JOURNAL OF MULTI-CRITERIA DECISION ANALYSIS
J. Multi-Crit. Decis. Anal. **23**, 15–27 (2016)
Published online 27 August 2015 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/mcda.1547

Development of a Robust Multicriteria Classification Model for Monitoring the Postoperative Behaviour of Heart Patients

MICHAEL DOUMPOS[a*], PANAGIOTIS XIDONAS[b], SOTIRIOS XIDONAS[c] and YANNIS SISKOS[d]
[a] *Financial Engineering Laboratory, School of Production Engineering and Management, Technical University of Crete, Chania 73100, Greece*
[b] *ESSCA, École de Management, 55 quai Alphonse Le Gallo, Paris 18534, France*
[c] *Second Department of Cardiology, Division of Cardiac Electrophysiology, Evaggelismos General Hospital, Athens Greece*
[d] *Department of Informatics, University of Piraeus, 80, M. Karaoli & A. Dimitriou St., Piraeus 18534, Greece*

ABSTRACT

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia occurring in 2% of the general population, while the assuming projected incidence in 2050 will rise to 4.3%. This paper presents a multicriteria methodology for the development of a model for monitoring the post-operative behaviour of patients who have received treatment for AF. The model classifies the patients in seven categories according to their relapse risk, on the basis of seven criteria related to the AF type and pathology conditions, the treatment received by the patients and their medical history. The analysis is based on an extension of the UTilités Additives DIScriminantes (UTADIS) method, through the introduction of a two-stage model development procedure that minimizes the number and the magnitude of the misclassifications. The analysis is based on a sample of 116 patients who had pulmonary veins isolation in a Greek public hospital. The classification accuracy of the best fitted models scores between 71% and 84%. Copyright © 2015 John Wiley & Sons, Ltd.

KEY WORDS: multiple criteria analysis; health care; disaggregation analysis; multicriteria classification

1. INTRODUCTION

Atrial fibrillation (AF) is the most common arrhythmia and can be either symptomatic or not. Its prevalence increases with age, and it appears that one in four adults older than 40 years has a lifetime risk of developing AF of approximately 25%. The major mechanism that initiates and perpetuates AF relies on rapid electrical discharges from the pulmonary veins (PV) that return oxygenated blood from the lungs to the left atrium (LA) of the heart. Electrical isolation of the PV with the application of high frequency current across the ostia of the PV is particularly effective for elimination of AF and is widely used in cardiac electrophysiology departments of tertiary hospitals.

However, even after PV isolation (PVI), AF often recurs. Recurrence is classified as early when it takes place 48 h after the operation, late when it occurs within 30 days and very late for cases more than 30 days after the operation. The efficacy of PVI depends on several medical variables, and the assessment of the AF recurrence risk is of major importance in order to decide the most suitable treatment for a patient. Analytic decision models can be particularly useful for defining post-operative AF treatment.

Empirical evidence has shown that medical decision support systems often improve signi ficantly the medical decision process (Garg *et al*., 2005; Kawamoto *et al*., 2005), in different ac tivities (e.g. diagnosis, therapy, monitoring and prevention) and contexts such as acute care, primary care and patient advice (Ammenwerth *et al*., 2013). Data mining, computational intel ligence and statistical pattern recognition techniques have been widely used for diagnostic purposes (for an overview, see the work of Hardin and Chhieng,

*Correspondence to: Michael Doumpos, Financial Engineering Laboratory, School of Production Engineering and Management, Technical University of Crete, Chania 73100, Greece. E-mail: mdoumpos@dpem.tuc.gr

Received 19 February 2015
Accepted 25 July 2015

2007). Such methods have also been used in predicting and detecting AF and other forms of cardiac arrhythmia (Alonso-Atienza *et al*., 2012; Chesnokov, 2008; Mohebbi and Ghassemian, 2012), mostly using complex machine learning/data mining models that emphasize the accuracy of the results rather than their inter pretability. However, as noted by Berner and La Lande (2007), many physicians are hesitant to use such systems because the reasoning behind them is not transparent and they are not built on the grounds of knowledge derived from the medical literature, or rules and guidelines issued by clinical associations based on clinical trials' results, registries and experts' consensus (see for instance, the work of Camm *et al*., 2012).

Multicriteria decision aid (MCDA) is well suited in this context, providing a constructive approach for developing medical support systems that combine the physicians' expert judgments with evidence-based clinical practice, in a patient-centred clinical decision-making context (Dolan, 2010). Medical applications of MCDA methods cover, among others, generic computer-aided diagnostic systems (Du Bois *et al*., 1989; Rahimi *et al*., 2007), specialized diagnostic and screening models (Belacel, 2000; Dolan and Frisina, 2002; Goletsis *et al*., 2004), decision aiding in evidence-based medicine (O'Sullivan *et al*., 2014; van Valkenhoef *et al*., 2013), medication risk analysis and appraisal (Goetghebeur *et al*., 2012; Tervonen *et al*., 2011), therapy planning (Hamacher and Küfer, 2002; Schlaefer *et al*., 2013) and and the setting of medical practice guidelines and policy interventions (Angelucci *et al*., 2008; Baltussen *et al*., 2010; Postmus *et al*., 2014).

In this paper, we present a novel MCDA approach for the construction of a decision model that supports the analysis of AF recurrence risk. The model can be used both preoperatively and post-operatively to assess possible options for performing the PVI operation, consult with patients regarding the operation and assess the status of patients after the operation in order to prescribe a proper post-operative pharmacological therapy when needed. The model provides estimates on the AF recurrence risk as well as insights into the factors that contribute to AF recurrence. These factors relate both to the characteristics of patients and the way the PVI operation are performed. The model is expressed in the form of an additive value function, which allows the modelling of nonlinear relationships between the considered factors and the AF recurrence risk, while retaining the interpretability of simpler linear models. The additive form of the model provides both overall risk estimates and the marginal effects due to each separate factor.

The analysis is based on a sample of 116 patients who have undergone PVI operation in a major Greek hospital and have been classified into seven recurrence risk categories according to their post-operative condition. The model is developed through a multicriteria classification approach in the context of disaggregation analysis (Jacquet-Lagrèze and Siskos, 2001) on the basis of the available data. Given the multi-category nature of the problem, a new mixed-integer programming formulation is introduced that takes into account not only the number of misclassifications but also their magnitude. These two model fitting criteria are handled through a lexicographic process, and the robustness of the model is also analysed. The results demonstrate that the proposed MCDA modelling approach can provide not only a useful medical decision aid model but also guidelines and insights into the role of the AF recurrence risk assessment criteria.

The rest of the paper is organized as follows. Section 2 describes the problem context regarding the assessment of AF recurrence risk and the prognostic attributes used in the modelling process. Section 3 is devoted to the proposed multicriteria methodology for constructing the prognostic medical decision model, whereas section 4 presents the application of the methodology and discusses the obtained results. Finally, Section 5 concludes the paper and proposes some future research directions.

