Development of a Multicriteria Decision Aiding Model for performance monitoring and evaluation of Health Care Units

Diana Isabel Fernandes Lopes
Mestrado Integrado em Engenharia Biomédica
Instituto Superior Técnico, Technical University of Lisbon, Portugal

Abstract - *Primary Health Care (PHC) is a key component of the Portuguese health system, representing the first level of contact of individuals with the National Health Service (NHS). In the scope of the PHC's reform that started in 2005 and is still in course, a contracting system was introduced with the objective of bringing more accountability and efficiency into the primary care system in the line with improving health outcomes given available resources. Given budget cuts and limited resources and the objectives of the Portuguese health care system, it is critical to have a correct identification of major problems of PHC providers and having tools to assist analyzing how to improve their performance. Within this context, the South West Primary Health Care Group (PHCG) is also facing the challenge of making decisions on how to best utilize scarce resources.*

This thesis aims to answer to the PHCG's call for help to build a system to evaluate and monitor the performance of Functional Units (FsU), as well as to prioritize corrective actions with a potential to improve FsU's performance. For monitoring and evaluating FsU's performance a multicriteria model based on the Choquet integral (CI) operator and MACBETH was developed; for prioritizing corrective actions, a resource allocation model constructed with the PROBE software was proposed. These models were constructed in a sequence of several interconnected steps built using a socio-technical approach.

This model was applied to three Family Health Units (Arandis, Dom Jordão and Gama) and to a Personalized Health Care Unit (Alenquer). The results support the PHCG in monitoring and evaluating the performance of the FsU as well as the prioritization of the corrective actions. Within the literature in the area, we conclude that the extension of MACBETH to the CI operators is still in its first steps, and needs to be developed to potentiate easier and more integrated applications; and that the proposed models are innovative in the context of health care literature.

Keywords: *Primary Health Care, Functional Units, Multicriteria Model, Choquet integral, Resource Allocation, Portugal.*

1. Introduction

Primary Health Care (PHC) is a key component of the Portuguese's health system, representing the first level of contact of individuals with the National Health Service (NHS) [1]. In most health systems, PHC is considered an essential type of universal care to be provided to all individuals and responsible for treatment and prevention of a large number of health problems of a population. Regarding the problems and challenges that the Portuguese PHC sector was facing, in 2005 a set of reforms was initiated. The objective has been to create more and better health care for citizens and to improve accessibility, proximity and quality in health care provision. This set of reforms is still in course and essentially aims to reorganize and modernize primary health care centres (PHCC), structuring them in functional units (FsU) with complementary missions. In addition, a contracting system was recently introduced aiming to bring more accountability and efficiency into the primary care system in the line with improving health outcomes given available resources [2].

Given budget cuts and limited resources and the objectives of the Portuguese health care system, it is critical to have a correct identification of major problems of PHC providers and having tools to assist analyzing how to improve their performance. Within this context, the South West Primary Health Care Group (PHCG) is also facing the challenge of making decisions on how to best utilize scarce resources.

Due to the scarcity of resources, the South West Primary Health Care Group (PHCG) wants to properly evaluate the performance of functional units subject to the contracting processes in order to identify the need for corrective actions. It also wants to know which set of actions have the greatest potential to improve FsU's performance. In fact, the correct identification of major problems in their FsU and appropriate management of available resources are imperative for proper decision making. Thus, this thesis was developed in an attempt to meet the request of the South West PHCG, having two main objectives: build a system to (1) evaluate and monitor the performance of Functional Units (FsU), as well as (2) to prioritize corrective actions with potential to improve FsU's performance.

To achieve these goals, it was made a literature review. The review has shown that literature is scarce and does not provide methods for, in a rational and transparent way, helping and supporting the PHCG's objectives in a satisfactory manner. Therefore this these has developed a multicriteria model based on a extension of MACBETH to the Choquet integral (CI) operators to evaluate and monitor the FsU's performance. This methodology intends to tackle the various generic faults and disadvantages of the methods found in literature and to explore, in an innovate way, the process of modeling interacting criteria. Once evaluated the performance of the FsU's under analysis, this thesis illustrates the necessary steps to select, in a rational and transparent way, the portfolio of actions with the highest potential to improve

FsU's performance. For prioritizing corrective actions, a resource allocation model constructed with the PROBE software was proposed, which allows to address constraints and synergies between actions [3].

This paper is structured as follows: Section 2 presents a brief review of related literature. Section 3 contains the proposed methodology. Section 4 describes the implementation of the proposed methodology. Section 5 shows the results of model's application. Finally section 6 presents some discussion and concluding remarks.

