool because of the transparent construction of value preferences in a collaborative manner, which facilitates the evaluation processes of groups, including the elicitation of value preferences on performance and trade-offs. Although we observed setting-specific differences in value perceptions, the rankings of drugs remained consistent across all countries. In alignment with the evidence used and generated in the study, a coverage decision based on this methodology might have pointed towards a different recommendation denoting differences in value between the first two treatments, in contrast with the cost minimisation approach adopted or the price parity attained between the two in real life.

Despite certain limitations relating to data and process issues and the existence of broader methodological challenges with the use of MCDA in HTA, the present study has demonstrated that an MCDA value framework can, in fact, provide meaningful valuations of novel health technologies, which in turn could inform coverage decisions.

The MCDA methodology adopted enabled participants in the study countries to reflect on certain value dimensions and incorporate these more explicitly in the deliberation process, supporting its use as a transparent value communication tool. Future research could involve similar cross-county case studies, the advancement of MCDA methods in alignment with HTA policy needs, or repeating the study with different participants to understand whether similarities and differences identified in this study can be replicated.

Acknowledgements

This paper has been conducted in the context of the Advance-HTA project, which has received funding from the European Commission's FP-7 Research Programme (grant agreement number: 305983). The views reflected in the paper do not represent the views of the European Commission. M.L. receives funding from the National Institute for Health Research (NIHR), UCL Hospitals Biomedical Research Centre, and CRUK UCL Experimental Cancer Medicine Centre. M.L. has received research funding from Bristol- Myers Squibb (BMS), AstraZeneca/MedImmune, Sanofi and Astellas and has received consulting fees from Janssen Pharmaceuticals, BioNTech and BMS.

We are grateful to Carlos Bana e Costa, Larry Phillips and Douglas Lundin for constructive feedback and suggestions on earlier versions of the paper. This research was made possible through the participation of a number of experts and stakeholders in four decision conferences that took place in collaboration with the Swedish Dental and Pharmaceutical Benefits Agency (TLV) in Stockholm, the Andalusian Health Technology Assessment Agency (AETSA) in Seville, the Polish Agency for Health Technology Assessment and Tariff System (AOTMiT) in Warsaw, and the Belgian Health Insurance Fund (INAMI-RIZIV) in Brussels. We are indebted to all participants for their valuable time and involvement. Thanks are also due to Olina Efthymiadou and Erica Visintin for valuable research assistance with drug costs and innovation spill-over effects calculations. Finally, we would like to thank all partners of the Advance-HTA consortium for their feedback throughout the project. All outstanding errors are our own.

Appendix

Model Building: Alternative treatments compared and evidence considered

The source of evidence used for identifying the performance of options across the evaluation criteria is shown in Table A1.

Model Building: Options performance and references levels

For the case of clinical therapeutic attributes, “lower” reference levels were based on best standard of care (BSC) performance, coming from the respective placebo arm of the *AFFIRM* trial, with the exception of the HRQoL attribute (EQ-5D utility score) that was based on the utility of stable disease with no treatment coming from past NICE TAs (NICE, 2012a, b). The “higher” reference levels were derived by adding a 20% absolute improvement to the performance level of the best performing option, besides for the case of the HRQoL attribute (EQ-5D utility score) that was based on the general Swedish population (Burström et al., 2001). The rationale was to design a value scale incorporating a “global” reference level (Belton and Stewart, 2002), reflecting an “ideal” performance (as proxied by the 20% improvement in best available performance), corresponding to the 100 anchor level of the value scale. This could also offer a flexibility margin to be able to incorporate the performance of future improved options within the same elicited value scale. Consequently, two reference levels within the attribute range were defined in most cases: i) the “lower” reference level (x_l) (i.e. BSC-based satisfactory performance), acting at the same time as the minimum limit of the attribute range (x_*); and ii) the “higher” reference level (x_h) (i.e. 20% better than the best performing option), acting at the same time as the maximum limit of the attribute range (x^*) to give x_* = x_l ≤ x_h = x^*.

A similar, but reverse, logic was used for setting the reference levels of the “treatment discontinuation” attribute in the safety cluster; the “lower” reference level was defined to be equal to the BSC (i.e. placebo) arm of the *AFFIRM* trial. However, contrary to the logic adopted so far for the therapeutic benefit criteria, the “higher” reference level was not set equal to 20% worse than the best performing option (because the lower the performance, the higher the value), but rather equal to the minimum, i.e. worst possible, natural limit of the attribute scale (i.e. 0%) which was regarded as an “ideal” level. In turn, the minimum limit of the scale was derived by worsening the performance of the worst performing treatment option by 20%. A similar approach was used for setting the reference levels of the qualitative “contraindications” attribute, defining the “higher” reference level equal to the maximum (i.e. most attractive) limit of the attribute scale (i.e. none known contraindications) and the “lower” reference level equal to the minimum (i.e. least attractive) limit of the attribute scale.