2. PROBLEM SETTING

Atrial fibrillation is the most common sustained cardiac arrhythmia occurring in 2% of the general population, while the projected incidence in 2050 will rise to 4.3%. The prevalence of AF increases with age, from $<0.5\%$ at 40–50 years, to 5%–15% at 80 years (Kirchhof *et al*., 2012; Wann *et al*., 2011). Men are more often affected than women. The lifetime risk of developing AF is approximately 25% in those who have reached the age of 40 years (You *et al*., 2012).

Atrial fibrillation is characterized electrocardio graphically by low-amplitude baseline oscillations (fibrillator, f-waves that lead to chaotic and irregular atrial rhythm) and an irregular ventricular rhythm, which leads to abnormal contraction and consequently

DOI: 10.1002/mcda

inadequate emptying of both atria of the heart in every cardiac cycle (heart beat). The abnormal and occasionally slow blood flow into the atria in AF patients leads to thrombus formation, thus increasing the risk of stroke, organ ischemia and other acute medical conditions that need hospitalization and have increased mortality risk.

The most common causes for developing AF are excessive alcohol intake, myocardial infarction, pericarditis, myocarditis pulmonary embolism and hyperthyroidism. Other risk factors include congestive heart failure, aortic and mitral valve disease, left atrial enlargement, obstructive sleep apnea and advanced age. On the other hand, the effects of AF in the cardiovascular system have been well studied; it doubles the risk of mortality, triples the risk for hospitalization and increases the risk of stroke nearly five times. Overall, AF promotes heart failure, and heart failure aggravates AF to worsen patients' overall prognosis.

Atrial fibrillation that terminates spontaneously within 7 days is termed paroxysmal, and AF of more than seven continuous days is called persistent. AF persistent for more than 1 year is termed longstanding, whereas longstanding AF refractory to electrical cardioversion is called permanent.

Depending on the characteristics of AF, its treatment can be based on pharmacological rate and rhythm control strategies. Left atrial catheter ablation is another option for long-term management involving patients who remain symptomatic despite other treatments. Catheter ablation is an electrophys iological operation during which multiple endocardial lesions are created by the multiple applications of high frequency current created by an external generator through ablation catheters (Figure 1). The aim of the operation is to isolate electrically the ostia (entrances) of the PV that return oxygenated blood from the lungs into the LA with the use of fluoroscopy and an electroanatomic mapping system for the navigation of ablation catheters in the heart.

The efficacy of AF ablation (PVI) varies widely depending mainly on medical variables like the type of AF, duration of AF, duration of the last AF episode, diameter and volume of the LA, the number of applications of high frequency current and the time of fluoroscopy. It is very important for electrophysiologists to choose the right patients, that is, with certain values of medical variables related to AF prior to PVI, who are more likely to benefit from the operation and remain free from arrhythmia for as long as possible, taking also into consideration the risk of adverse events due to the operation (which is estimated to by about 1%–3%).

In this context, this study employs a multicriteria methodology to determine the risk of AF recurrence. The analysis is based on a sample of 116 patients

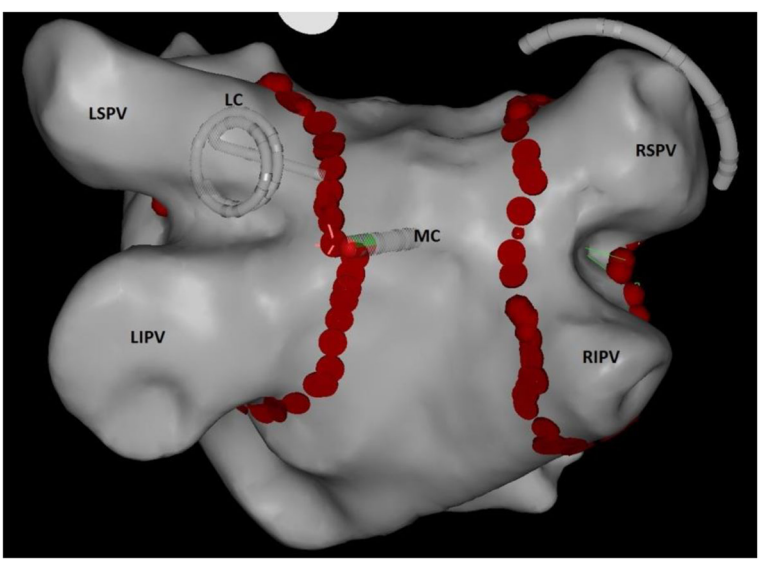

Figure 1. Postero-anterior view of left atrium (endocardial lesions—red dots—created by application of high frequency current through the irrigated tip ablation mapping catheter (MC); lasso catheter (LC) records endocardial potentials at the ostia of left superior pulmonary vein (LSPV), left inferior pulmonary vein (LIPV), right superior pulmonary vein (RSPV) and right inferior pulmonary vein (RIPV)).

DOI: 10.1002/mcda

who had PVI in a Greek public hospital. The condition of the patients was monitored after the operation, and AF recurrence was characterized as 'early' for cases in which it occurred during the first 48 h post-operatively, 'late' when in the first month and 'very late' if it occurred more than a month after the operation. Thus, the patients were classified into seven ordinal recurrence risk categories, ranging from high risk cases (class YYY), corresponding to patients for whom AF recurrence occurred at all three-time windows (early, late and very late), to patients for which PVI was successful as no recurrence occurred (class NNN).[1] The risk order of the classes was defined in cooperation with a cardiologist with experience on the treatment of AF and the PVI operation. Table I illustrates the definition of the recurrence risk categories and the number of sample patients in each class. It should be noted that early recurrence is often observed without any future complications whereas late or very late recurrence is more likely to be associated with cases that may require additional treatment. Thus, patients in category YNN are considered to be of lower risk than those in category NNY. Furthermore, patients with late recurrence are more likely to require additional treatment compared with patients with no late recurrence. This is why the top three risk categories (YYY, NYY and NYN) all correspond to patients with late recurrence.