2. Literature Review

Performance measurement is fundamental to any management control. Due to the increasing restrictions on the health sector, the development of management tools, that enable the evaluation and communication of the extent to which various aspects of health system meet key objectives (performance measurement), has assumed critical importance to the delivery of efficient and effective health care [4].

Records of performance measurement efforts in health systems can be traced back to at least 250 years ago. However, just over the past thirty years there has been a dramatic growth in health system performance measurement. A confluence of forces has led to this growth. On the supply side, great advances in information technology have made it much cheaper and easier to collect and process data. On the demand side, health systems have come under intense cost containment pressures and there has emerged a growing demand for increased oversight and accountability in health professions and health institutions [5].

Some of the most well known performance measurement frameworks are Six Sigma approach, European Foundation for Quality Management model, Performance Pyramid, Results and Determinants Framework and Balanced Scorecard. In fact, these frameworks emphasize the need of measurement systems to make explicit the trade-offs between the various performance measures, but are vague on how to deal with these trade-offs. Furthermore, there is no information about its aggregation mechanisms and some of the frameworks is related to high costs of implantation [6, 7].

Thus, we believe that the use of a Multicriteria Decision Analysis (MCDA) approach can be helpful in this context. Up to our knowledge, multicriteria models found in the health literature are scarce and the existing ones do not take into account possible synergies or interactions between criteria (nor have been used for performance measurement of health care providers). Thus, these models are inaccurate or limited to a few cases where there is, in fact, independence and absence of interactions between criteria [8]. In the industrial sector, however, some studies exploited in an integrated manner the interaction phenomena between criteria, with the use of the CI operator [9-12].

The literature found for the prioritization of actions is also scarce, and the existing ones (economical and non-economic approaches) have restrictive problems that preclude their application in our study. Some of these problems are related to the lack of transparency which can lead to a misunderstanding of results and to not taking into account two basic principles in prioritization: the opportunity cost and the incremental change principles [13-15]. Similarly, the review found out that there are few applications of multicriteria models for this purpose and existing ones do not take into account possible interactions between criteria.

In other to tackle the gaps mentioned above, this study developed a MCDA model based on the extension of MACBETH to the CI operators. In fact, this framework allows: (1) to build a transparent model through an iterative and social process; (2) to take into account multiple criteria relevant to the decision makers; (3) to consider the interactions between criteria; (4) to determinate the FsU's performance in different levels of specifications; and (5) to prioritize corrective actions.

3. Methodological Framework

The construction of the multicriteria framework aimed to determine FsU's performance (as well as the incremental benefit of management actions) and involves several and interconnected steps built using a socio-technical approach: (1) model structuring; (2) construction of the value functions; (3) formulation of the aggregation operators to collapse the multiple dimensions of performance into a global one, making possible the determination of global performance of the FsU's under analysis; (4) determination of area and criteria weights; and (5) construction of performance categories, to more easily distinguish the need for corrective action.

