Journal of Biomedical Informatics 59 (2016) 201–208

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Combining multi-criteria decision analysis and mini-health technology assessment: A funding decision-support tool for medical devices in a university hospital setting

Nicolas Martelli [a,b,*], Paul Hansen [c], Hélène van den Brink [b], Aurélie Boudard [d], Anne-Laure Cordonnier [d], Capucine Devaux [a], Judith Pineau [a], Patrice Prognon [a], Isabelle Borget [b,e]

[a] Pharmacy Department, Georges Pompidou European Hospital, AP-HP, 20 rue Leblanc, 75015 Paris, France
[b] University Paris-Sud, GRADES, Faculty of Pharmacy, 5 rue Jean-Baptiste Clément, 92290 Châtenay-Malabry, France
[c] Department of Economics, University of Otago, Dunedin 9054, New Zealand
[d] Therapeutic Evaluation Unit, General Agency of Equipment and Health Products, AP-HP, 7, rue du Fer à Moulin, 75005 Paris, France
[e] Department of Health Economics, Gustave Roussy Institute, 114, rue Edouard-Vaillant, 94805 Villejuif, France

ARTICLE INFO

Article history:
Received 17 August 2015
Revised 29 November 2015
Accepted 7 December 2015
Available online 15 December 2015

Keywords:
Medical devices
Hospital
Multi-criteria decision analysis
Mini-health technology assessment

ABSTRACT

Background: At the hospital level, decisions about purchasing new and oftentimes expensive medical devices must take into account multiple criteria simultaneously. Multi-criteria decision analysis (MCDA) is increasingly used for health technology assessment (HTA). One of the most successful hospital-based HTA approaches is mini-HTA, of which a notable example is the Matrix4value model.
Objectives: To develop a funding decision-support tool combining MCDA and mini-HTA, based on Matrix4value, suitable for medical devices for individual patient use in French university hospitals – known as the IDA tool, short for 'innovative device assessment'.
Methods: Criteria for assessing medical devices were identified from a literature review and a survey of 18 French university hospitals. Weights for the criteria, representing their relative importance, were derived from a survey of 25 members of a medical devices committee using an elicitation technique involving pairwise comparisons. As a test of its usefulness, the IDA tool was applied to two new drug-eluting beads (DEBs) for transcatheter arterial chemoembolization.
Results: The IDA tool comprises five criteria and weights for each of two over-arching categories: risk and value. The tool revealed that the two new DEBs conferred no additional value relative to DEBs currently available.
Conclusions: Feedback from participating decision-makers about the IDA tool was very positive. The tool could help to promote a more structured and transparent approach to HTA decision-making in French university hospitals.

1. Introduction

Health technology assessment (HTA) is increasingly performed by health care agencies worldwide to support decision-making concerning the uptake of new health technologies such as drug therapies, equipment and medical devices. HTA, which is a multi-disciplinary field bridging scientific evidence and policy-making [1], considers a wide range of aspects, including medical, social, ethical and economic implications of the development, diffusion and use of health technologies [2]. HTA has spread beyond just national health care agencies; many hospitals have developed local HTA models with respect to purchasing new and oftentimes expensive health technologies [3].

One of the most successful hospital-based HTA approaches is mini-HTA [4,5]. This decision tool is based on a checklist designed for rapid assessment of four central aspects: technology, patient, organization and economy. The key to the success of mini-HTA is its capacity to integrate the views of end-users more effectively into hospital policy actions, which has been identified as a decisive factor in implementing hospital-based HTA [6,7].

Several authors have suggested that the future development of HTA will necessarily incorporate multi-criteria decision analysis (MCDA) [8–13]. MCDA is a methodology for helping

* Corresponding author at: Pharmacy Department, Georges Pompidou European Hospital, AP-HP, 20, rue Leblanc, 75015 Paris, France. Tel.: +33 156092575; fax: +33 156093657.
E-mail address: nicolas.martelli@aphp.fr (N. Martelli).

http://dx.doi.org/10.1016/j.jbi.2015.12.002
1532-0464/

decision-makers to evaluate alternatives in the context of considering multiple criteria simultaneously [8]. Multi-attribute value theory is the MCDA method most widely used in HTA, particularly with respect to approaches based on additive models for aggregating alternatives' performance across the multiple criteria [8]. A wide range of techniques for eliciting decision-makers' preferences with respect to weights on the criteria, representing their relative importance, is available [14–16].

Experiments testing approaches combining HTA and MCDA have recently been performed, including the well-known "EVIDEM" project [17,18]. Approaches combining MCDA and hospital-based HTA have also emerged, several of which use the mini-HTA model – notably the Matrix4value model [5,19]. This model comprises six criteria extracted from the mini-HTA form for each of two over-arching categories: risk and value [5]. For each of these categories, an overall score for each device under consideration is produced by multiplying the expected performance score on each criterion by its weight and then summing the weighted part-scores. These two overall scores for each device can be illustrated graphically in a similar fashion to a cost-effectiveness quadrant, which can help decision-makers to discriminate, and ultimately choose, between competing alternatives.

A tool combining MCDA and mini-HTA, such as Matrix4value, could potentially improve hospital-based HTA activities in French university hospitals by supporting a common, formal and transparent framework for evaluating new medical devices [20]. However, earlier tools and methods have several important weaknesses.