The assessment of the AF recurrence risk is based on the seven prognostic criteria noted in the succeeding texts, which have been selected in cooperation with an expert medical decision-maker and existing medical guidelines on the risk factors of AF and its treatment (Camm *et al*., 2012; Kirchhof *et al*., 2012). In particular, the assessment criteria involve the following measures:

Table I. Definition of the atrial fibrillation recurrence risk categories

Recurrence period				
Early	Late	Very late	Class labels	No. of cases
Yes	Yes	Yes	YYY	8
No	Yes	Yes	NYY	8
No	Yes	No	NYN	9
Yes	No	Yes	YNY	2
No	No	Yes	NNY	12
Yes	No	No	YNN	3
No	No	No	NNN	74

- AF type (paroxysmal, persistent and permanent): ordinal criterion, such that a permanent type is associated with higher recurrence risk whereas a paroxysmal type is associated to lower risk.
- Duration of AF (number of years since first episode of AF): positively associated to recurrence risk (i.e. the larger the duration of AF, the higher the risk).
- Duration of the last AF episode (in days): positively associated to recurrence risk.
- LA diameter (measured in two-dimension echocardiogram in millimetre): positively associated to recurrence risk.
- LA volume (calculated automatically by echocardiograph software using longitudinal and transverse dimensions in cubic centimetre): positively associated to recurrence risk.
- Number of applications of high frequency current (each application lasts 60 s): negatively associated to recurrence risk (i.e. the risk of recurrence decreases with the number of applications of high frequency current).
- Time of fluoroscopy (duration of fluoroscopy used in order to visualize catheters and navigate them across cardiac chambers and PV in minutes): positively associated to recurrence risk.

The aforementioned assessment criteria combine attributes about the nature of the arrhythmia in each patient as well as attributes that are related to the PVI operation. The combination of such factors in an aggregate recurrence risk assessment model can support medical doctors in a number of ways, both preoperatively and post-operatively. First, it allows them to assess the risk of recurrence preoperatively based on the cardiovascular diagnostic characteristics of patients and the parameters that define how the PVI operation can be performed. In that respect, a model combining such factors can guide medical doctors to differentiate the ablation strategy during operation (i.e. increase the number of applications of high frequency current or make additional lesion lines in patients with longstanding AF and dilated LA) and provide patients with personalized estimated success rate during preoperative consultation. Furthermore, through such a model, cardiologists can change the type and duration of post-operative pharmacological therapy (i.e. more potent antiarrhythmic drugs in high risk patients).

First, it allows them to assess (post-operatively) additional treatments that may improve the condition of the patients and minimize the AF recurrence risk. Furthermore, through such a model, medical doctors

DOI: 10.1002/mcda

can analyse the trade-offs between the factors that define the nature of the PVI operation (applications of high frequency current and fluoroscopy time) while controlling for the characteristics of the AF for each patient. This allows doctors to decide on make informed decisions about the best way to perform the operation in order to minimize the recurrence risk.

3. MULTICRITERIA METHODOLOGY

In the context of the problem setting described in the previous section, the development of a decision model that facilitates the monitoring of the patients can be considered as a multicriteria classification problem. Multicriteria classification problems have received much interest among MCDA researchers over the past couple of decades, and several modelling approaches have been developed (Zopounidis and Doumpos, 2002). In this study, we employ an additive value function model. In particular, denoting by $\mathbf{x}_i=(x_{i1}, \ldots, x_{in})$ the vector with the available data for patient i on a set of n recurrence risk attributes, the patient's overall recurrence risk is assessed with the following additive function:

$$V(\mathbf{x}_i) = \sum_{j=1}^{n} w_j v_j\left(x_{ij}\right), \text{ with } \sum_{j=1}^{n} w_j = 1, \qquad (1)$$

where w_j is the (non-negative) weight for criterion j (the weights represent the trade-offs the decision-maker is willing to make among the criteria) and $v_j(\cdot)$ is the marginal value function for criterion j, normalized in [0, 1]. The additive model is well founded from a theoretical point of view (Keeney and Raiffa, 1993) and has been used in a wide range of multicriteria evaluation problems. The additive form of the model makes it easy to use and comprehend. The comprehensibility of the model is an important feature that greatly helps medical doctors to understand the model's logic, thus improving the practical usefulness of the model. More complex modelling forms (e.g. a multi-linear value function) take into account interactions between the decision criteria at the expense of yielding models, which are difficult to construct and understand.

In the modelling setting followed in this study, it is assumed that the higher the global value $V(\mathbf{x}_i)$ of patient i, the higher is his/her recurrence risk. Thus, with the additive model (1), a patient i is classified into risk group k if and only if $t_k < V(\mathbf{x}_i) < t_{k-1}$, where $t_0 > 1 > t_1 > t_2 > \cdots > t_{q-1} > t_q > 0$ is a set of thresholds that distinguish between the q recurrence risk categories $C_1, \ldots, C_q$ (e.g. $q=7$ for the sample used in this study).[2] In accordance with the aforementioned interpretation of the additive value model, the categories are risk-ordered such that C_1 corresponds to high risk patients (i.e. category YYY in Table I) and C_q to low risk ones (category NNN in Table I).

The construction of the additive model and the estimation of the separating thresholds are performed using a preference disaggregation approach (Jacquet-Lagrèze and Siskos, 2001), namely, the UTADIS II method (Doumpos and Zopounidis, 2002), which adapts the framework of the UTilités Additives (UTA) method (Jacquet-Lagrèze and Siskos, 1982) to classification problems. In this context, the evaluation model (1) is fitted on a set of data (reference set) for m patients already classified in q recurrence risk categories. The objective of the model-fitting process is to construct a decision model that is as compatible as possible with the predefined classification of the patients in the reference set. The constructed model can then be calibrated (if needed) through an interactive process with the medical decision-maker and then used to evaluate the risk for patients in a real time setting. In the UTADIS II approach, the fitting of the model is based on the solution of the following mixed-integer programme (MIP):

$$\begin{aligned}
&\min \frac{1}{q}\sum_{k=1}^{q}\frac{1}{m_k}\sum_{\forall i\in C_k}\left(\sigma_i^+ + \sigma_i^-\right)\\
&\text{s.t. } V(\mathbf{x}_i) - t_k + \sigma_i^+ \geq \delta && \forall\, i\in C_k\ (k=1,\ \ldots,\ q-1)\\
&\quad V(\mathbf{x}_i) - t_{k-1} - \sigma_i^- \leq -\delta && \forall\, i\in C_k\ (k=2,\ \ldots,\ q)\,, \qquad (2)\\
&\quad V(\mathbf{x}_*) = 0,\ V(\mathbf{x}^*) = 1\\
&\quad t_{k-1} - t_k \geq \varepsilon && k=1,\ \ldots,\ q-1\\
&\quad \sigma_i^+, \sigma_i^- \in \{0,\ 1\} && i=1,\ \ldots,\ m,
\end{aligned}$$

where m_k denotes the number of patients in the reference set from category C_k whereas σ_i^+ and σ_i^- are binary slack variables associated with patients misclassified by the additive model. In particular, σ_i^+ equals one if a patient is misclassified in a lower risk category compared to the one he/she actually belongs to (i.e. when the model underestimates the actual recurrence risk), whereas σ_i^- denotes the misclassification into higher risk classes (i.e. overestimation of risk). The first two constraints define these error variables on the basis of the threshold-based classification rule. In both constraints, δ is a small user-defined positive constant used to handle ambiguous classification results, which arise when the risk score of a patient equals one of the