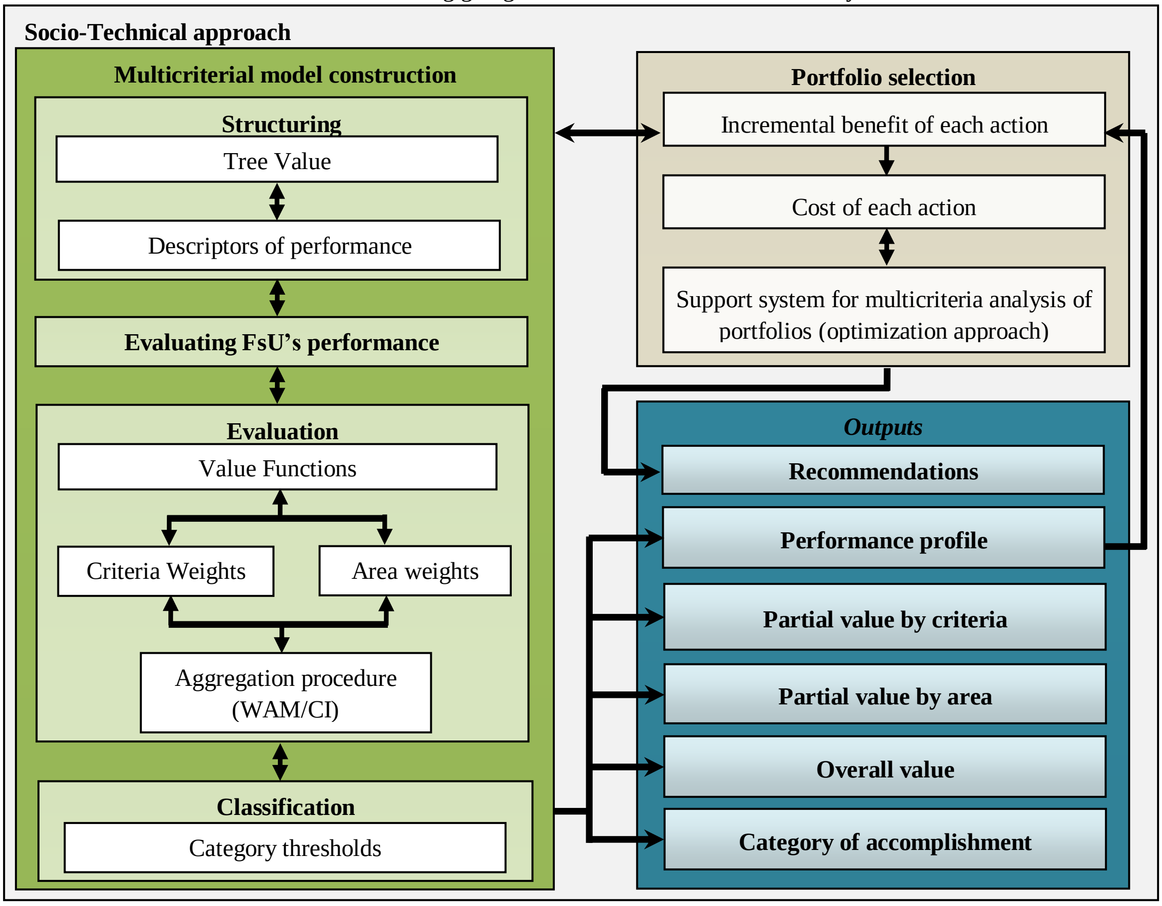
The social component included various meetings with the Executive Director (ED) of the South West in the model structuring and construction of performance categories as well as two decision conferences with the group of decision-makers (ED and the Clinic Council President of the South West PHCG, the coordinator of the PHCU of Torres Vedras and two doctors) for the construction of value functions and determination of area and criteria weights.

In the second step, we illustrate a support system based on resource allocation to help PHCG prioritizing corrective actions, in view to improve FsU's performance. To that end, it was used the PROBE software. The methodology and its various activities are specified in figure 1.

3.1. Model structuring

The structuring the multicriteria model consisted of an interactive learning process between the decision-makers, ED and Clinic Council President of the South West PHCG, and the facilitator (the author), in order to (1)

Figure 1: Activities involved in the construction of the methodology. Note that the double arrows mean that this is an iterative model, allowing going back and forth whenever necessary.



specify the relevant criteria for the decision-makers and (2) operationalize each criterion.

To that end, firstly, it was necessary to discuss the PHCG's strategic objectives in order to recognize the main areas of concern. Then, it was necessary to reflect about these areas in order to identify specific criteria. In the sense of providing a practical visual overview of the structure of concerns in several levels of increasing specification, it is useful to group the criteria in a value tree according to the areas of concern [16].

Secondly, it was necessary to operationalize each criterion, by assigning a descriptor of performance and defining levels of reference. A descriptor is an ordered set of plausible impact levels in terms of a criterion, intended to serve as a basis to appraise, as much as possible objectively, FsU's performance on each criterion [17]. Depending on the context, descriptors may be quantitative (continuous or discrete) or qualitative [16]. The definition of levels of reference, like "target" and "base" used in this study, has three objectives [18]:

1) it contributes significantly to the intelligibility of the criteria;
2) it objectifies the notion of intrinsic attractiveness of each FU, by assigning it to one of the following categories: (a) "very positive" FU's performance, when it is at least as attractive as a fictitious 'target' FU's performance, (b) "positive" FU's performance, when it is at least as attractive as a 'neutral' fictitious FU's performance, but less attractive than a fictitious 'target' FU's performance, and (c) "negative" FU's performance, if it is less attractive than a 'neutral' fictitious FU's performance;
3) it allows the use of a criteria-weighting procedure.

3.2. Construction of Value Functions

The process of building a value function for each criterion permits the translation of performance into value. Thus, it measures the FsU's attractiveness at the criterion level. Note that at the two reference levels defined previously, designated as 'target' and 'base' levels of performance, are assigned 100 and 0 points, respectively (see equation (2)) [19].