First, as we pointed out in an earlier study [20], some of the criteria within the mini-HTA model are unsuitable for medical devices for individual patient use. This issue is important because local HTAs in French hospitals do not necessarily involve the same processes or stakeholders for medical devices with respect to collective and individual patient use respectively [21]. Second, the simple weight-elicitation technique used by Matrix4value based on Likert scales has several drawbacks; for example, scales have been shown to suffer from biases associated with decision-makers not employing the scale's full range to represent their preferences [22]. These drawbacks could be remediated by using a more robust methodology [17,22]. Finally, in order to encourage medical device committees of French hospitals to adopt them, methodologies need to be easy-to-use, cost-effective and reproducible over time (which has not always been so in the past).

Thus, the objective of the study reported in this article is to design and apply a mini-HTA/MCDA tool suitable for assessing medical devices for individual patient use in French university hospitals.

2. Methods

Consistent with most MCDA approaches in use internationally [23] and based on the same rationale underpinning the development of Matrix4value, we decided to build simple additive models for the over-arching risk and value categories respectively. Such 'compensatory' models, where an alternative's strengths on one criterion can offset its weaknesses on one or more other criteria, have been found to accurately reflect decision-makers' preferences [24].

We followed a three-step approach to develop the mini-HTA/MCDA tool – which we named IDA, short for 'innovative device assessment'. First, we selected relevant criteria for assessing new medical devices for individual patient use based on a literature review and a survey of 18 French university hospitals. We then implemented a weight-elicitation technique that appropriately balances methodological rigor and ease of use. Finally, as a test of its usefulness, we applied the IDA tool under real-world conditions in a university hospital by assessing two new medical devices relative to one currently available.

2.1. Selecting decision criteria

We performed a literature review of mini-HTA-like models to identify criteria considered in local HTAs for medical devices. We also surveyed 18 French university hospitals to identify criteria for assessing new medical devices. Details about the review and survey are available in an earlier article [20].

The criteria identified from both sources were coded independently by the first and sixth authors (NM and CD) and assigned to one of the four mini-HTA perspectives mentioned earlier: technology, patient, organization and economy. Criteria that, in essence, referred to the same concept were grouped under a common code; for example, "medical benefit", "clinical benefit" and "health benefit" can be grouped together under the code "CLINICAL BENEFIT". Earlier studies of health decision criteria standardized terms in a similar fashion [25].

We analyzed the criteria from both sources most frequently considered for each mini-HTA perspective, and compared their similarities and differences. Only criteria suitable for medical devices for individual patient use were selected; for example, we discarded criteria concerning devices' physical-space impact. We further reduced the number of criteria by retaining only those from both the literature review and the survey. Finally, consistent with conventional MCDA modelling guidelines, we did our best to ensure that the criteria selected are complete, non-redundant, operational and mutually independent [26].

2.2. Determining criterion weights

As explained in the Introduction, our goal was to use a weight-elicitation technique that is more methodologically robust than relying on evaluation (e.g. Likert) scales while also being easy-to-use, cost-effective and reproducible. After considering the wide range of techniques available [23], often supported by specialized software [27,28], we decided to use the PAPRIKA method [29] implemented by 1000Minds software [30].[1]

As explained below, the PAPRIKA method – a partial acronym for '**P**otentially **A**ll **P**airwise **R**an**K**ings of all possible **A**lternatives' – is based on pairwise comparisons, and so it is less cognitively burdensome for decision-makers than other methods. Another advantage is that it yields a set of weights for each participant, in contrast to other methods which produce aggregated data only, thereby permitting us to compare weights across participant subgroups. Earlier applications of PAPRIKA and 1000Minds in the area of health technology prioritisation include [31,32]; other health applications include prioritising patients for elective surgery [33,34], disease classification [35–37] and measuring clinical trial outcomes [38–40].

The PAPRIKA method begins by identifying (performed by the software) all pairs of, in the present context, hypothetical medical devices defined on two criteria at-a-time and involving a trade-off. Each participating decision-maker is repeatedly presented with pairs of devices in random order and asked to choose which device she prefers (has greater priority). An example of a pairwise-ranking question appears in Fig. 1. Each time the decision-maker ranks a pair of devices, all other hypothetical devices that can be pairwise ranked via transitivity are identified and eliminated; for example, if a decision-maker prefers device X over Y and then she prefers Y over Z, then – by transitivity – X is also prioritised over Z (and so the method would not ask a question relating to this third pair of devices).

[1] Co-invented by the second author (PH), the method and software is freely available for academic and non-commercial use from him or via www.1000minds.com.

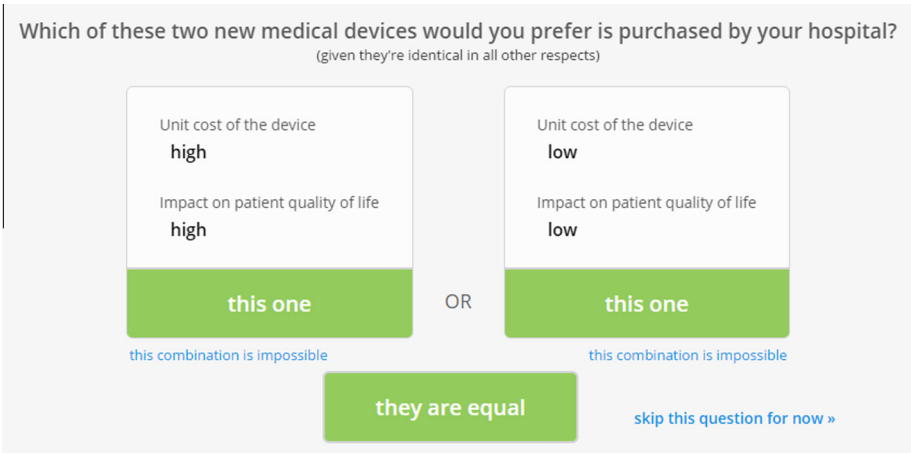

Fig. 1. Example of a pairwise-ranking question (a screenshot from 1000Minds software).