DOI: 10.1002/mcda

classification thresholds (in the analysis, we set $\delta = 0.0001$). The third set of constraints normalizes the additive model in [0, 1], such that a low risk patient (denoted by $\mathbf{x}_*$) is assigned a risk score of 0, whereas a patient with the highest risk (denoted by $\mathbf{x}^*$) is assigned the maximum risk score of 1. These two extremes ($\mathbf{x}_*$ and $\mathbf{x}^*$) can either be defined through medical expertise or through the data used in the analysis. In this study, we followed the latter approach, defining $\mathbf{x}_*$ and $\mathbf{x}^*$ by the minimum and maximum levels, respectively, of the criteria described in the previous section (except for the number of applications of high frequency current, which is negatively related to recurrent risk; for this criterion, $\mathbf{x}^*$ was defined by the minimum level of the criterion and $\mathbf{x}_*$ by its maximum). Thus, a high risk patient ($\mathbf{x}^*$) has permanent AF, large durations, large LA diameter/volume, small number of applications of high frequency current during the PVI operation and large fluoroscopy time.

Finally, the fourth constraint of problem (2) defines the minimum difference between two consecutive classification thresholds, with ε being a user-defined positive constant (in this study, we used $\varepsilon = 0.02$). The objective function of problem (2) minimizes the total weighted classification error for the patients in the reference set. The weighting of the errors for each patient i from risk class C_k by $1/m_k$ imposes a balance among all risk categories, thus ensuring that the classifications of the fitted model will not be biased towards classes with a large number of patients. The aforementioned optimization problem can be formulated as a linear MIP, through the modelling of the marginal value functions of the additive model (1) as piecewise linear functions of the data (for details, see Doumpos and Zopounidis, 2002; Jacquet-Lagreze and Siskos, 1982).

Even though problem (2) is easy to solve for medium-size reference sets (with existing powerful MIP solvers), it fails to distinguish between the magnitude of the classification errors, which is an important issue in multi-category ordinal classification problems such as the one considered in this study. In such cases, instead of using the total number of misclassifications as the modelling fitting criterion, the mean absolute error is a more meaningful objective. Imposing weights to account for the imbalances in the number of patients in each risk category in the reference set, the mean-weighted absolute error (MWAE) is defined as follows:

$$\frac{1}{q}\sum_{k=1}^{q}\frac{1}{m_k}\sum_{\forall i\in C_k}|\hat{y}_i - y_i|, \tag{3}$$

where $y_i = \{1, 2, \ldots, q\}$ is the actual risk category for patient i and $\hat{y}_i$ is classification of the patient by the decision model. The construction of an additive value function model that optimizes this fitting measure for a given reference set can be performed with the following MIP formulation:

$$\begin{aligned}
\min\quad & \sum_{k=1}^{q}\frac{1}{m_k}\sum_{\forall i\in C_k}\sum_{\ell=1}^{q}\left(\xi_{i\ell}^{+}+\xi_{i\ell}^{-}\right) & & \\
\text{s.t.}\quad & V(\mathbf{x}_i) - t_k + \xi_{ik}^{+} \geq \delta & & \forall i\in\{C_1,\ \ldots,\ C_k\},\ k = 1,\ \ldots,\ q-1 \\
& V(\mathbf{x}_i) - t_{k-1} - \xi_{ik}^{-} \leq -\delta & & \forall i\in\{C_k,\ \ldots,\ C_q\},\ k = 2,\ \ldots,\ q \\
& t_k - t_{k-1} \geq \varepsilon & & k = 1,\ldots,q-1 \\
& V(\mathbf{x}_*) = 0,\ V(\mathbf{x}^*) = 1 & & \\
& \xi_{ik}^{+}, \xi_{ik}^{-}\in\{0,\ 1\} & & i = 1,\ \ldots,\ m,\ k = 1,\ \ldots,\ q
\end{aligned} \tag{4}$$

Compared to model (2), this formulation distinguishes between the possible misclassifications for a patient i from risk category C_k through the binary error variables ξ^+ and ξ^-. More specifically, the first constraint compares the global value of every patient belonging in the set of risk categories $\{C_1, C_2, \ldots,$

DOI: 10.1002/mcda

$C_k\}$ (for each, $k=1, \ldots, q-1$), against the lower threshold t_k of category C_k. For instance, a patient from the high risk category C_1 (i.e. $k=1$) is compared (successively) against t_1 (the lower threshold of category C_1) and t_2 (the lower threshold of category C_2), up to t_{q-1} (the lower threshold of category C_{q-1}). Each of these comparisons is associated with a different error variable $\xi_{i\ell}^{+}$ ($\ell=k, \ldots, q-1$), which equals to 1 if and only if a patient i that actually belongs to the set of risk categories $\{C_1, C_2, \ldots, C_k\}$ is misclassified in any of the risk categories $\{C_{\ell+1}, C_{\ell+2}, \ldots, C_q\}$. For example, if a patient from the risk category C_1 is assigned into category C_3, then $\xi_{i1}^{+} = \xi_{i2}^{+} = 1$, thus indicating that there is a two-notch difference between the actual and the estimated classification.

In a similar manner, the second constraint compares the global value of every patient from the set of risk categories $\{C_k, C_{k+1}, \ldots, C_q\}$ (for each, $k=2, \ldots, q$), against the upper threshold t_{k-1} of category C_k. For instance, a patient from the low risk category C_q is compared (successively) against t_{q-1} (the upper threshold of category C_q) and t_{q-2} (the upper threshold of category C_{q-1}), up to t_2 (the upper threshold of category C_2). These comparisons are associated with error variables $\xi_{i\ell}^{-}$ ($\ell=2, \ldots, k$), which equal to 1 if and only if a patient i that actually belongs to the set of risk categories $\{C_k, C_{k+1}, \ldots, C_q\}$ is assigned (misclassified) into any of the risk categories $\{C_1, C_2, \ldots, C_{\ell-1}\}$.

Thus, the sum $\xi_{ik}^{+} + \xi_{i,k+1}^{+} + \cdots + \xi_{i,q-1}^{+}$ for a patient from risk category C_k equals the difference $\hat{y}_i - y_i$ (as in the example noted previously), when the patient is assigned into a lower risk category compared to its actual risk level (i.e. $\hat{y}_i > y_i$), whereas the sum $\xi_{i2}^{-} + \cdots + \xi_{ik}^{-}$ equals the difference $y_i - \hat{y}_i$, when the patient is assigned into a higher risk category compared to its actual risk level (i.e. $\hat{y}_i < y_i$). Obviously, the ordinal definition of the risk categories implies that $\xi_{i\ell}^{+} = 1$ whenever $\xi_{i,\ell+1}^{+} = 1$ and $\xi_{i\ell}^{-} = 1$ whenever $\xi_{i,\ell-1}^{-} = 1$.