The adoption of the MACBETH approach for this purpose in the South West PHCG context was motivated by the deliberate intent of avoiding the difficulties of some decision-makers (DM) to elicit quantitative judgments. In fact, the MACBETH procedure (with the help of M-MACBETH software) allows constructing the value functions by means of qualitative judgments of differences in attractiveness between performance levels, two at a time. To facilitate each pairwise comparison the DM is asked to choose one of the MACBETH qualitative categories of difference in attractiveness: 'is the difference very weak, weak, moderate, strong, very strong, or extreme? During this questioning process, the facilitator fills in a matrix with the categorical judgments of DM. Note that judgment disagreement or hesitation between two or more categories is also allowed. Each time a qualitative judgment is elicited, the consistency is verified and suggestions are offered to solve eventual inconsistencies. After the consistency verification, the software derives a value function at each criterion, which has to be validated by the decision makers.

In order to evaluate the FsU's attractiveness at the level of each subarea, the value scores on the criteria should be aggregated [19]. To that end, firstly, it is necessary to determine criteria weights.

3.3. Weighting criteria

Trade-off, Swing-Weighting and MACBETH are the main methods of eliciting weights. This thesis proposes the determination of criteria weights, in an innovative way, through the mechanism of 2-additive CI. Unlike the previous procedures, this mechanism allows to consider the possible interactions between criteria. However, since the use of this mechanism becomes impractical when applied to areas with more than four criteria, it is also described the procedure that is supposed to use in that cases (MACBETH procedure). In subsection A is described the MACBETH procedure for the determination of criteria weights (to use when the mechanism of 2-additive IC is impractical) as well as its mechanism of aggregation (Weighted Arithmetic Mean - WAM). In subsection B, it is described the 2-additive CI procedure for criteria weights determination as well as its related aggregation operator to determine the overall performance of FsU (as well as the incremental benefit of corrective actions).

A. MACBETH weighting process

The adoption of this procedure implies an additive value model, which implicitly assumes 'difference independence', which means: the difference in attractiveness between performance level on one criterion does not depend and can be measured independently of the levels on other criteria [19]. So, the first step, in this procedure, consists on verifying this hypothesis.

The mathematical formulation of its aggregation operator (WAM) is given by the expression:

$$P_{Ag}^{u}(x_{1u},\dots,x_{nu}) = \sum_{i=1}^{n} w_i p_i(x_{iu})\,, with\ \sum_{i=1}^{n} w_i = 1 \quad (1)$$

$$and\ w_i > 0\ and\ \begin{cases} p_i(target_i) = 100 \\ p_i(base_i) = 0 \end{cases} \quad (2)$$

where P_{Ag}^{u} is the performance value of the FU u, which measures its attractiveness in terms of each area; n is the number of criteria involved; x_{iu} is the impact of u in the criterion i; w_i is the weight of criterion i; $p_i(x_{iu})$ is the partial value of u in the ith criterion; $target_i$ and $base_i$ are, respectively, the target and base impact levels in the descriptor of the ith criterion.

Let $FU^F = \{FU^0, FU^1, \dots, FU^j, \dots, FU^n\}$ be a reference set of fictitious FsU, where FU^0 is a reference fictitious FU with the base impacts in all criteria, and $FU^j (j = 1,..,n)$ is a reference fictitious FU with the target impact in the jth criterion and the base impacts in all the other criteria. Note that n is the number of criteria present in each area.

After this mathematical overview, the MACBETH procedure for weighting the criteria consists in asking the DM: (1) first, to order the FU^j in terms of attractiveness; (2) second, to judge qualitatively the difference of attractiveness between any two reference FsU [17]. These judgments are introduced in a matrix present in the M-MACBETH software, being, each at a time, consistently verified. After that, the software derives a weighting scale, which needs to be discussed, possibly adjusted, and then validated by the DM.

This procedure provides a simple and transparent approach in solving complex multicriteria problems, hence its wide applicability to most methods for MCDA [17-21]. However, this mechanism of aggregation does not consider possible interactions between criteria. Consequently, this approach is limited to certain cases (which there is no interaction between criteria) or may lead to inaccurate results. Thus, in the next subsection is describe the 2-additive CI mechanism which enables the determination of criteria weights, taking into account the possible interactions between them.