For the innovation attributes, the “higher” reference level was set either equal to 20% better than the best performing option for the case of natural quantitative attributes (e.g. number of new indications for which the technology is investigated in a given clinical development stage), or equal to the maximum, i.e. best possible, limit of the scale for the case of constructed qualitative attributes (e.g. the existence of any special instructions, the technology's relative market entrance in regards to its ATC Level). Given that the BSC performance was irrelevant to be used as satisfactory level in the innovation attributes and any efforts to derive a "satisfactory" level would be subjective in nature, the minimum limit of the scale for each attribute was used as a "lower" reference level. Therefore the "lower" reference level was based on the worst performance plausible as inferred from the lowest possible limits of the scales, both for the case of natural quantitative attributes (e.g. 0 number of new indications for which the technology is investigated in a given clinical development stage), and the case of constructed qualitative attributes (e.g. worst possible combination of special instructions, 5th entrance at an ATC level).

For the socioeconomics attribute (impact on direct costs), the "higher" reference level was based on BSC's impact on cost (i.e. £0 impact on costs), given that by definition impact on costs for all treatment options are incremental to BSC, and the "lower" reference level was derived by adding a 20% absolute increment to the worst performing option (i.e. to the one with the biggest impact on costs).

"Lower" and "higher" performance reference levels for all criteria at the pre-workshop stage and the basis of their selection are outlined in Table A2 (assuming no impact of luteinizing hormone-releasing hormone analogue).

Model Assessment and Appraisal: Decision conferences

On the day of each decision conference the preliminary model was validated with the participants by revising it cluster by cluster through an open discussion, seeking group consensus and adopting an iterative and interactive model-building process where debate was encouraged and differences of opinion were actively sought.

In terms of the decision-aiding methodology used, the lead author acted as an impartial facilitator with the aim of enhancing content and process interaction, while refraining from contributing to the content of the group's discussions, essentially guiding the group in how to think about the issues but not what to think (Phillips and Bana e Costa, 2007; Schein, 1999).

In terms of facilities, the rooms of the decision conferences had Π-shaped table(s) layout for all the participants to have direct eye to eye contact, with an overhead projector screen and a second portable projector or large TV screen. The M-MACBETH software (more information provided in the MCDA technique section of the main text and below) was operated using a laptop, the screen of which was connected to the projector, and the second screen was used to show the list of the evaluation criteria together with their "lower" and "higher" reference levels.

The decision conferences took place over a full working day or over two half working days; in the former case, there was one lunch break and two coffee breaks throughout the day, whereas in the latter case a coffee break took place around the middle of each session. In each decision conference, the day started with an overview of the MCDA methodology adopted and the description of the preliminary version of the value tree which was then analysed cluster by cluster. At the beginning of each cluster the value tree was validated; the various criteria were explained, followed by a group discussion relating to their relevance and completeness. As a result of this iterative process, some of the criteria were not included because they were perceived as irrelevant or non-fundamental. Schematic illustrations of the final versions of the value trees are shown in Fig. A1. Then, value functions were elicited for the different criteria and relative weights were assigned within the clusters. Finally, relative weights were assigned across clusters, enabling the calculation of the options' overall WPV scores.

Model Assessment and Appraisal: MCDA technique

MACBETH uses seven semantic categories ranging between "no difference" to "extreme difference", in order to distinguish between the value of different performance levels. Based on these qualitative judgements of difference and by analysing judgmental inconsistencies it facilitates the move from ordinal preference modelling, a cognitively less demanding elicitation of preferences, to a quantitative value function. The approach has evolved through the course of theoretical research and real world practical applications, making it an interactive decision support system that facilitates decision-makers' communication. An example of the type of questioning being asked would be "What do you judge to be the difference of value between x' and x"?" where x' and x" are two different performance reference levels of criterion x, across the plausible range (i.e. $x_* \leq x'$, $x'' \leq x^*$). The value judgements matrix for the radiographic tumour progression, i.e. progression free survival (PFS) criterion and its conversion into a value function is provided as an example in Fig. A2.

Following the elicitation of value functions, criteria baseline weights can be elicited. Questions of direct importance for a criterion such as "*How important is a given criterion*?" are known to be as one of the most common mistakes when making value trade-offs because they are assessing them independent of the respective consequences of the options (Keeney, 2002). In contrast, an indirect weighting technique that assesses value trade-offs in tandem with the respective ranges of attributes, i.e. performance reference levels, should be employed. For example, the quantitative swing weighting technique asks for judgments of relative value between 'swings' (i.e. changes from standard lower level x_* to higher reference level x^* on each xth attribute) taking the form "*How would you rank the relative importance of the criteria considering their attributes ranges*, *relative to 100 for the highest-ranked criterion considering its range*?". Each swing, i.e. a relative change from a lower performance level to a higher performance level, is valued between 0 and 100, with the most valuable swing anchored as 100 (von Winterfeldt and Edwards, 1986). Normalised weights are then calculated, as a proportion of each swing weight, so the normalised weights sum to 100%. Relative criteria weights were calculated using an alternative qualitative swing weighting protocol, by using the MACBETH procedure to elicit the differences in attractiveness between the lower and higher reference levels of the different attributes, initially at individual level and then at criteria cluster level (Bana e Costa et al., 2016b; Bana E Costa et al., 2012).

Finally, criteria preference value scores and the respective weights can be combined together through an additive aggregation approach as described in equation (2) (if the adequate conditions of complete and transitive preferences are met as well as multi-attribute preferential independence conditions (von Winterfeldt and Edwards, 1986)).