The elimination procedure explained above ensures that the number of questions a decision-maker is asked is minimized (as reported in the next section, on average, 18 questions per person); and yet the decision-maker ends up having pairwise ranked all hypothetical devices differentiated on two criteria at-a-time, either explicitly or implicitly (by transitivity). Finally, from the decision-maker's explicit pairwise rankings, the software uses mathematical methods based on linear programming to derive weights for the criteria (for technical details, see [29]). These weights are reported at the individual decision-maker level and also averaged across the group of participants.

The overall scores for risk and value respectively are calculated using Eqs. (1) and (2) below. Thus, for each device j, multiplying the mean weights W_i for each criterion i (normalized across the 10 criteria so the weights sum to one) by the performance scores for the risk- and value-based criteria respectively, $R_{i,j}$ and $V_{i,j}$, gives overall scores for risk and value, S_{Rj} and S_{Vj}.

$$S_{Rj} = \sum_{i=1}^{5} W_i \times R_{ij} \quad (1)$$

$$S_{Vj} = \sum_{i=1}^{5} W_i \times V_{ij} \quad (2)$$

2.3. Applying the IDA tool

To test the usefulness of the IDA tool, we worked with the medical devices committee *Comité des dispositifs médicaux stériles* (CODIMS) of the association *Assistance Publique-Hôpitaux de Paris* (AP-HP), which represents all 38 university hospitals in Paris and is the largest health care provider in France, to apply the tool. CODIMS is responsible for reviewing the evidence to make recommendations on technologies' proper and for deciding which medical devices should be adopted by the AP-HP network [41].

CODIMS's scientific secretariat is responsible for performing HTAs, which are based on local data and involve comparisons of new devices vis-a-vis devices currently available in the AP-HP network. Based on this evidence, CODIMS's committee of representatives, comprising 25 permanent members (14 medical doctors and 11 hospital pharmacists) and also other invited members, makes recommendations about the adoption of new devices by the network.

We applied the IDA tool in an evaluation of two new drug-eluting beads (DEBs) for transcatheter arterial chemoembolization, a treatment for hepatocellular carcinoma. Although DEBs are already funded in the AP-HP network, interventional radiologists had expressed interest in two new DEBs that recently became available. They requested a HTA from CODIMS, asking in particular whether the purchase of these new DEBs could extend the clinical indications for using DEBs. A systematic review of the literature performed by the scientific secretariat did not find any clinical studies directly comparing the DEBs.

The 25 permanent members of CODIMS were invited to participate in an online pairwise-ranking survey administered by the 1000Minds software discussed earlier. As well as generating mean criterion weights for the IDA tool, weights at the individual-participant level enabled us to compare weights from medical doctors versus hospital pharmacists (as mentioned earlier, both groups are members of CODIMS). This comparison was performed via a Mann-Whitney U test, using R software, version 2.14.1 [42].

Based on the scientific secretariat's HTA report on the two new DEBs, two CODIMS hospital pharmacists – the fourth and fifth authors (AB and ALC) – and an independent researcher – the first author (NM) – independently rated the two new DEBs, DEB_A and DEB_B, on the 10 decision criteria relative to DEBs currently in use, DEB_0. For DEB_A and DEB_B respectively, a score for each criterion was assigned, which could take three possible values: "1" for a higher performance (i.e. relative to DEB_0), "0" for the same performance or it is not possible to reach a conclusion, or "−1" for a lower performance. If the three scorers were not unanimous in their ratings, the disagreement was discussed until a consensus score was agreed to.

Finally, applying Eqs. (1) and (2) above, the consensus scores for DEB_A and DEB_B were multiplied by the corresponding criterion weights and summed to produce overall scores for the risk and value categories respectively. These overall values were plotted on a graph with "value" on the horizontal axis and "risk" on the vertical axis using Microsoft Office Excel 2010 [43]. To illustrate the degree of uncertainty of the scoring process and to provide a visual sensitivity analysis for the CODIMS committee of representatives, this graph also included overall scores arising from the three individual scorers.

3. Results

3.1. Decision criteria

Upon completion of the two-step selection procedure, 26 criteria were initially identified, as reported in Table 1. Ensuring as much as possible that the criteria are complete, non-redundant, operational and mutually independent (see Table 1 again) necessitated the culling of more than half of these initial criteria, leaving

Table 1
26 evaluation criteria identified in the literature review and survey of French university hospitals.