B. The 2-additive Choquet integral

The operators of the CI family were introduced by Murofushi and Sugeno and are interesting because they include a lot of generalized mean operators [11]. In particular, they can be written under the form of a conventional weighted mean modified by effects coming from interactions between *k* of *n* criteria, being called '*k*-additive'. In this sense, *k*-additivity property fixes the degree of interaction between criteria: 1-additivity does not permit interaction (and this operator acts as a WAM operator); 2-additivity allows interaction up to 2 criteria, etc. Note that a *k*-additive operator needs $\sum_{a=1}^{k}\binom{n}{a}$ coefficients to be defined, which makes, in practice, the 2-additivity the best compromise between low complexity and richness of the model [22]. Thus, in this study we proposed the use of the 2-additive CI operator. This operator is based on two types of parameters [11]:

- Importance of each criterion in relation to all the other contributions to the overall performance evaluation by the so-called Shapley parameters ($v_i's$) that satisfy the condition $\sum_{i=1}^{n} v_i = 1$, which is a natural condition for decision makers where v_i acts a weight in the WAM operator.
- The interaction parameters I_{ij} of any pair of criteria, that range in [-1,1]:
  - A positive value of I_{ij} implies a conjunctive behavior between criteria *i* and *j*, so that simultaneous satisfaction of both criteria is significant for the global score.
  - A negative value I_{ij} implies a disjunctive behavior between criteria *i* and *j*, so that the

satisfaction of either one of them is sufficient to have a significant effect on the global score.

- A null I_{ij} implies that no interaction exists and, subsequently, v_i acts as the weights in a common WAM.

The associated aggregation function is given by:

$$P_{Ag}^{u} = \sum_{i=1}^{n} v_i\, p_i(x_{iu}) - \frac{1}{2}\sum_{\substack{\{i,j\}\\ i\neq j}} I_{ij}\,|\,p_i(x_{iu}) - p_j(x_{ju})| \quad (3)$$

With the property:

$$\left(v_i - \frac{1}{2}\sum_{i\neq j}|I_{ij}|\right) \geq 0, \forall i\, \in [1,n] \;\; and\;\; j \neq i \qquad (4)$$

However, it is important to note that this mathematical formulation is coherent with the performance values on an unipolar interval scale (on the universe [0,1]) [11]. In fact, these scales are not always appropriate, but there are not practical alternatives to the formulation referred adaptive to other types of scales [11, 22]. The performance values, used in this study, concern negative values as positive values superior to one, that is: the performance values are on an unlimited bipolar scale. So, to use this operator, it is necessary to convert the scale proposed by MACBETH (value function) in the unipolar scale required. Note that the MACBETH scale is an interval scale, so it is invariant under positive linear transformations, and therefore it can be transformed into another.

In order to determine the CI parameters, the DM has to compare fictitious FsU, which correspond to particular cases where all the performances are in the target and/or in the base levels. The objective is to make such comparisons more realistic and simpler. The number of fictitious FsU is intended to be equal to the sum of the CI parameters plus one (α, which is a coefficient necessary to guaranty that performance values are in the range [0,1]) [10]. In the case of two criteria, it involves the existence of 3 CI parameters (v_1, v_2 and I_{12}) and, thus, 4 fictitious FsU. These fictitious FsU should be FU(B,B), FU(B,T), FU(T,B) and FU(T,T), where FU(B,T) represents a fictitious FU with performances in the base and target levels in the first and second criteria, respectively. The next step consists in their ordination in terms of attractiveness and to judge semantically the difference of attractiveness between two consecutive ordered fictitious FsU. Note that these judgments are introduced in a matrix. After that, and with the help of equations (3) and (4), it is possible to construct an equation system, which permits the determination of criteria weights and interaction parameters. Consider that the DM says that: "$P_{Ag}^{(T,T)}$ is moderately preferred to $P_{Ag}^{(B,T)}$". This piece of information leads to write the following equation:

$$\begin{aligned} &P_{Ag}^{(T,T)} - P_{Ag}^{(B,T)} = 3\alpha = \\ &= [p_1(T_1)v_1 + p_2(T_2)v_2 - 0{,}5 \times |p_1(T_1) - p_2(T_2)v_2|] \qquad (5)\\ &-[p_1(B_1)v_1 + p_2(T_2)v_2 - 0{,}5 \times |p_1(B_1) - p_2(T_2)v_2|] \end{aligned}$$

Other equations are written in the same logical, originating an equation system. Let A be the matrix of the coefficients for the equation systems; V the column matrix of variables, namely the Shapley and interactions parameters and α; and e_i, the column vector where all the elements except ith is null. The equation system to resolve is given by:

$$A \times V = e_i \qquad (6)$$

The resolution of the system gives the values of the Shapley and interactions parameters and α. Then the aggregated performance (by area) can be determined by using equation (3).