The M-MACBETH software automatically performs consistency checking between the qualitative judgements expressed, and in addition a second consistency check was manually performed by the lead author which acted as the facilitator to validate the cardinality, i.e. interval nature, of the emerging value scale. This was done by comparing the sizes of the intervals between the proposed scores and inviting participants to adjust them if necessary (Fasolo and Bana e Costa, 2014), a requirement which is essential for the application of simple additive value models.

Table A1
Criteria definition and sources of evidence.

Cluster	Criterion	Definition	Evidence source		
			Abiraterone	Cabazitaxel	Enzalutamide
THERAPEUTIC BENEFIT	Overall survival	The median time from treatment randomisation to death	de Bono et al. (2011)	de Bono et al. (2010)	Scher et al. (2012)
	Health related quality of life	Health related quality of life using the EQ-5D score	Sullivan et al., (2007); TA 255; TA259; TA316	N/A – assumed, based on Sullivan et al., (2007); TA 255; TA259; TA316	Sullivan et al., (2007); TA 255; TA259; TA316
	Radiographic tumour progression	The median survival time on which patients have not experienced disease progression (using RECIST criteria)	de Bono et al. (2011)	de Bono et al. (2010)	Scher et al. (2012)
	PSA response	The proportion of patients having a ≥50% reduction in PSA	Fizazi et al. (2012)	de Bono et al. (2010)	Scher et al., 2012
SAFETY PROFILE	Treatment discontinuation	The proportion of patients discontinuing treatment due to AEs	de Bono et al. (2011)	de Bono et al. (2010)	Scher et al. (2012)
	Contra-indications	The existence of any type of contra-indication accompanying the treatment	EPAR, Prescribing info	EPAR, Prescribing info	EPAR, Prescribing info
INNOVATION LEVEL	ATC Level 1	The technology's relative market entrance in regards to its ATC Level 1 (Anatomical)	WHO ATC index	WHO ATC index	WHO ATC index
	ATC Level 2	The technology's relative market entrance in regards to its ATC Level 2 (Therapeutic)	WHO ATC index	WHO ATC index	WHO ATC index
	ATC Level 3	The technology's relative market entrance in regards to its ATC Level 3 (Pharmacological)	WHO ATC index	WHO ATC index	WHO ATC index
	ATC Level 4	The technology's relative market entrance in regards to its ATC Level 4 (Chemical)	WHO ATC index	WHO ATC index	WHO ATC index
	ATC Level 5	The technology's relative market entrance in regards to its ATC Level 5 (Molecular)	WHO ATC index	WHO ATC index	WHO ATC index
	Phase 1	The number of new indications for which the technology is investigated in Phase 1 clinical trials	ClinicalTrials.gov	ClinicalTrials.gov	ClinicalTrials.gov
	Phase 2	The number of new indications for which the technology is investigated in Phase 2 clinical trials	ClinicalTrials.gov	ClinicalTrials.gov	ClinicalTrials.gov
	Phase 3	The number of new indications for which the technology is investigated in Phase 3 clinical trials	ClinicalTrials.gov	ClinicalTrials.gov	ClinicalTrials.gov
	Marketing authorisation	The number of new indications that the technology has gained an approval for at the stage of marketing authorisation	ClinicalTrials.gov	ClinicalTrials.gov	ClinicalTrials.gov
	Delivery posology	The combination of the delivery system (RoA and dosage form) with the posology (frequency of dosing and duration of administration) of the treatment	EPAR, Prescribing info	EPAR, Prescribing info	EPAR, Prescribing info
	Special instructions	The existence of any special instructions accompanying the administration of the treatment	EPAR, Prescribing info	EPAR, Prescribing info	EPAR, Prescribing info
SOCIOECONOMIC IMPACT	Medical costs impact	The impact of the technology on direct medical costs excluding the purchasing costs of the technology*	BNF 69, Prescribing info, Connock et al., (2011), Riemsa et al., (2013), TA259	BNF 69, Prescribing info, de Bono et al., (2010), TA255	BNF 69, TA316

Notes: * These costs include i) concomitant medications, ii) outpatient visits, diagnostic/laboratory tests, hospitalisations and other monitoring costs (including management AEs), and iii) terminal care.
Source: The authors.

Table A2
Pre-decision conference performance reference levels and basis of selection.