Perspective	Evaluation criteria identified (code)	Definition	Mutual independence	Operationality	Decision criteria selected (code)
Economy	ECOREVIEW	Review of economic evaluation studies	✓		
	UNITCOST	Unit cost of the device	✓	✓	UNITCOST
	COST PER PATIENT	Cost per patient excluding the unit cost of the device	✓	✓	COST PER PATIENT
	BUDGETIMP	Budgetary impact of the adoption of the device		✓	
	DRGTARIFF	DRG tariff including device cost	✓	✓	DRGTARIFF
	ADDPAYMENT	Additional payment for the device	✓	✓	ADDPAYMENT
Organization	PATIENTNB	Number of patients treated with the device per year	✓	✓	PATIENTNB
	EXPERTISE	Need for specific expertise in the users	✓	✓	EXPERTISE
	TRAINING	Need for user training	✓	✓	TRAINING
	ADDMD	Need for additional medical devices to use with the new device	✓		
	WORKFLOW	Impact on work flow within the hospital	✓	✓	WORKFLOW
Patient	CONFLINT	Conflict of interest	✓		
	QUALITYOFLIFE	Impact on patient quality of life	✓	✓	QUALITYOFLIFE
Technology	CURRENTHT	Description of the currently available technology	✓		
	REVIEW	Review of clinical evaluation studies	✓		
	CLINICAL BENEFIT	Expected health gains for the patient	✓	✓	CLINICAL BENEFIT
	DESCRIPTION	Description of the new device	✓		
	OTHERHOSP	Adoption of the device at other hospitals	✓		
	PREVIOUS USE	Previous uses of the device within the hospital (e.g. clinical trials)	✓		
	SAFETY	Expected risks and/or adverse events related to the device	✓	✓	SAFETY
	INDICATION	Indications for the device	✓		
	INNOVATIVENESS	Novelty of the device	✓		
	EVIDENCE	Quality of evidence on the device	✓	✓	EVIDENCE
	OPINION	Opinions from a HTA agency or scientific society about the device	✓	✓	OPINION
	STUDY	Ongoing studies on the device	✓		
	TYPE	Type of device (diagnosis, treatment…)	✓		

DRG: diagnosis-related group; HTA: health technology assessment

Table 2
Decision criteria included in the IDA tool.

Decision criterion	Definition
Risk-based criteria	
TRAINEXP	Need for specific expertise and/or training of users
WORKPAT	Number of patients and impact on work flow within the hospital
COST PER PATIENT	Cost per patient excluding the unit cost of the device
COVERAGE	DRG tariff or additional payments for the device
UNIT COST	Unit cost of the device
Value-based criteria	
SAFETY	Expected risks and/or adverse events related to the device
CLINICAL BENEFIT	Expected health gains for the patient
QUALITY OF LIFE	Impact on patient quality of life
OPINION	Opinion of a HTA agency or scientific society concerning the device
EVIDENCE	Quality of the evidence concerning the device

DRG: diagnosis-related group; HTA: health technology assessment.

12 criteria: seven in the risk category and five in the value category. In order to match the numbers of risk-based and value-based criteria (five each) two pairs of criteria in the risk category were combined to generate two composite criteria. The final five criteria in each category are reported in Table 2.

3.2. Criterion weights

In the four weeks before the triannual CODIMS meeting in June 2014, all 25 permanent members of CODIMS completed the online pairwise-ranking survey. Each participant was required to answer 18 pairwise-ranking questions on average, typically taking 5–10 min in total per person. The mean criterion weights are presented in Fig. 2. It can be seen that the mean weights for the value-based criteria are very similar to each other: in the 12–15% range. The mean weights for the risk-based criteria are lower: in the 4–8% range. There are no statistically-significant differences between medical doctors and hospital pharmacists with respect to their weights ($p > 0.05$).

3.3. Application of the IDA tool

The scoring scale used by the three scorers to independently rate the performance of the two new DEBs (DEB_A and DEB_B) on the 10 decision criteria relative to DEBs currently in use (DEB_0) is explained in Table 3. The scorers' ratings, a consensus rating and the overall scores on the risk- and value-based criteria respectively are reported in Table 4.

According to the IDA tool, as represented in Fig. 3, DEB_A delivers the same amount of value and slightly less risk than DEB_0, whereas DEB_B is less valuable and more risky than DEB_0. Clearly, DEB_B is dominated by the DEBs currently available, and so the committee of representatives rejected it. After deliberating, the committee rejected DEB_A too – because the committee considered it did not offer new therapeutic options and the risk reduction was negligible (relative to DEBs currently available).

4. Discussion

In the hospital context, multiple criteria must be considered simultaneously when deciding whether or not to adopt new medical devices, and the analytical framework needs to be structured and transparent about how decisions are reached; this is an advantage of MCDA over traditional decision-making processes [12]. Based on the concept of mini-HTA and applying relevant decision

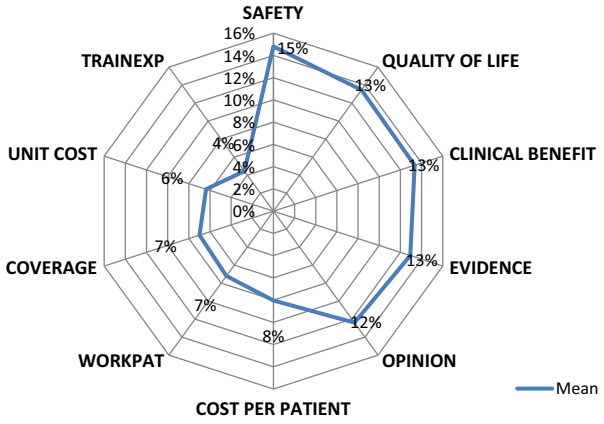

Fig. 2. Radar chart of the mean criterion weights for the 25 CODIMS members. TRAINEXP: need for specific expertise and/or training of users; WORKPAT: number of patients and impact on work flow within the hospital; COST PER PATIENT: cost per patient excluding the unit cost of the device.

criteria to medical devices for individual patient use inspired from the Matrix4value model, we developed the IDA ('innovative device assessment') decision-support tool. This combination of MCDA and HTA seems to result in fairness of the priority-setting process, an essential condition for an ethical HTA framework: accountability for reasonableness [17,24,44].