3.4. Weighting areas

By the previous step, it is possible to measure the attractiveness at the second level of specification (subareas), where the first level corresponds to the criteria. To measure the attractiveness of FsU in terms of areas belonging to the 3rd level of specification, it is proposed to use an hierarchical additive model. Thus, the scores obtained in the subareas of the 2nd level are aggregated through the WAM operator, yielding a single score in each area (the third level). To obtain a value of overall performance for each state the procedure is identical: the scores obtained in each area are aggregated through the WAM operator, originating an overall performance value [19]. This model is proposed because it is quite simple and has wide acceptance in the literature [23].

In order to determine the weights from the subareas present in the second level of specification, it is proposed the swing weighting procedure. In a bottom-up approach, this procedure, firstly, considers the most elementary areas (2nd level) and then hierarchically areas superior to them, and so successively until reaching the top of the hierarchy. The description of this procedure can be seen in [23].

3.5. Construction of categorical performances

In order to easily identify the needs for corrective actions, the status of the FU in each subarea, area and in overall terms can be classified within a category scale. In this study it was used a five category scale: 'Very Weak', 'Weak', 'Acceptable', 'High' and 'Very High'. These designations are common to all areas, but their description varies from area (subarea) to area (subarea) and also in overall terms. These categorical performances were constructed using the method proposed in [19].

3.6. Portfolio Analysis

The second stage of the methodology consisted in the construction of the resource allocation model, in particular in its illustration by describing the steps required for the selection of the portfolio of actions which, more efficiently, improve the FsU's performance. A portfolio is a set of actions that can be funded by the PHCG. A portfolio is efficient when there is no other portfolio with a global benefit so great without spending more money. The selection of the portfolio should be performed with the DM and the first step to find every efficient portfolio is to determine the incremental benefit of each action in relation to the status quo (given by the difference between the FU's performance value improved by an action and the FU's performance value in the status quo). The mechanisms of aggregation consisted of the same applied in the previous section. Then, it was necessary to determine the cost associated with each action.

Later, taking into account the available resources, the optimization approach, which solves the knapsack problem, finds all efficient portfolios, according to the expressions:

$$\text{Maximize:} \quad \sum_{j=1}^{m} p_j l_j \qquad (7)$$
$$\text{subject to:} \quad \sum_{j=1}^{m} c_j l_j \leq B \qquad (8)$$
$$l_j \epsilon \{0,1\}, \qquad (9)$$
$$j = 1, \dots, t \qquad (10)$$

where: p_j is the incremental benefit of the action j, c_j is the cost of the action (in this study, is the number of monthly working hours of health professionals); l_j is a binary variable that takes the value 1 if an action is selected and 0 otherwise; and B is the budget, which in this case corresponds to the total amount of monthly working hours of health professionals. The maximization of (7) represents the knapsack problem.

Note that the mathematical formulation of knapsack covers other than the mentioned budget constraints. These constraints can be translated as: inclusion and/or exclusion of certain actions in the portfolio; dependencies between actions; simultaneous integration of two actions; and exclusive integration of actions.

There are several software programs that aim to support the selection of an efficient portfolio. In this paper, we propose the PROBE since it allows, in an innovative way, to visualize the efficient frontier as well as take into account synergies and constraints between actions [3].

4. Application of the Framework

4.1. Model structuring

The final model is composed of 3 areas, each one integrating a coherent and an exhaustive set of non-redundant criteria (23 in total) selected with the Executive Director (ED) and the Clinic Council President of the South West PHCG (see figure 2).

The initial phase of structuring the criteria showed the existence of three main areas of concern: *Access*, *Performance of care* and *Efficiency*, which were further explored in a top down approach. Starting by the first branch, Access, the DM explained that this concern was mostly related to the use of family planning services, general practice and home visits. A total of 5 criteria were identified in this branch. The second branch, *Performance of care*, involves aspects of *Adult Health*, *Maternal Health* and *Child Health*, which correspond to the three subareas in this branch. In brief terms, this area includes criteria associated with prevention, care and accompaniment of diseases. This area covers 16 criteria. Finally, in the *Efficiency*, it is important to evaluate the policy of medicines and means of diagnosis and therapeutics prescription. Besides the inherent financial component, this area indirectly infers about the proper fulfillment of the clinical guidelines for relevant diseases in the PHC level. It is important to note that the majority of the considered criteria correspond to the indicators present in the contracting process. The full structure of the value tree is represented in Figure 2.

After defining the value tree, it was necessary to define a descriptor to each criterion. Given the nature of the elicited criteria, all of the descriptors were quantitative and described by a mathematical formulation. However, it is important to note that, in addition to the two reference levels assigned in each criteria (base and target), it was assigned three performance levels to the majority of the descriptors, with the same meaning (plausible minimum, desirable and plausible maximum). The definition of these five performance levels are in table 1.