Cluster	Criterion	Attribute	Lower level	Basis	Higher level	Basis
THERAPEUTIC BENEFIT	Overall survival	months	13.6	Best supportive care (BSC)	22.1	20% higher than the best performing option
	Health related quality of life	utility (EQ-5D)	0.72	Utility used for stable disease	0.82	Utility scores of general population
	Radiographic tumour progression	months	2.9	BSC	10.6	20% higher than the best performing option
	PSA response	% patients	1.5	BSC	64.8	20% higher than the best performing option
SAFETY PROFILE	Treatment discontinuation (% of patients)	% patients	10	BSC	0	Highest possible limit of the scale
	Contra-indications	types of contra-indications	Hypersensitivity + hepatic impairment + low neutrophil counts	Lowest possible limit of the scale	None known contra-indications	Highest possible limit of the scale
INNOVATION LEVEL	ATC Level 1	relative market entrance	5th	Lowest possible limit of the scale	1st	Highest possible limit of the scale
	ATC Level 2	relative market entrance	5th	Lowest possible limit of the scale	1st	Highest possible limit of the scale
	ATC Level 3	relative market entrance	5th	Lowest possible limit of the scale	1st	Highest possible limit of the scale
	ATC Level 4	relative market entrance	5th	Lowest possible limit of the scale	1st	Highest possible limit of the scale
	ATC Level 5	relative market entrance	5th	Lowest possible limit of the scale	1st	Highest possible limit of the scale
	Phase 1	number of new indications	0	Lowest possible limit of the scale	10	20% higher than the best performing option
	Phase 2	number of new indications	0	Lowest possible limit of the scale	16	20% higher than the best performing option
	Phase 3	number of new indications	0	Lowest possible limit of the scale	2	20% higher than the best performing option
	Marketing authorisation	number of new indications	0	Lowest possible limit of the scale	1	20% higher than the best performing option
	Delivery Posology	types of delivery system & posology combinations	Oral, every day - one off + IV, every 3 weeks - 1 h*	Lowest possible limit of the scale	Oral, every day - one off*	Highest possible limit of the scale
	Special instructions	types of special instructions	No food + concomitant and/or pre-medication*	Lowest possible limit of the scale	None*	Highest possible limit of the scale
SOCIOECONOMIC IMPACT	Medical costs impact	GBP	10,000	20% higher than the worst performing option (rounded up)	0	BSC

Note: * Assuming no impact on Luteinizing hormone-releasing hormone (LHRH) analogue.
Source: The authors based on the literature.

Fig. A1. Final value trees for metastatic prostate cancer across the four HTA settings. Images produced using the M-MACBETH (beta) software version 3.0.0.

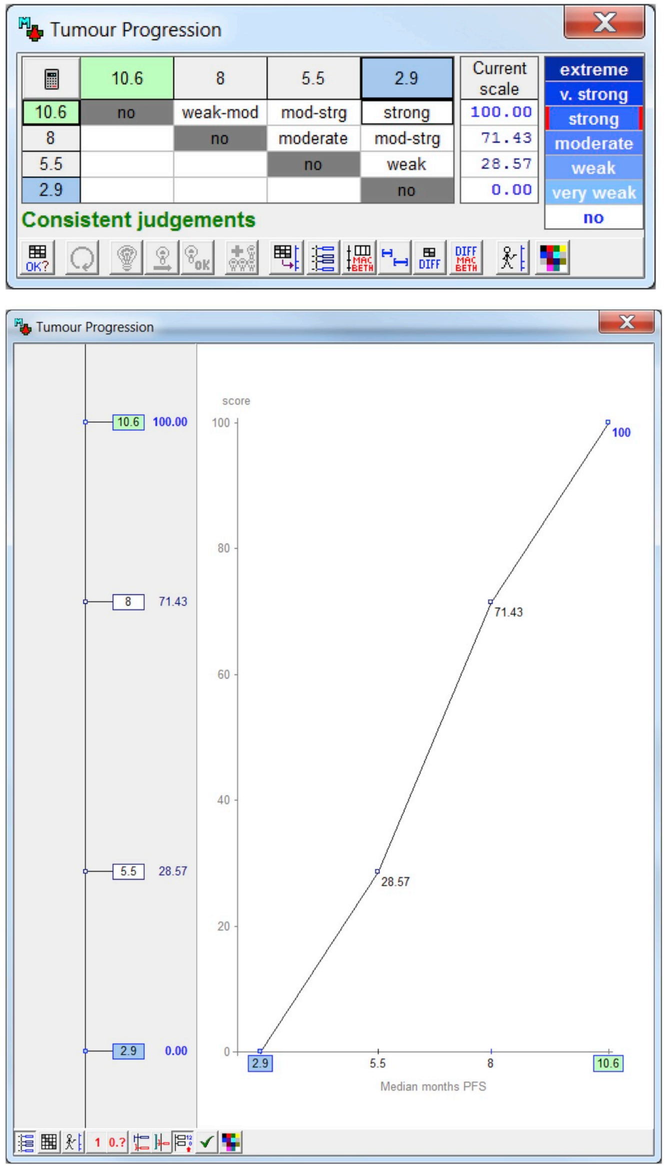

Fig. A2. Example of value judgements matrix for the 'Radiographic tumour progression' i.e. progression free survival (PFS) criterion measured in months and its conversion into a value function.

Note: In the Radiographic tumour progression' i.e. progression free survival (PFS) example, measured in median months, the question asked was the following: "*What do you judge to be the difference of value between 2.9 and 10.6 months PFS? No difference, very weak, weak, moderate, strong, very strong, or extreme?*" Once a decision was reached (by consensus or majority voting), the next question came along: "*What do you judge to be the difference of value between 5.5 and 10.6 months PFS? No difference, very weak, weak, moderate, strong, very strong, or extreme?*" The same process was followed until value judgments for all the different combinations of attribute levels were elicited, filling in the different rows from the right-hand side (i.e. lower range) to the left-hand side (i.e. higher range). Image produced using the M-MACBETH (beta) software version 3.0.0.

References

AEMPS, 2015. Informe de Posicionamiento Terapéutico Pt-Enzalutamida/V1/30072015. Agencia Española de Medicamentos y Productos Sanitarios.