We used the PAPRIKA method implemented by 1000Minds software to elicit decision-makers' preferences with respect to weights on the criteria, representing their relative importance. This method proved easy to implement and was judged to be highly satisfactory by all participants. PAPRIKA's defining characteristic is that it is based on answering pairwise-ranking questions (see Fig. 1 again), which has the advantage that pairwise ranking – choosing one alternative (here, medical devices) from two – is a natural type of decision activity that everyone has experience of in their daily lives. As Drummond *et al.* [45] argue in support of such methods in general: "The advantage of choice-based methods is that choosing ... is a natural human task at which we all have considerable experience, and furthermore it is observable and verifiable."

In contrast, most other weight-elicitation techniques, of which a well-known example is the analytical hierarchy process (AHP) [46], are based on 'scaling' or 'ratio' measurements of decision-makers' preferences. We preferred PAPRIKA over AHP because AHP is based on asking decision-makers to express ratio-scale measurements of their preferences – e.g. "On a 1–9 scale, how much more or less is one criterion (or alternative) preferred to another?" – which is cognitively demanding relative to PAPRIKA's simple ordinal (ranking) measurements, as discussed above.

A potential limitation of techniques based on pairwise ranking, such as PAPRIKA, is the relatively large number of comparisons required which may be excessively burdensome for decision-makers. However, in this study no participants made any comments concerning the time required to complete the pairwise-ranking survey (typically 5–10 min per person and involving 18 pairwise-ranking questions on average). On the contrary, the feedback obtained was very positive with most participants finding the survey to be user-friendly and simple, as in other applications of 1000Minds software [47].

The EVIDEM team recently suggested that weight-elicitation techniques based on pairwise comparisons (as for PAPRIKA) are likely to offer better discrimination with respect to the weight estimates produced than other methods [47]. Indeed, we observed that the participants in our pairwise-ranking survey gave more weight to value-based criteria than to risk-based ones. The Matrix4value study, whose weight-elicitation technique was based on Likert scales, reported similar findings, but with more homogeneous results both within and between the two over-arching value and risk categories [5]. We speculate that Matrix4value's poor discrimination was due to the inferior weight-elicitation technique used.

Table 3
Comparative scoring scale used in the IDA tool.

Decision criterion	Score		
	−1	0	1
Risk-based criteria			
TRAINEXP	Less expertise or training required than for the device currently available	The same performance, or not possible to draw a conclusion	More expertise or training required than for the device currently available
WORKPAT	Lower impact on the number of patients and on work flows than for the device currently available		Higher impact on the number of patients and on work flow than for the device currently available
COST PER PATIENT[a]	Cost per patient treated lower than for the device currently available		Cost per patient treated higher than for the device currently available
COVERAGE	Better coverage than for the device currently available		Poorer coverage than for the device currently available
UNITCOST	Unit cost of the device lower than for the device currently available		Unit cost of the device higher than for the device currently available
Value-based criteria			
SAFETY	Less secure than the device currently available	The same performance, or not possible to draw a conclusion	More secure than the device currently available
CLINICAL BENEFIT	Lower clinical benefits than for the device currently available		Greater clinical benefits than for the device currently available
QUALITY OF LIFE	Lower impact on patient quality of life than the device currently available		Greater impact on patient quality of life than the device currently available
OPINION	Negative opinion about the device from a HTA agency or scientific society		Positive opinion about the device from a HTA agency or scientific society
EVIDENCE	Less high-quality evidence than for the device currently available		More high-quality evidence than for the device currently available

[a] Cost per patient excluding the unit cost of the device.

Table 4
Scores for the two new drug-eluting beads.

Decision criterion	Scorer 1		Scorer 2		Scorer 3		Consensus score	
	DEB_A	DEB_B	DEB_A	DEB_B	DEB_A	DEB_B	DEB_A	DEB_B
Risk-based criteria								
TRAINEXP	1 × 0.04	1 × 0.04	1 × 0.04	1 × 0.04	0	0	1 × 0.04	1 × 0.04
WORKPAT	0	0	0	0	0	0	0	0
COST PER PATIENT	0	0	0	0	0	0	0	0
COVERAGE	0	0	0	0	0	0	0	0
UNITCOST	−1 × 0.06	1 × 0.06	−1 × 0.06	−1 × 0.06	1 × 0.06	1 × 0.06	−1 × 0.06	1 × 0.06
Overall score: Risk	−0.02	0.10	−0.02	−0.02	0.06	0.06	−0.02	0.10
Value-based criteria								
SAFETY	0	0	0	0	0	0	0	0
CLINICAL BENEFIT	0	0	0	0	0	0	0	0
QUALITY OF LIFE	0	0	0	0	0	0	0	0
OPINION	0	0	0	0	0	0	0	0
EVIDENCE	0	−1 × 0.13	0	−1 × 0.13	−1 × 0.13	−1 × 0.13	0	−1 × 0.13
Overall score: Value	0	−0.13	0	−0.13	−0.13	−0.13	0	−0.13

DEB_A: drug-eluting beads A; DEB_B: drug-eluting beads B; TRAINEXP: need for specific expertise and/or training of users; WORKPAT: number of patients and impact on work flow within the hospital; COST PER PATIENT: cost per patient excluding the unit cost of the device.
Criterion weights are shown in parentheses. Scores are assigned relative to the drug eluting-beads currently available in the hospital network: "1" for higher performance, "0" for the same performance, or it is not possible to draw a conclusion, "−1" for lower performance. The scores are multiplied by the corresponding criterion weights and summed to yield an overall score.

The PAPRIKA method may have enabled a more accurate representation of the criteria's relative importance.