Table 1: Definition of the five performance levels present in the majority of the considered criteria.

Performance levels	Definition
Plausible Minimum	Represents the plausible worst level.
Base	Corresponds to the average of the realized values by the PHCU, in December 2010.
Target	Corresponds to the contracted value with the HRA (Health Regional Administration) or, in its absence, to the average of the values contracted in 2011 in the HFU and PHCU.
Desirable	Corresponds to the average of the values contracted in 2011 only in the FHU.
Plausible Maximum	Represents the plausible best level.

The criteria which constitute exceptions to these five level descriptors involve:

- The *Diabetes Prevalence* and *Hypertension Prevalence* criteria. Note that these criteria do not integrate the set of indicators present in the contracting process. Hence, it was defined a four level descriptor (without desirable level) where the levels' definition differ from the previous mentioned. Thus, it was discussed what would be a plausible maximum/minimum, target and base levels, in these criteria;
- The *Cost of invoiced medical bills* and *Cost of invoiced means of diagnosis and therapeutics* criteria. In these criteria the difference is on the plausible limits, that is: the plausible maximum represents the plausible worst level;

4.2. Value measurement

Like explained in the section 3.2., the construction of value functions was performed by means of the MACBETH judgments matrix. This activity was carried out in a decision conference with the Executive Director and the Clinic Council President of the South West PHCG, the coordinator of the PHCU of Torres Vedras and two doctors. When the value scales for all criteria were built, the next step was the elicitation of the criteria and

Figure 2: Representation of the value tree. The blue-shadowed nodes, in the final branches of the tree, represent criteria.

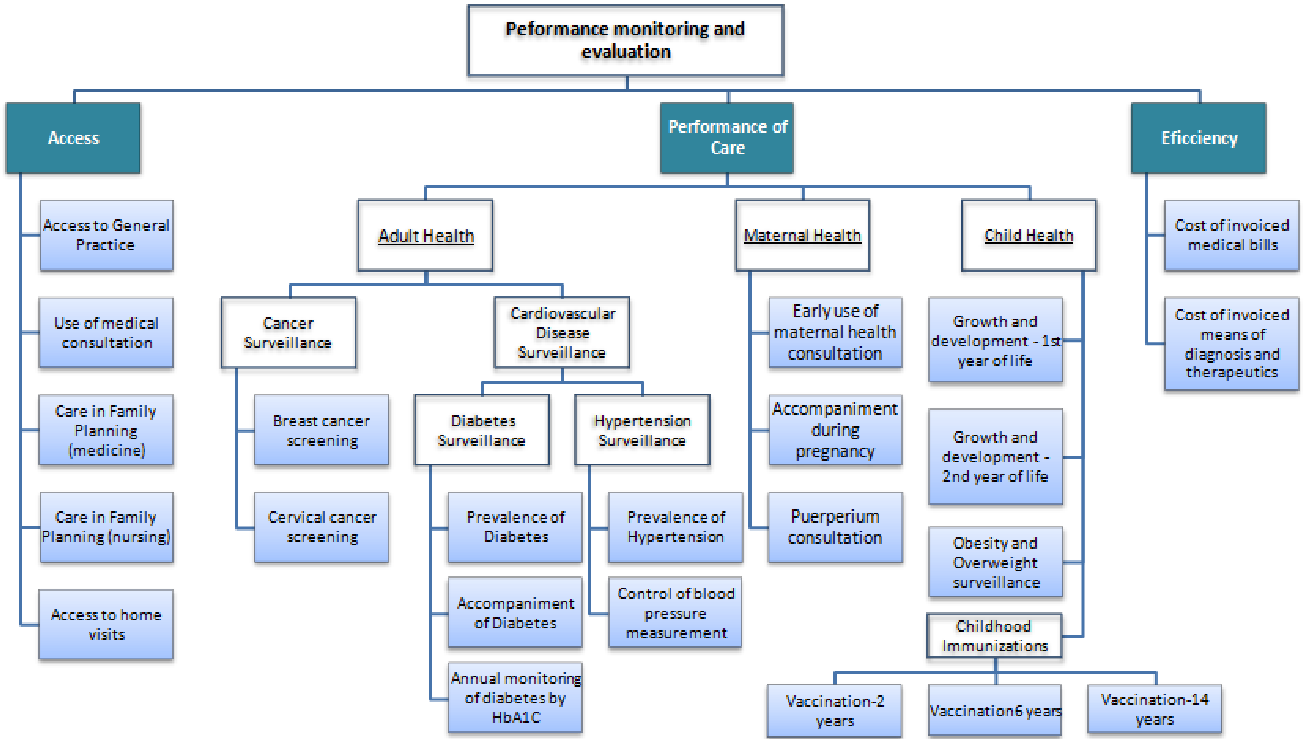


Table 2: Obtained CI parameters.