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., 2018. Evaluating the benefits of new drugs in health technology assessment using multiple criteria decision analysis: a case study on metastatic prostate cancer with the dental and pharmaceuticals benefits agency (TLV) in Sweden. MDM Pol. Pract. 3.

Angelis, A., Kanavos, P., 2016. 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. https://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., Kanavos, P., 2017. Multiple criteria decision analysis (MCDA) for evaluating new medicines in health technology assessment and beyond: the advance value framework. Soc. Sci. Med. 188, 137–156.

Angelis, A., Kanavos, P., Montibeller, G., 2017. Resource allocation and priority setting in health care: a multi-criteria decision analysis problem of value? Glob. Pol. 8 (52), 76–83. https://doi.org/10.1111/1758-5899.12387.

Angelis, A., Lange, A., Kanavos, P., 2018. 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. 19 (1), 123–152. https://doi.org/10.1007/s10198-017-0871-0.

Angelis, A., Montibeller, G., Hochhauser, D., Kanavos, P., 2017. Multiple criteria decision analysis in the context of health technology assessment: a simulation exercise on metastatic colorectal cancer with multiple stakeholders in the English setting. BMC Med. Inf. Decis. Mak. 17.

AOTMiT, 2017. Rekomendacja Nr 19/2017 Z Dnia 27 Marca 2017 R. Prezesa Agencji Oceny Technologii Medycznych I Taryfikacji W Sprawie Objęcia Refundacją Produktu Leczniczego Xtandi, Enzalutamidum, Kapsułki Miękkie, 40 Mg, 112 Kaps. w ramach programu lekowego „Leczenie opornego na kastrację raka gruczołu krokowego z przerzutami (ICD-10 C-61). http://bipold.aotm.gov.pl/assets/files/zlecenia_mz/2017/008/REK/RP_19_2017_Xtandi.pdf.

Bach, P., 2015. DrugAbacus App. Memorial Sloan Kettering Cancer Center.

Baltussen, R., Jansen, M.P.M., Bijlmakers, L., Grutters, J., Kluytmans, A., Reuzel, R.P., et al., 2017. Value assessment frameworks for HTA agencies: the organization of evidence-informed deliberative processes. Value Health 20, 256–260.

Bana e Costa, C., Beinat, E., 2005. Model-structuring in Public Decision-Aiding. Operational Research Working Papers. London School of Economics and Political Science, London.

Bana e Costa, C., De Corte, J., Vansnick, J., 2016a. M-MACBETH website. http://m-macbeth.com/.

Bana e Costa, C., De Corte, J., Vansnick, J., 2016b. On the mathematical foundations of MACBETH. In: Greco, S., Ehrgott, M., Figueira, J. (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys. Springer, New York.

Bana e Costa, C., Lourenço, J., Oliveira, M., Bana e Costa, J., 2014. A socio-technical approach for group decision support in public strategic planning: the pernambuco PPA case. Group Decis. Negot. 23, 5–29.

Bana e Costa, C., Vansnick, J., 1999. The MACBETH approach: basic ideas, software, and an application. In: Meskens, N., Roubens, M. (Eds.), Advances in Decision Analysis. Springer Netherlands, pp. 131–157.

Bana e Costa, C.A., Corrêa, É.C., De Corte, J.-M., Vansnick, J.-C., 2002. Facilitating bid evaluation in Public call for tenders: a socio-technical approach. Omega 30, 227.

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., Ensslin, L., Corrêa, É.C., Vansnick, J.-C., 1999. Decision Support Systems in action: integrated application in a multicriteria decision aid process. Eur. J. Oper. Res. 113, 315–335.

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.

Belton, V., Stewart, T., 2002. Multiple Criteria Decision Analysis: an Integrated Approach. Kluwer Academic Publishers, Dordrecht.

Berthold, D.R., Pond, G.R., Soban, F., de Wit, R., Eisenberger, M., Tannock, I.F., 2008. Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer: updated survival in the TAX 327 study. J. Clin. Oncol.: Offic. J. Am. Soc. Clin. Oncol. 26, 242–245.

Bianchini, D., Lorente, D., Rodriguez-Vida, A., Omlin, A., Pezaro, C., Ferraldeschi, R., et al., 2014. Antitumour Activity of Enzalutamide (MDV3100) in Patients with Metastatic Castration-Resistant Prostate Cancer (CRPC) Pre-treated with Docetaxel and Abiraterone. vol. 50. European Journal Of Cancer, Oxford, England, pp. 78–84 1990.

BNF, 2015. British National Formulary, vol. 69. https://www.bnf.org/.

Burström, K., Johannesson, M., Diderichsen, F., 2001. Swedish population health-related quality of life results using the EQ-5D. Qual. Life Res.: Int. J. Qual. Life Asp. Treat. Care Rehabilitat. 10, 621–635.

Center, M.M., Jemal, A., Lortet-Tieulent, J., Ward, E., Ferlay, J., Brawley, O., et al., 2012. International variation in prostate cancer incidence and mortality rates. Eur. Urol. 61, 1079–1092.

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.: Offic. J. Eur. Soc. Med. Oncol. ESMO 26, 1547–1573.