In addition, the good discrimination of the PAPRIKA method made it possible for us to compare weights between participating hospital pharmacists and medical doctors, a potentially important property in terms of the sustainability of the IDA tool. No statistically significant differences were observed, but this situation may change over time as CODIMS's members are replaced, and monitoring of this aspect is therefore required.

The results obtained with the IDA tool, revealing that the two new DEBs confer no additional value relative to DEBs currently available, greatly assisted the CODIMS committee in its deliberations. Careful identification of an appropriate comparator (i.e. the device currently available) highlights the importance of defining the scope of the HTA [2]. The motivation for the present HTA was interventional radiologists' interest in the potential added value of the new DEBs being considered. Instead of asking whether the new DEBs could be used to replace the device currently available, they focused on whether the new DEBs offered new therapeutic possibilities. Before using the IDA tool, decision-makers should clearly identify the over-arching question, including the counterfactual, to be addressed.

Like all decision-support tools combining MCDA and HTA, the results obtained depend on the quality of the underlying HTA reports. The data evaluated must be reliable and comprehensive in order for the new devices under consideration to be accurately scored on the decision criteria without introducing judgment biases. This is the main reason why we requested CODIMS's involvement in this study; CODIMS has a lot of expertise in hospital-based HTA, ensuring that the IDA tool could be tested under proper conditions [41].

Our findings highlight some of the prerequisites for the correct implementation of the IDA tool in other French university hospitals. Indeed, the IDA tool cannot be implemented alone because it is intended to complement a well-conducted HTA process by ensuring all important criteria are considered by decision-makers in a more transparent manner. Consequently, the methodology used for collecting and assessing evidence for medical devices must respect good HTA practice to ensure that the evidence submitted to decision-makers is of high quality. Nonetheless, some criteria remained difficult to score despite rigorous adherence to HTA guidelines. For example, quantifying the learning curve for the new medical device is important when considering organizational impacts, such as expertise and training [48]; unfortunately,

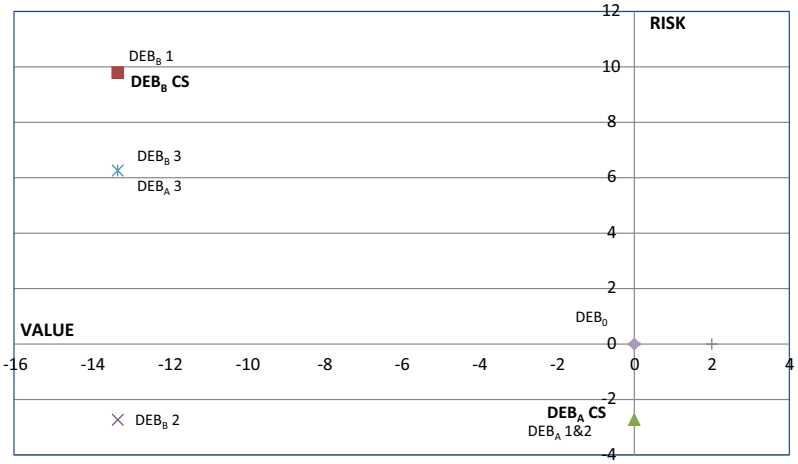

Fig. 3. Comparison of the three drug-eluting beads based on the IDA tool results. DEB_A: drug-eluting beads A; DEB_B: drug-eluting beads B; DEB_0: drug-eluting beads currently available. "1", "2", "3"represent the three scorers and "CS" the consensus score.

however, the data available for quantifying learning curves are often very poor [49]. In general, high quality clinical data for medical devices are rarely available due to methodological difficulties associated with designing trials (mainly involving randomization and blinding). Consequently, studies available for recently marketed devices are mostly case series and not often large comparative studies or randomized controlled studies.

This study has several limitations worthy of discussion. First, selection of the criteria was a relatively long and difficult process due to the considerable variability of mini-HTA terminology. Two researchers independently classified and coded the criteria, but this process was limited by their subjective interpretations. Another limitation concerned the testing of the IDA tool with CODIMS of AP-HP (*Assistance Publique-Hôpitaux de Paris*), which is not necessarily representative of all medical device committees across France. We intend to carry out further tests in other university hospitals to explore the tool's strengths and weaknesses. As already pointed out, the methodology used for collecting evidence is critical to ensuring the correct scoring of the criteria. Nonetheless, checklists or scales are not always available or appropriate for capturing the evidence for all criteria, and this is especially true for medical devices. Indeed, we are often forced to use checklists originally designed for medicines that do not closely match the specifications of medical devices. Finally, the scoring was performed by three scorers, but we believe that this step could be improved by involving the whole committee in the process, which would also improve CODIMS members' understanding of the process.

In conclusion, research on combinations of MCDA and hospital-based HTA is in its infancy. Matrix4value opened up new possibilities in this area, and we hope that this case study applying the IDA tool helps to demonstrate the value of MCDA as a complement to HTA at the hospital level. Before MCDA can be more widely used, further studies would be useful to adapt and develop HTA in the hospital context, as in the AdHopHTA project, for example [3].

Potential competing interests

The second author (PH), together with Franz Ombler, invented the 1000Minds software and PAPRIKA method referred to in the article. Via their company 1000Minds Ltd, PH and FO earn income from software licenses; they also make the software available for free to unfunded academic users (to-date more than 1000 researchers and students worldwide), including the present study.

Acknowledgments

Thank you to the 25 CODIMS members for their participation.