4. RESULTS

4.1. Empirical setting

In this study, the two model fitting formulations described in the previous section are employed in a lexicography manner. In particular, model (2) is first used to obtain an additive evaluation model that minimizes the total weighted number of misclassifications while ignoring their magnitude. Then, at a second stage, problem (4) is solved to minimize the MWAE while controlling for the number of misclassified patients on the basis of the solution of model (2), that is, by adding the following constraint to problem (4):

$$\sum_{\forall i \in C_k} \left(\xi_{ik}^{\pm} + \xi_{i,k-1}^{\pm} \right) = E^{*}, \tag{5}$$

where $E^{*} = \sum_i \left(\sigma_i^{+} + \sigma_i^{-} \right)$ is the total number of misclassifications corresponding to the solution of model (2).

All optimization problems are solved with a quad-core personal computer with an Intel i7-2600 K/3.4 GHz processor and 16 GB of RAM, using the Gurobi 6 solver. With this computational environment, the mixed integer linear programming formulation of UTADIS II was easily solved to optimality, whereas problem (4) was much more challenging due to its increased complexity. In that respect, a time limit of 1 h was imposed during the solution process.

Table II. Model fitting metrics

	UTADIS II	MWAE	MWAE-Lex	WOLR
Overall classification accuracy	0.724	0.569	0.724	0.345
Average classification accuracy	0.843	0.691	0.793	0.244
Mean-weighted absolute error	0.679	0.530	0.588	1.627

MWAE, mean-weighted absolute error; WOLR, weighted ordinal logistic regression model.

4.2. Analysis of results

Table II presents some main model fitting measures for three different additive evaluation models, including the model resulting from the solution of the UTADIS II problem (2), the one obtained from the MWAE problem (4), the MWAE-Lex model obtained from the combination of the previous two approaches through the aforementioned lexicographic scheme and a weighted ordinal logistic regression model (King and Zeng, 2001). For each evaluation model, three fitting indices are calculated, namely: (i) the overall classification accuracy, defined as the percentage of patients correctly classified by the

DOI: 10.1002/mcda

Table III. Classification matrices for the UTADIS II and MWAE-Lex models (all entries in %)

			Model's classification						
			YYY	NYY	NYN	YNY	NNY	YNN	NNN
Actual classification	UTADIS II	YYY	100.0	0.0	0.0	0.0	0.0	0.0	0.0
		NYY	0.0	100.0	0.0	0.0	0.0	0.0	0.0
		NYN	0.0	0.0	100.0	0.0	0.0	0.0	0.0
		YNY	0.0	0.0	0.0	100.0	0.0	0.0	0.0
		NNY	33.3	8.3	0.0	0.0	58.3	0.0	0.0
		YNN	33.3	0.0	0.0	0.0	0.0	66.7	0.0
		NNN	4.1	17.6	5.4	2.7	2.7	2.7	64.9
	MWAE-Lex	YYY	100.0	0.0	0.0	0.0	0.0	0.0	0.0
		NYY	12.5	50.0	12.5	0.0	12.5	12.5	0.0
		NYN	0.0	0.0	66.7	0.0	11.1	0.0	22.2
		YNY	0.0	0.0	0.0	100.0	0.0	0.0	0.0
		NNY	16.7	0.0	8.3	0.0	66.7	0.0	8.3
		YNN	0.0	0.0	0.0	0.0	0.0	100.0	0.0
		NNN	2.7	1.4	6.8	8.1	4.1	5.4	71.6

model; (ii) the average classification accuracy, defined by the objective function of problem (2) and (iii) the MWAE index (3).

The basic UTADIS II performs best in terms of the average classification accuracy, which corresponds to the objective function of the MIP (2), whereas the MWAE model minimizes the MWAE on the basis of the optimization problem (4). The performance of the MWAE model, however, on the two classification accuracy criteria is significantly lower compared to the results of UTADIS II. The MWAE-Lex approach provides a good balance between the two other models. In particular, compared to UTADIS II, MWAE-Lex has slightly lower average classification accuracy (by about 6% in relative terms) while improving the weighted absolute error by about 13% (again in relative terms). On the other hand, compared to MWAE, the MWAE-Lex model has a bit higher weighted absolute error but yields much higher classification accuracies. Finally, the ordinal logistic regression model performs consistently worst than all multicriteria models.

Table IV. Weights of the criteria in the decision models developed with UTADIS II and the lexicographic approach

	UTADIS II	MWAE-Lex
Type of AF	0.00	2.04
AF duration	14.29	10.70
AF episode duration	16.32	21.32
LA diameter	18.41	12.48
LA volume	12.33	13.03
Applications	12.19	19.21
Fluoroscopy time	26.46	21.21

MWAE, mean-weighted absolute error; AF, atrial fibrillation; LA, left atrium.

The detailed classification matrices for the results of the additive decision models constructed with the UTADIS II and the MWAE-Lex approaches are presented in Table III. It is evident that the UTADIS II model performs very well for patients in the high risk categories YYY–YNY but it leads to some significant misclassifications. For instance, about 33% of the patients from the low risk category YNN are classified as very risky cases (category YYY), whereas 17.6% of the patients with no recurrence indications (category NNN) are classified as high risk patients in category NYY. Overall, the UTADIS II model is clearly biased towards overestimating the recurrence risk as all classification errors involve cases misclassified into higher risk categories. On the other hand, the decision model constructed with the lexicographic scheme provides more balanced results with a considerable reduction of the major misclassifications noted previously.

Detailed results for the weights of the criteria in the UTADIS II and MWAE-Lex models are presented in Table IV. Both models indicate that the duration of AF episodes, the number of applications of high frequency current during AF ablation, the fluoroscopy time and the LA volume are major factors contributing to the decisions regarding the monitoring and evaluation of a patient's condition. The type of AF on the other hand, seems to be a less important factor.