Chongtrakul, P., Sumpradit, N., Yoongthong, W., 2005. ISafE and the Evidence-Based Approach for Essential Medicines Selection in Thailand. Essential Drugs Monitor, pp. 18–19.

Claxton, K., Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S., et al., 2015. Methods for the estimation of the National Institute for Health and care excellence cost- effectiveness threshold. Health Technol. Assess. 19, 1–503.

Cleemput, I., Devriese, S., Kohn, L., Devos, C., van Til, J., Groothuis-Oudshoorn, C.G.M., et al., 2018. What does the public want? Structural consideration of citizen preferences in health care coverage decisions. MDM Pol. Pract. 3.

Clement, F.M., Harris, A., Li, J.J., Yong, K., Lee, K.M., Manns, B.J., 2009. Using effectiveness and cost-effectiveness to make drug coverage decisions: a comparison of Britain, Australia, and Canada. J. Am. Med. Assoc. 302, 1437–1443.

Cohen, D., 2017. Cancer drugs: high price, uncertain value. BMJ 359.

Collins, R., Fenwick, E., Trowman, R., Perard, R., Norman, G., Light, K., et al., 2007. A systematic review and economic model of the clinical effectiveness and cost-effectiveness of docetaxel in combination with prednisone or prednisolone for the treatment of hormone-refractory metastatic prostate cancer. Health Technol. Assess. 11, iii.

Connock, M., Cummins, E., Shyangdan, D., Hall, B., Grove, A., Clarke, A., 2011. Abiraterone Acetate for the Treatment of Metastatic, Castrate-Resistant Prostate Cancer Following Previous Cytotoxic Chemotherapy: A Single Technology Appraisal. Warwick Evidence.

Currim, I.S., Sarin, R.K., 1984. A comparative evaluation of multiattribute consumer preference models. Manag. Sci. 30, 543–561.

Danner, M., Hummel, J.M., Volz, F., van Manen, J.G., Wiegard, B., Dintsios, C.-M., et al., 2011. Integrating patients' views into health technology assessment: analytic hierarchy process (AHP) as a method to elicit patient preferences. Int. J. Technol. Assess. Health Care 27, 369–375.

de Bono, J.S., Logothetis, C.J., Molina, A., Fizazi, K., North, S., Chu, L., et al., 2011. Abiraterone and increased survival in metastatic prostate cancer. N. Engl. J. Med. 364, 1995–2005.

de Bono, J.S., Oudard, S., Ozguroglu, M., Hansen, S., Machiels, J.-P., Kocak, I., et al., 2010. Prednisone plus cabazitaxel or mitoxantrone for metastatic castration-resistant prostate cancer progressing after docetaxel treatment: a randomised open-label trial. Lancet 376, 1147–1154.

Devlin, N., Lorgelly, P., 2017. QALYs as a measure of value in cancer. J. Canc. Pol. 11, 19–25.

Devlin, N., Sussex, J., 2011. Incorporating Multiple Criteria in HTA: Methods and Processes. Office of Health Economics, London.

Drummond, M., Tarricone, R., Torbica, A., 2013. Assessing the added value of health technologies: reconciling different perspectives. Value Health: J. Int. Soc. Pharmacoecon. Outcomes Res. 16, S7–S13.

Efthymiadou, O., Mossman, J., Kanavos, P., 2019. Health related quality of life aspects not captured by EQ-5D-5L: results from an international survey of patients. Health Policy 123, 159–165.

EMA, 2016a. Jevtana (Cabazitaxel) EPAR - Product Information. European Medicines Agency.

EMA, 2016b. Xtandi (Enzalutamide) EPAR - Product Information. European Medicines Agency.

EMA, 2016c. Zytiga (Abiraterone) EPAR - Product Information. European Medicines Agency.

Faden, R.R., Chalkidou, K., Appleby, J., Waters, H.R., Leider, J.P., 2009. Expensive cancer drugs: a comparison between the United States and the United Kingdom. Milbank Q. 87, 789–819.

Fasolo, B., Bana e Costa, C.A., 2014. Tailoring value elicitation to decision makers' numeracy and fluency: expressing value judgments in numbers or words. Omega 44, 83–90.

Fizazi, K., Scher, H.I., Molina, A., Logothetis, C.J., Chi, K.N., Jones, R.J., et al., 2012. Abiraterone acetate for treatment of metastatic castration-resistant prostate cancer: final overall survival analysis of the COU-AA-301 randomised, double-blind, placebo-controlled phase 3 study. Lancet Oncol. 13, 983–992.

Franco, L., Montibeller, G., 2010a. Problem Structuring for Multicriteria Decision Analysis Interventions Analysis Interventions. Wiley Encyclopedia of Operations Research and Management Science.

Franco, L.A., Montibeller, G., 2010b. Facilitated modelling in operational research. Eur. J. Oper. Res. 205, 489–500.

GHEMA, 2016. ENZALUTAMIDA en cáncer próstata metastásico resistente a la castración (previo a quimioterapia). Grupo de Evaluación de Novedades, EStandarización e Investigación en Selección de Medicamentos.

Goetghebeur, 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.

Goetghebeur, M.M., Wagner, M., Khoury, H., Rindress, D., Grégoire, J., Deal, C., 2010. Combining multicriteria decision analysis, ethics and health technology assessment: applying the EVIDEM decisionmaking framework to growth hormone for Turner syndrome patients. Cost Eff. Resour. Allocation 8, 1.