References

[1] M.V. Garrido, F.B. Kristensen, C.P. Nielsen, Health Technology Assessment and Health Policy-Making in Europe: Current Status, Challenges and Potential, WHO Regional Office Europe, 2008 (198 p).
[2] M.F. Drummond, J.S. Schwartz, B. Jönsson, B.R. Luce, P.J. Neumann, U. Siebert, et al., Key principles for the improved conduct of health technology assessments for resource allocation decisions, Int. J. Technol. Assess. Health Care 24 (3) (2008) 244–258.
[3] M.-P. Gagnon, Hospital-based health technology assessment: developments to date, Pharmacoeconomics 32 (9) (2014) 819–824.
[4] K. Kidholm, L. Ehlers, L. Korsbek, R. Kjaerby, M. Beck, Assessment of the quality of mini-HTA, Int. J. Technol. Assess. Health Care 25 (1) (2009 Jan) 42–48.
[5] L. Sampietro-Colom, I. Morilla-Bachs, S. Gutierrez-Moreno, P. Gallo, Development and test of a decision support tool for hospital health technology assessment, Int. J. Technol. Assess. Health Care 28 (4) (2012 Oct) 460–465.
[6] M. McGregor, J.M. Brophy, End-user involvement in health technology assessment (HTA) development: a way to increase impact, Int. J. Technol. Assess. Health Care 21 (2) (2005) 263–267.
[7] L. Ehlers, M. Vestergaard, K. Kidholm, B. Bonnevie, P.H. Pedersen, T. Jørgensen, et al., Doing mini-health technology assessments in hospitals: a new concept of decision support in health care?, Int J. Technol. Assess. Health Care 22 (3) (2006) 295–301.
[8] P. Thokala, A. Duenas, Multiple criteria decision analysis for health technology assessment, Value Health 15 (8) (2012 Dec) 1172–1181.
[9] N. Devlin, J. Sussex, Incorporating multiple criteria in HTA, Methods and Processes London: Office of Health Economics [Internet], 2011. <http://healthpolicy.fsi.stanford.edu/sites/default/files/ohe_hta_methods.pdf> (cited 2015 Mar 11).
[10] M.M. Goetghebeur, M. Wagner, H. Khoury, D. Rindress, J.-P. Grégoire, C. Deal, Combining multicriteria decision analysis, ethics and health technology assessment: applying the EVIDEM decisionmaking framework to growth hormone for Turner syndrome patients, Cost Eff. Resour. Alloc. 8 (1) (2010) 4.
[11] S. Gurtner, Making the right decisions about new technologies: a perspective on criteria and preferences in hospitals, Health Care Manage. Rev. 39 (3) (2014 Sep) 245–254.
[12] V. Diaby, R. Goeree, How to use multi-criteria decision analysis methods for reimbursement decision-making in healthcare: a step-by-step guide, Exp. Rev. Pharmacoeconom. Outcomes Res. 14 (1) (2014) 81–99.
[13] I. Ivlev, P. Kneppo, M. Bartak, Multicriteria decision analysis: a multifaceted approach to medical equipment management, Technol. Econ. Develop. Econ.. 20 (3) (2014) 576–589.
[14] T. Sullivan, P. Hansen, Determining Benefits-related Criteria and Weights for Prioritising Health Technologies. Centre for Health Systems, University of Otago, Dunedin, 2014.
[15] M. Ryan, K. Gerard, Using discrete choice experiments to value health care programmes: current practice and future research reflections, Appl. Health Econ. Health Policy 2 (1) (2003) 55–64.
[16] E. Lancsar, J. Louviere, Conducting discrete choice experiments to inform healthcare decision making, Pharmacoeconomics 26 (8) (2008) 661–677.
[17] M.M. Goetghebeur, M. Wagner, H. Khoury, R.J. Levitt, L.J. Erickson, D. Rindress, Bridging health technology assessment (HTA) and efficient health care decision making with multicriteria decision analysis (MCDA): applying the EVIDEM framework to medicines appraisal, Med. Decis. Making 32 (2) (2012) 376–388.
[18] G. Radaelli, E. Lettieri, C. Masella, L. Merlino, A. Strada, M. Tringali, Implementation of EUnetHTA core model® in Lombardia: the VTS framework, Int. J. Technol. Assess. Health Care 22 (2014 Jan) 1–8.
[19] M. Kaltoft, G. Kirketerp, J. Dowie, Literature Review on the Use of Handhelds in Acute Care: Towards a Patient-centred Health Technology Assessment (HTA) [Internet]. <http://www.academia.edu/1668470/Literature_review_on_the_use_of_handhelds_in_acute_care_towards_a_patient-centred_Health_Technology_Assessment_HTA_> (cited 2015 Aug 14).
[20] N. Martelli, M. Billaux, I. Borget, J. Pineau, P. Prognon, H. van den Brink, Introduction of innovative medical devices at French university hospitals: an overview of hospital-based health technology assessment initiatives, Int. J. Technol. Assess. Health Care 31 (1–2) (2015) 12–18.
[21] N. Martelli, A.-S. Lelong, P. Prognon, J. Pineau, Hospital-based health technology assessment for innovative medical devices in university hospitals and the role of hospital pharmacists: learning from international experience, Int. J. Technol. Assess. Health Care 29 (2) (2013 Apr) 185–191.
[22] M. Pöyhönen, R.P. Hämäläinen, On the convergence of multiattribute weighting methods, Eur. J. Oper. Res. 129 (3) (2001) 569–585.
[23] V. Belton, T. Stewart, Multiple Criteria Decision Analysis: An Integrated Approach, Kluwer Academic Publishers, 2002.
[24] R. Baltussen, L. Niessen, Priority setting of health interventions: the need for multi-criteria decision analysis, Cost Eff. Resour. Alloc. 21 (4) (2006) 14.
[25] L.A. Guindo, M. Wagner, R. Baltussen, D. Rindress, J. van Til, P. Kind, et al., From efficacy to equity: literature review of decision criteria for resource allocation and healthcare decisionmaking, Cost Eff. Resour. Alloc. 10 (1) (2012) 9.
[26] National Economic Research Associates: Multi-Criteria Analysis Manual for Making Government Policy [Internet], 2000. <https://www.gov.uk/government/publications/multi-criteria-analysis-manual-for-making-government-policy> (cited 2015 Aug 13).
[27] P. McGinley, Decision Analysis Software Survey. OR/MS Today 39. OR/MS Today 39 2012 January. <http://www.orms-today.org/surveys/das/das.html>.
[28] H. Weistroffer, Y. Li, Multiple criteria decision analysis software, in: M. Ehrgott, J. Figueira, S. Greco (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys, second ed., Springer, 2014.
[29] P. Hansen, F. Ombler, A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives, J. Multi-Crit. Decis. Anal. 15 (3–4) (2008) 87–107.
[30] F. Ombler, P. Hansen, 1000Minds Software, 2015. <http://www.1000minds.com>.
[31] O. Golan, P. Hansen, G. Kaplan, O. Tal, Health technology prioritization: which criteria for prioritizing new technologies and what are their relative weights?, Health Policy 102 (2) (2011 Oct) 126–135
[32] O. Golan, P. Hansen, Which health technologies should be funded? A prioritization framework based explicitly on value for money, Isr. J. Health Policy Res. 1 (1) (2012) 44.
[33] P. Hansen, A. Hendry, R. Naden, F. Ombler, R. Stewart, A new process for creating points systems for prioritising patients for elective health services, Clin. Gov. Int. J. 17 (3) (2012) 200–209.
[34] A. Fitzgerald, C. de Coster, S. McMillan, R. Naden, F. Armstrong, A. Barber, et al., Relative urgency for referral from primary care to rheumatologists: the Priority Referral Score, Arthritis Care Res. (Hoboken) 63 (2) (2011) 231–239.