DOI: 10.1002/mcda

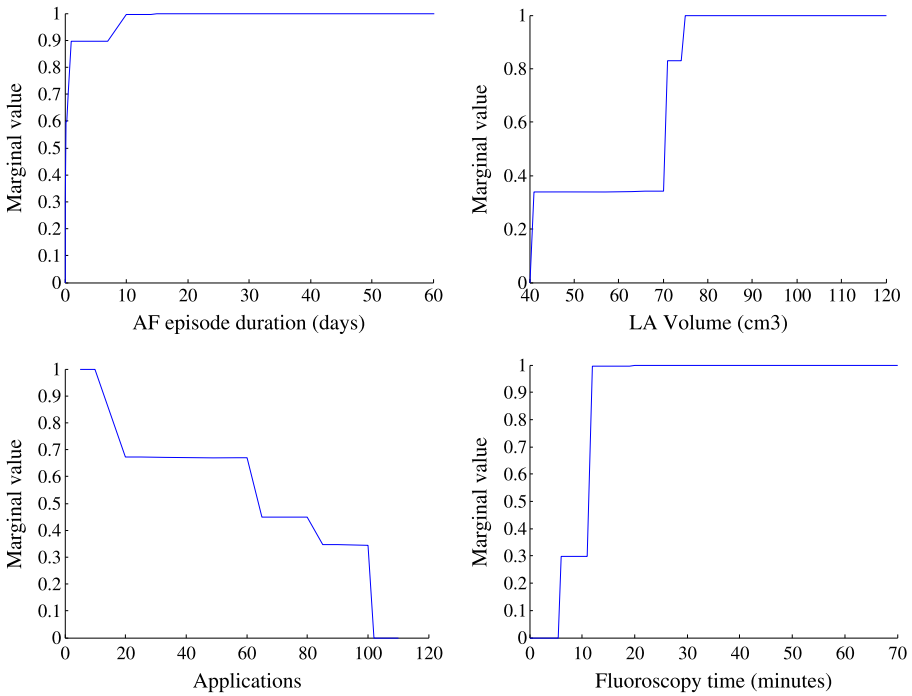

Figure 2. Marginal value functions for the four evaluation criteria with the highest weights. AF, atrial fibrillation; LA, left atrium.

The marginal value functions for the criteria with the highest weights in the MWAE-Lex decision model are illustrated in Figure 2. The function for the AF episode duration criterion has a concave form indicating that the recurrence risk increases rapidly even for cases with low AF episode duration and remains at high levels for cases with duration above 1 day. A similar concave form is also evident for the fluoroscopy time criterion, according to which the recurrence risk increases significantly in cases where the fluoroscopy time is more than 10 min. All high risk patients had fluoroscopy time greater than 10 min. This may be influenced by difficulties faced with navigating and positioning the catheters into the PV for patients with high LA volume. Additionally, patients with longstanding AF episodes have more intense and chaotic electric disorganization of the atria, demanding prolonged and repeated lesions for PVI, which are associated with longer operational times and therefore longer fluoroscopy times. On the other hand, the marginal value function for the LA volume criterion reveals that recurrence risk increases significantly for patients with LA volume above 70 cm^3. It is worth noting that under normal conditions, LA volume ranges between 25 and 58 cm^3. Therefore, the model does confirm that the ablation operation is likely to be unsuccessful for patients with LA volume much higher than normal levels. Finally, the function for the number of applications of high frequency current during AF ablation has a decreasing form, with the recurrence risk being much lower when there are more than 100 applications. These insights provide cardiologists with a disaggregated view of the global recurrence risk assessment result for each particular patient, in terms of his/her medical status on each one of the prognostic attributes. This is valuable information that strengthens the medical decision-maker's confidence on the model's reasoning and results, facilitates their qualitative analysis and supports the process for providing sound medical treatment to individual patients.

4.3. Robustness analysis

In a preference disaggregation context, such as the one adopted in this study for the inference of preferential information from a set of decision instances, the robustness of the obtained conclusions is a critical issue. The robustness concern (Roy, 2010) has recently received considerable attention among MCDA researchers. The MCDA literature related to the robustness concern in disaggregation techniques can be categorized into two main streams. The first

DOI: 10.1002/mcda

focuses on providing a range of recommendations (instead of single point results) based on the full set of models compatible with the information provided by the decision-maker (see, for instance, Greco *et al*., 2010). When inconsistencies exist in the data (i.e. classification errors), these are resolved (Mousseau *et al*., 2003) prior to the formulation of the recommendations. An alternative approach adopts a post-optimality perspective focusing on investigating the existence of multiple optimal or near-optimal models, after a decision model has been constructed using a set of reference examples (Siskos and Grigoroudis, 2010).

In this study, we adopt the latter approach in order to examine the existence of alternative decision models that describe the classification of the given patients in the available sample in the same way the obtained MWAE-Lex model does. If other very different models exist, that would raise concerns about the validity of the recommendations derived with the MWAE-Lex model for patients outside the reference sample.

Similarly to the post-optimality analysis often employed in the context of UTA-like methods (Jacquet-Lagrèze and Siskos, 1982; Siskos and Grigoroudis, 2010), in order to examine the existence of other optimal models, we first fix all the classification assignments obtained from the MWAE-Lex model and then check the variability of different models that provide the same assignments for the patients in the sample. More specifically, let $\hat{C}_1$, $\hat{C}_2$, …, $\hat{C}_7$ denote the sets of patients assigned by the MWAE-Lex model in each of the seven recurrence risk classes. Then, all additive value models that are compatible with the assignments of the MWAE-Lex model should satisfy the following constraints:

$$\begin{aligned} &V(\mathbf{x}_i) \geq t_k + \delta && \forall i \in \hat{C}_k,\ k = 1,\ \ldots,\ 6 \\ &V(\mathbf{x}_i) \leq t_{k-1} - \delta && \forall i \in \hat{C}_k,\ k = 2,\ \ldots,\ 7 \\ &t_k - t_{k-1} \geq \varepsilon && k = 1,\ \ldots,\ 6 \\ &V(\mathbf{x}_*) = 0,\ V(\mathbf{x}^*) = 1 \end{aligned} \quad . \tag{6}$$

In order to explore the robustness of the solutions in the polyhedron defined by these constraints, we follow two approaches. First, a post-optimality analysis (Jacquet-Lagrèze and Siskos, 1982) is employed to identify extreme solutions corresponding to the maximization and minimization of the weight for each criterion (separately). Additionally, we also examine the divergence between the weights of the criteria in the developed MWAE-Lex model and the ones that correspond to the analytic centre of the aforementioned polyhedron. As noted by Bous *et al*. (2010), decision models close to the centre of feasible polyhedron are more robust representations (compared to solutions near the boundaries) of the preferential information embodied in a set of reference examples. The identification of analytic centre can be easily performed through the solution of an optimization problem with linear constraints and logarithmic barrier objective function (Bous *et al*., 2010). A similar approach for the construction of a robust and representative sorting tool is outlined by Greco *et al*. (2011).

The criteria weights obtained from the aforementioned two approaches are shown in Table V (the post-optimality results include the minimum, maximum and the average of each criterion's weight). The results obtained from the post-optimality approach indicate that there are only very minor variations in the weights of the criteria between different models compatible with the assignments of the MWAE-Lex model. Furthermore, both the post-optimality results as well as those obtained from the analytic centre are extremely similar to the ones of the MWAE-Lex model (cf. Table IV). The robustness of decision model developed with the lexicographic approach was also verified with the average stability index (ASI) proposed by Grigoroudis and Siskos (2002), which provides a comprehensive measure of the robustness of an inferred additive value model taking into account not only the weights of the criteria but also variations with respect to the marginal value functions. By definition, ASI ranges in a 0%–100% scale, with higher values indicating more stable models. In the context of the data in this study, the ASI of the MWAE-Lex model was found to be 99.63%, slightly improved over the ASI for the UTADIS II model (99.18%).