Horwich, A., Parker, C., de Reijke, T., Kataja, V., 2013. Prostate cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann. Oncol.: Offic. J. Eur. Soc. Med. Oncol. ESMO 24 (Suppl. 6), vi106–vi114.

Ijzerman, M.J., van Til, J.A., Snoek, G.J., 2008. Comparison of two multi-criteria decision techniques for eliciting treatment preferences in people with neurological disorders. Patient 1, 265–272.

INAMI, 2019. Remboursement de médicaments : Décisions ministérielles et rapports d'évaluation de la CRM. https://www.inami.fgov.be/fr/programmes-web/Pages/applications-rapports-crm.aspx.

Jansen, J.P., Fleurence, R., Devine, B., Itzler, R., Barrett, A., Hawkins, N., et al., 2011. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health: J. Int. Soc. Pharmacoecon. Outcomes Res. 14, 417–428.

Jaramillo, H.E.C., Goetghebeur, M., Moreno-Mattar, O., 2016. Testing multi-criteria decision analysis for more transparent resource-allocation decision making in Colombia. 32, 307–314.

Kanavos, P., Angelis, A., 2013. Multiple Criteria Decision Analysis for Value Based Assessment of New Medical Technologies: a Conceptual Framework. The LSE Health Working Paper Series in Health Policy and Economics. London School of Economics and Political Science.

Kearns, B., Lloyd Jones, M., Stevenson, M., Littlewood, C., 2013. Cabazitaxel for the second-line treatment of metastatic hormone-refractory prostate cancer: a NICE single technology appraisal. PharmacoEconomics 31, 479–488.

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., 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.

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.