[35] D. Aletaha, T. Neogi, A.J. Silman, J. Funovits, D.T. Felson, C.O. Bingham, et al., 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative, Arthritis Rheum. 62 (9) (2010) 2569–2581.
[36] F. van den Hoogen, D. Khanna, J. Fransen, S.R. Johnson, M. Baron, A. Tyndall, et al., 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative, Arthritis Rheum. 65 (11) (2013 Nov) 2737–2747.
[37] S.R. Johnson, R.P. Naden, J. Fransen, F. van den Hoogen, J.E. Pope, M. Baron, et al., Multicriteria decision analysis methods with 1000Minds for developing systemic sclerosis classification criteria, J. Clin. Epidemiol. 67 (6) (2014) 706–714.
[38] W.J. Taylor, J.A. Singh, K.G. Saag, N. Dalbeth, P.A. MacDonald, N.L. Edwards, et al., Bringing it all together: a novel approach to the development of response criteria for chronic gout clinical trials, J. Rheumatol. 38 (7) (2011) 1467–1470.
[39] W.J. Taylor, M. Brown, O. Aati, M. Weatherall, N. Dalbeth, Do patient preferences for core outcome domains for chronic gout studies support the validity of composite response criteria?, Arthritis Care Res (Hoboken) 65 (8) (2013) 1259–1264.
[40] F. Dobson, R.S. Hinman, E.M. Roos, J.H. Abbott, P. Stratford, A.M. Davis, et al., OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis, Osteoarthr. Cartil. 21 (8) (2013) 1042–1052.
[41] H. Beaussier, H. Junot, S. Lancrenon, P. Faure, New medical device hospital assessment: what kind of clinical data?, Ann Pharm. Fr. 70 (1) (2012) 35–45.
[42] R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2011. <http://www.R-project.org>.
[43] Microsoft, Microsoft Excel, Redmond, Washington: Microsoft, 2010, Computer Software. <https://products.office.com/fr-fr/excel>.
[44] N. Daniels, Accountability for reasonableness: Establishing a fair process for priority setting is easier than agreeing on principles, BMJ 321 (7272) (2000) 1300.
[45] M.F. Drummond, M.J. Sculpher, G.W. Torrance, B.J. O'Brien, G.L. Stoddart, Methods for the Economic Evaluation of Health Care Programmes, third ed., OUP Oxford, Oxford, 2005.
[46] M. Ritrovato, F.C. Faggiano, G. Tedesco, P. Derrico, Decision-oriented health technology assessment: one step forward in supporting the decision-making process in hospitals, Value Health 18 (4) (2015 Jun) 505–511.
[47] J. van Til, C. Groothuis-Oudshoorn, M. Lieferink, J. Dolan, M. Goetghebeur, Does technique matter; a pilot study exploring weighting techniques for a multi-criteria decision support framework, Cost Eff. Resour. Allocation 12 (1) (2014) 22.
[48] C.R. Ramsay, S.A. Wallace, P.H. Garthwaite, A.F. Monk, I.T. Russell, A.M. Grant, Assessing the learning curve effect in health technologies. Lessons from the nonclinical literature, Int. J. Technol. Assess. Health Care 18 (1) (2002) 1–10.
[49] E. Lettieri, C. Masella, Priority setting for technology adoption at a hospital level: relevant issues from the literature, Health Policy 90 (1) (2009 Apr) 81–88.