Table V. Robustness analysis results for the weights of the criteria

	Post-optimality (min, mean and max)	Analytic centre
Type of AF	[2.00, 2.05, 2.07]	2.05
AF duration	[10.64, 10.75, 11.07]	10.74
AF episode duration	[21.24, 21.29, 21.59]	21.31
LA diameter	[10.48, 11.42, 12.82]	11.32
LA volume	[12.94, 14.12, 15.28]	14.20
Applications	[19.08, 19.14, 19.33]	19.18
Fluoroscopy time	[21.14, 21.23, 21.49]	21.21

AF, atrial fibrillation; LA, left atrium.

DOI: 10.1002/mcda

5. CONCLUSIONS AND FUTURE PERSPECTIVES

The development of medical decision aiding models is a challenging issue with important practical implications. In this study, we presented a real-world case study involving the development of such a model for monitoring the post-operative condition of AF patients. The model combines a number of medical factors that are potential predictors of AF recurrence and classify patients into risk categories.

A preference disaggregation approach was used to develop an appropriate model, combining two main fitting criteria through a lexicographic scheme. This lexicographic approach was found to lead to a good trade-off between the fitting criteria, resulting to a model with a small number and magnitude of misclassifications, with the overall accuracy rate ranging higher than 70%. The model performed very well in identifying high risk patients, whereas low-risk cases were found to be more difficult to be evaluated accurately, thus indicating the more detailed analysis is further needed for such cases. In that regard, it could be particularly beneficial to combine the model's results with the expertise and judgement of expert cardiologists as well as to examine the usefulness of additional prognostic criteria. Among the recurrence risk criteria used in the analysis, the duration of the most recent AF, the volume of the LA and the two criteria related to the PVI treatment (number of applications of high frequency current and fluoroscopy time) were found to contribute to the assessment of AF recurrence risk. The conducted robustness analysis verified the validity of these results. These findings are in accordance with complex nature of AF recurrence, which is due to a combination of factors regarding the nature of a patient's AF, his/her physical characteristics and the PVI operation. According to an expert medical doctor, the results of the model were found to be satisfactory, both in terms of their classification performance as well as in terms of their interpretation, their implications in practice and the insights that it provides.

Future research can focus on the consideration of a number of different variables of the aforementioned medical procedure or other interventional methods. On the methodological side, other model fitting criteria could be considered model, focusing, for instance, on eliminating/reducing important errors for specific risk categories or patient cases, which are explained poorly by the constructed model. The use of efficient optimization techniques (e.g. meta-heuristics) is also a point that could be considered in order to improve the computational efficiency of the model construction process. Comparisons with other multicriteria and data mining techniques could also be considered, focusing on the robustness of the results for patients outside the reference set (out of sample generalization ability).

ACKNOWLEDGEMENT

This research has been co-financed by the European Union (European Social Fund) and Greek national funds through the Operational Programme 'Education and Lifelong Learning'.

ENDNOTES

1. The case YYN is missing from the analysis, as it is highly unlikely a patient with early and late atrial fibrillation recurrence to go asymptomatic at the very late time window (there was no such case in our data sample).
2. When $V(\mathbf{x}_i) = t_k$ for some $k = 1, \ldots, q-1$, then the classification of patient i is arbitrary. In such cases, we assume that patient i is assigned to risk group k (no such cases were observed in our application).

REFERENCES

Alonso-Atienza F, Rojo-Álvarez JL, Rosado-Muñoz A, Vinagre JJ, García-Alberola A, Camps-Valls G. 2012. Feature selection using support vector machines and bootstrap methods for ventricular fibrillation detection. *Expert Systems with Applications* **39**: 1956–1967.

Ammenwerth E, Nykänen P, Rigby M, de Keizer N. 2013. Clinical decision support systems: need for evidence, need for evaluation. *Artificial Intelligence in Medicine* **59**: 1–3.

Angelucci E, Barosi G, Camaschella C, Cappellini MD, Cazzola M, Galanello R, Marchetti M, Piga A, Tura S. 2008. Italian Society of Hematology practice guidelines for the management of iron overload in thal assemia major and related disorders. *Haematologica* **93**: 741–752.

Baltussen R, Youngkong S, Paolucci F, Niessen L. 2010. Multi-criteria decision analysis to prioritize health interventions: capitalizing on first experiences. *Health Policy* **96**: 262–264.

Belacel N. 2000. Multicriteria assignment method PROAFTN: methodology and medical application. *European Journal of Operational Research* **125**: 175–183.

Berner ES, La Lande TJ. 2007. Overview of clinical decision support systems. In *Clinical Decision Support Systems* Berner ES (ed). Springer: New York; 3–22.

DOI: 10.1002/mcda

Bous G, Fortemps P, Glineur F, Pirlot M. 2010. ACUTA: a novel method for eliciting additive value functions on the basis of holistic preference statements. *European Journal of Operational Research* **206**: 435–444.

Camm AJ, Lip GYH, De Caterina R, Savelieva I, Atar D, Hohnloser SH, Hindricks G, Kirchhof P. 2012. 2012 focused update of the ESC Guidelines for the management of atrial fibrillation: an update of the 2010 ESC Guidelines for the management of atrial fibrillation. Developed with the special contribution of the European Heart Rhythm Association. *European Heart Journal* **33**: 2719–2747.

Chesnokov YV. 2008. Complexity and spectral analysis of the heart rate variability dynamics for distant prediction of paroxysmal atrial fibrillation with artificial intelligence methods. *Artificial Intelligence in Medicine* **43**: 151–165.

Dolan JG. 2010. Multi-criteria clinical decision support: a primer on the use of multiple criteria decision making methods to promote evidence-based, patient-centered healthcare. *Patient* **3**: 229–248.

Dolan JG, Frisina S. 2002. Randomized controlled trial of a patient decision aid for colorectal cancer screening. *Medical Decision Making* **22**: 125–139.

Doumpos M, Zopounidis C. 2002. *Multicriteria Decision Aid Classification Methods, Applied Optimization*, Springer: New York.

D Bois P, Brans JP, Cantraine F, Mareschal B. 1989. MEDICIS: an expert system for computer-aided diagnosis using the PROMETHEE multicriteria method. *European Journal of Operational Research* **39**: 284–292.

Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, Sam J, Haynes RB. 2005. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *Journal of the American Medical Association* **293**: 1223–1238.