London School of Economics, 2019. ADVANCE_HTA Project. http://www.lse.ac.uk/lse-

health/research/projects/advance-hta.
Loriot, Y., Bianchini, D., Ileana, E., Sandhu, S., Patrikidou, A., Pezaro, C., et al., 2013. Antitumour activity of abiraterone acetate against metastatic castration-resistant prostate cancer progressing after docetaxel and enzalutamide (MDV3100). Ann. Oncol.: Offic. J. Eur. Soc. Med. Oncol. ESMO 24, 1807–1812.
Malone, D.C., Berg, N.S., Claxton, K., Garrison Jr., L.P., Ijzerman, M., Marsh, K., et al., 2016. International society for pharmacoeconomics and outcomes research comments on the American society of clinical oncology value framework. J. Clin. Oncol.: Offic. J. Am. Soc. Clin. Oncol. 34, 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. Pharmacoecon. 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.
Marsh, K., Sculpher, M., Caro, J.J., Tervonen, T., 2018. The Use of MCDA in HTA: Great Potential, but More Effort Needed. Value in Health.
Montibeller, G., Winterfeldt, D., 2015. Cognitive and motivational biases in decision and risk analysis. Risk Anal.: Int. J. 35, 1230–1251.
Morton, A., 2017. Treacle and smallpox: two tests for multicriteria decision analysis models in health technology assessment. Value Health 20, 512–515.
Nadler, E., Eckert, B., Neumann, P.J., 2006. Do oncologists believe new cancer drugs offer good value? The Oncologist 11, 90–95.
NICE, 2012. Abiraterone for castration-resistant metastatic prostate cancer previously treated with a docetaxel-containing regimen. Technology Appraisal Guidance, vol. 259 National Institute for Health and Care Excellence.
NICE, 2012. Cabazitaxel for hormone-refractory metastatic prostate cancer previously treated with a docetaxel-containing regimen. Technology Appraisal Guidance, vol. 255 National Institute for Health and Care Excellence.
NICE, 2014. Enzalutamide for metastatic hormone-relapsed prostate cancer previously treated with a docetaxel - containing regimen. Technology Appraisal Guidance, vol. 316 National Institute for Health and Care Excellence.
Nicod, E., Kanavos, P., 2012. Commonalities and differences in HTA outcomes: a comparative analysis of five countries and implications for coverage decisions. Health Policy 108, 167–177.
NIH, 2016. ClinicalTrials.gov. US National Institutes of Health.
Nord, E., Daniels, N., Kamlet, M., 2009. QALYs: some challenges. Value Health 12, S10–S15.
Obwieszczenie, M., 2017. Obwieszczenie Ministra Zdrowia Z Dnia 25 Października 2017 R. W Sprawie Wykazu Refundowanych Leków, Środków Spożywczych Specjalnego Przeznaczenia Żywieniowego Oraz Wyrobów Medycznych (DZ. URZ. Min. Zdr. 2017.105). http://www.bip.mz.gov.pl/legislacja/akty-prawne-1/obwieszczenie-ministra-zdrowia-z-dnia-25-pazdziernika-2017-r-w-sprawie-wykazu-refundowanych-lekow-srodkow-spozywczych-specjalnego-przeznaczenia-zywieniowego-oraz-wyrobow-medycznych-na-1-listopada-20/.
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.
Phelps, C., Madhavan, G., 2018. Resource allocation in decision support frameworks. Cost Eff. Resour. Allocation 16.
Phillips, L., 1984. A theory of requisite decision models. Acta Psychol. 56, 29–48.
Phillips, L., 2007. Decision conferencing. In: Edwards, W., Miles, R., von Winterfeldt, D. (Eds.), Advances in Decision Analysis: from Foundations to Applications. Cambridge University Press, Cambridge.
Phillips, L., Phillips, M., 1993. Facilitated work groups: theory and practice. J. Oper. Res. Soc. 44.
Phillips, L.D., 2017. Best practice for MCDA in healthcare. In: Marsh, K., Goetghebeur, M., Thokala, P., Baltussen, R. (Eds.), Multi-Criteria Decision Analysis to Support Healthcare Decisions. Springer.
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.
Radaelli, G., Lettieri, E., Masella, C., Merlino, L., Strada, A., Tringali, M., 2014. Implementation of EUnetHTA core Model® in Lombardia: the VTS framework. Int. J. Technol. Assess. Health Care 30, 105–112.
Riemsa, R., Ramaekers, B., Tomini, F., Wolff, R., van Asselt, A., Joore, M., et al., 2013. Abiraterone for the Treatment of Chemotherapy Naïve Metastatic Castration-Resistant Prostate Cancer: a Single Technology Appraisal. Kleijnen Systematic Reviews Ltd.
Rodrigues, T.C., Montibeller, G., Oliveira, M.D., Bana E Costa, C.A., 2017. Modelling multicriteria value interactions with Reasoning Maps. Eur. J. Oper. Res. 258, 1054–1071.
Schein, E., 1999. Process Consultation Revisited: Building the Helping Relationship. Addison–Wesley, Reading.
Scher, H.I., Fizazi, K., Saad, F., Taplin, M.-E., Sternberg, C.N., Miller, K., et al., 2012. Increased survival with enzalutamide in prostate cancer after chemotherapy. N. Engl. J. Med. 367, 1187–1197.
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.: Offic. J. Am. Soc. Clin. Oncol. 33, 2563–2577.
Sculpher, M., Claxton, K., Pearson, S.D., 2017. Developing a value framework: the need to reflect the opportunity costs of funding decisions. Value Health 20, 234–239.
Shih, Y.C.T., Ganz, P.A., Aberle, D., Abernethy, A., Bekelman, J., Brawley, O., et al., 2013. Delivering high-quality and affordable care throughout the cancer care continuum. J. Clin. Oncol. 31, 4151–4157.
Sullivan, P., Mulani, P., Fishman, M., Sleep, D., 2007. Quality of life findings from a multicenter, multinational, observational study of patients with metastatic hormone-refractory prostate cancer. Int. J. Qual. Life Asp. Treat. Care Rehabilit. Offic. J. Int. Soc. Qual. Life Res. 16, 571–575.
Sulmasy, D., Moy, B., 2014. Debating the oncologist's role in defining the value of cancer care: our duty is to our patients. J. Clin. Oncol. 32, 4039–4041.
Sussex, J., Rollet, P., Garau, M., Schmitt, C., Kent, A., Hutchings, A., 2013. A pilot study of multicriteria decision analysis for valuing orphan medicines. Value Health 16, 1163–1169.
Tannock, I.F., de Wit, R., Berry, W.R., Horti, J., Pluzanska, A., Chi, K.N., et al., 2004. Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer. N. Engl. J. Med. 351, 1502–1512.
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. 35, 859–871 An International Journal of the Society for Medical Decision Making.
Thokala, P., 2011. Multi Criteria Decision Analysis for Health Technology Assessment: Report by the Decision Support Unit. University of Sheffield, Sheffield.
TLV, 2014. Designation 2775/2013. Tandvårds- Och Läkemedelsförmånsverket.
TLV, 2015. Designation 4774/2014. Tandvårds- Och Läkemedelsförmånsverket.
TLV, 2015. Designation 4852/2014. Tandvårds- Och Läkemedelsförmånsverket.
Tony, M., Wagner, M., Khoury, H., Rindress, D., Papastavros, T., Oh, P., et al., 2011. Bridging health technology assessment (HTA) with multicriteria decision analyses (MCDA): field testing of the EVIDEM framework for coverage decisions by a public payer in Canada. BMC Health Serv. Res. 11, 329.
Torre, L.A., 2015. Global cancer statistics, 2012. CA A Cancer J. Clin. 65, 87.
van Valkenhoef, G., Tervonen, T., Zhao, J., de Brock, B., Hillege, H.L., Postmus, D., 2012. Multicriteria benefit-risk assessment using network meta-analysis. J. Clin. Epidemiol. 65, 394–403.
von Winterfeldt, D., Edwards, W., 1986. Decision Analysis and Behavioral Research. Cambridge University Press, Cambridge.
Wagner, M., Khoury, H., Bennetts, L., Berto, P., Ehreth, J., Badia, X., et al., 2017. Appraising the holistic value of Lenvatinib for radio-iodine refractory differentiated thyroid cancer: a multi-country study applying pragmatic MCDA. BMC Canc. 17.
World Health Organisation Collaborating Centre, 2016. ATC/DDD Index 2016. World Health Organisation Collaborating Centre.
Wouters, O.J., Naci, H., Samani, N.J., 2015. QALYs in cost-effectiveness analysis: an overview for cardiologists. Heart (British Cardiac Society) 101, 1868–1873.