Goetghebeur MM, Wagner M, Khoury H, Levitt RJ, Erickson LJ, Rindress D. 2012. Bridging health technology assessment (HTA) and efficient health care decision making with multicriteria decision analysis (MCDA): applying the EVIDEM framework to medicines appraisal. *Medical Decision Making* **32**: 376–388.

Goletsis Y, Papaloukas C, Fotiadis DI, Likas A, Michalis LK. 2004. Automated ischemic beat classification using genetic algorithms and multicriteria decision analysis. *IEEE Transactions on Biomedical Engineering* **51**: 1717–1725.

Greco S, Kadziński M, Słowiński R. 2011. Selection of a representative value function in robust multiple criteria sorting. *Computers and Operations Research* **38**: 1620–1637.

Greco S, Mousseau V, Słowiński R. 2010. Multiple criteria sorting with a set of additive value functions. *European Journal of Operational Research* **207**: 1455–1470.

Grigoroudis E, Siskos Y. 2002. Preference disaggregation for measuring and analysing customer satisfaction: the MUSA method. *European Journal of Operational Research* **143**: 148–170.

Hamacher HW, Küfer K-H. 2002. Inverse radiation therapy planning—a multiple objective optimization approach. *Discrete Applied Mathematics* **118**: 145–161.

Hardin MJ, Chhieng DC. 2007. Data mining and clinical decision support systems. In *Clinical Decision Support Systems*, Berner ES (ed). Springer: New York; 44–63.

Jacquet-Lagrèze E, Siskos J. 1982. Assessing a set of additive utility functions for multicriteria decision making: the UTA method. *European Journal of Operational Research* **10**: 151–164.

Jacquet-Lagrèze E, Siskos Y. 2001. Preference disaggregation: 20 years of MCDA experience. *European Journal of Operational Research* **130**: 233–245.

Kawamoto K, Houlihan CA, Balas EA, Lobach DF. 2005. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* **330**: 765.

Keeney RL, Raiffa H. 1993. *Decisions with Multiple Objectives: Preferences and Value Trade-offs*, Cambridge University Press: Cambridge.

King G, Zeng L. 2001. Logistic regression in rare events data. *Political Analysis* **9**: 137–163.

Kirchhof P, Lip GYH, Van Gelder IC, Bax J, Hylek E, Kaab S, Schotten U, Wegscheider K, Boriani G, Brandes A, Ezekowitz M, Diener H, Haegeli L, Heidbuchel H, Lane D, Mont L, Willems S, Dorian P, Aunes-Jansson M, Blomstrom-Lundqvist C, Borentain M, Breitenstein S, Brueckmann M, Cater N, Clemens A, Dobrev D, Dubner S, Edvardsson NG, Friberg L, Goette A, Gulizia M, Hatala R, Horwood J, Szumowski L, Kappenberger L, Kautzner J, Leute A, Lobban T, Meyer R, Millerhagen J, Morgan J, Muenzel F, Nabauer M, Baertels C, Oeff M, Paar D, Polifka J, Ravens U, Rosin L, Stegink W, Steinbeck G, Vardas P, Vincent A, Walter M, Breithardt G, Camm AJ. 2012. Comprehensive risk reduction in patients with atrial fibrillation: emerging diagnostic and therapeutic options—a report from the 3rd Atrial Fibrillation Competence NETwork/European Heart Rhythm Association consensus conference. *Europace* **14**: 8–27.

Mohebbi M, Ghassemian H. 2012. Prediction of paroxysmal atrial fibrillation based on non-linear analysis and spectrum and bispectrum features of the heart rate variability signal. *Computer Methods and Programs in Biomedicine* **105**: 40–49.

Mousseau V, Figueira J, Dias L, Gomes da Silva C, Clímaco J. 2003. Resolving inconsistencies among constraints on the parameters of an MCDA model. *European Journal of Operational Research* **147**: 72–93.

O'Sullivan D, Wilk S, Michalowski W, Słowiński R, Thomas R, Kadziński M, Farion K. 2014. Learning the preferences of physicians for the organization of result lists of medical evidence articles. *Methods of Information in Medicine* **53**: 344–356.

Postmus D, Tervonen T, van Valkenhoef G, Hillege HL, Buskens E. 2014. A multi-criteria decision analysis perspective on the health economic evaluation of medical interventions. *Eur. J. Heal. Econ.* **15**: 709–716.

DOI: 10.1002/mcda

Rahimi S, Gandy L, Mogharreban N. 2007. A web-based high-performance multicriteria decision support system for medical diagnosis. *International Journal of Intelligence Systems* **22**: 1083–1099.

Roy B. 2010. Robustness in operational research and decision aiding: a multi-faceted issue. *European Journal of Operational Research* **200**: 629–638.

Schlaefer A, Viulet T, Muacevic A, Fürweger C. 2013. Multicriteria optimization of the spatial dose distribution. *Medical Physics* **40**: 121720.

Siskos Y, Grigoroudis E. 2010. New trends in aggregation-disaggregation approaches. In *Handbook of Multicriteria Analysis* Zopounidis C, Pardalos PM (eds). Springer: Berlin Heidelberg; 189–214.

Tervonen T, van Valkenhoef G, Buskens E, Hillege HL, Postmus D. 2011. A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. *Statistics in Medicine* **30**: 1419–1428.

Van Valkenhoef G, Tervonen T, Zwinkels T, de Brock B, Hillege H. 2013. ADDIS: a decision support system for evidence-based medicine. *Decision Support Systems* **55**: 459–475.

Wann LS, Curtis AB, January CT, Ellenbogen KA, Lowe JE, Estes NAM, Page RL, Ezekowitz MD, Slotwiner DJ, Jackman WM, Stevenson WG, Tracy CM, Fuster V, Rydén LE, Cannom DS, Le Heuzey J-Y, Crijns HJ, Olsson SB, Prystowsky EN, Halperin JL, Tamargo JL, Kay GN, Jacobs AK, Anderson JL, Albert N, Hochman JS, Buller CE, Kushner FG, Creager MA, Ohman EM, Ettinger SM, Guyton RA, Tarkington LG, Yancy CW. 2011. 2011 ACCF/AHA/HRS focused update on the management of patients with atrial fibrillation (updating the 2006 guideline): a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation* **123**: 104–123.

You JJ, Singer DE, Howard PA, Lane DA, Eckman MH, Fang MC, Hylek EM, Schulman S, Go AS, Hughes M, Spencer FA, Manning WJ, Halperin JL, Lip GYH. 2012. Antithrombotic therapy for atrial fibrillation: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* **141**: e531S–75S.

Zopounidis C, Doumpos M. 2002. Multicriteria classi fication and sorting methods: a literature review. *European Journal of Operational Research* **138**: 229–246.

DOI: 10.1002/mcda