Subarea/ Area	Criteria	Shapley Parameters	Interaction parameters
Cancer Surveillance	*1: Breast cancer surveillance*	0,512	$I_{12} = 0{,}14$
	2: Cervical cancer surveillance	0,488	
Diabetes Surveillance	*1: Prevalence of diabetes*	0,370	$I_{12} = -0{,}1138$
	2: Accompaniment of diabetes	0,170	$I_{13} = -0{,}20279$
	3: Annual monitoring of diabetes by HbA1C	0,460	$I_{23} = -0{,}08854$
Hypertension Surveillance	*1: Prevalence of hypertension*	0,560	$I_{12} = 0{,}42$
	2: Control of blood pressure measurement	0,440	
Maternal Health	*1: Early use of maternal health consultation*	0,689	$I_{12} = 6{,}8 \times 10^{-15}$
	2: Accompaniment during pregnancy.	0,235	$I_{13} = -2{,}5 \times 10^{-16}$
	3: Puerperium consultation	0,076	$I_{23} = -1{,}9 \times 10^{-16}$
Efficiency	*1: Cost of invoiced medical bills*	0,587	$I_{12} = 0{,}092$
	2: Cost of invoiced means of diagnosis and therapeutics	0,413	

area weights, which was also built in a decision conference with the same people mentioned in the previous activity.

It was adopted the MACBETH procedure to determine the criteria weights present in the *Access* area and in the *Child Health* subarea. This adoption results from the number of criteria present in each of these areas (5 and 6, respectively), where the 2-additive CI procedure would be impractical. Note that, before this adoption, it was verified the ‘difference independence’ in these criteria. The 2-additive CI procedure was applied to the rest of criteria, being determined the corresponding Shapley parameters and interactions parameters (see table 2).

From the preference ordering of the fictitious FsU and the strength of that preference order given by the DM, it was concluded that there were interactions, some of which cannot be ignored, between criteria, as it can be seen in table 2. The obtained positive interaction values in the criteria present in *Cancer Surveillance* subarea, show that for the DM, a FU must be successful from both the criteria in order to be considered successful (complementary phenomenon). The same was observed for the criteria present in *Hypertension Surveillance* and *Efficiency* areas.

In order to determine the weights from the subareas and areas, it was used the swing weighting procedure. In a bottom-up approach, firstly, it was considered the most elementary areas – subareas - (2^{nd} level) and then hierarchically areas superior to them, and so successively until reaching the top of the hierarchy. It is important to note some difficulties felt by the DM to quantitatively judge the difference of attractiveness between two stimuli.

4.3. Construction of categorical performances

The construction of the five category scale to classify the FsU's performance was realized by the bottom-up procedure described in the study [19]. Due to time constraints, this activity was only realized with the ED of the South West PHCG.

4.4. Evaluation of FU's performance

After having constructed the evaluation model, it was necessary to perform its implementation. In this sense, this model was applied on three Familiar Health Units (FHsU) – Arandis, Dom Jordão and Gama - and one Personalized Health Care Unit (PHCU) – Alenquer. Note that these FsU have distinct characteristics that must be considered, since they can influence its performance at the criteria level. These characteristics include the percentages of woman and elder people registered in each FU. In fact, high percentages imply, for example, better results on the use of general practice. In that sense it was constructed a fictitious FU to serve as a comparison base for the evaluation of FsU's performance. Since the variations were insignificant, it was not necessary to make approaches.

4.5. Portfolio Analysis

Due to time constraints, it was only shown, with the help of an example, the necessary steps for selecting the portfolio of actions with potential to improve the FsU's performance. To that end, it was, firstly, considered three hypothetical actions to improve the PHCU of Alenquer: (1) sexual health promotion: family planning; (2) breast and cervical cancers awareness campaigns; and (3) motherhood project. Then, it was specify the cost of each action (in terms of the number of monthly working hours of health professionals) and it was determined the incremental benefit of each action in relation to the status quo. After that, with the help of PROBE software, it was possible to determine every efficient portfolio.

Figure 3: States of the four FsU under analysis, in each area (*Access* (a), *Performance Care* (b) and *Efficiency* (c)) and in overall terms (d). Note that the red color represents a very weak state; orange, a weak state; yellow, a acceptable state; light green, a high state; and dark green, a very high state.

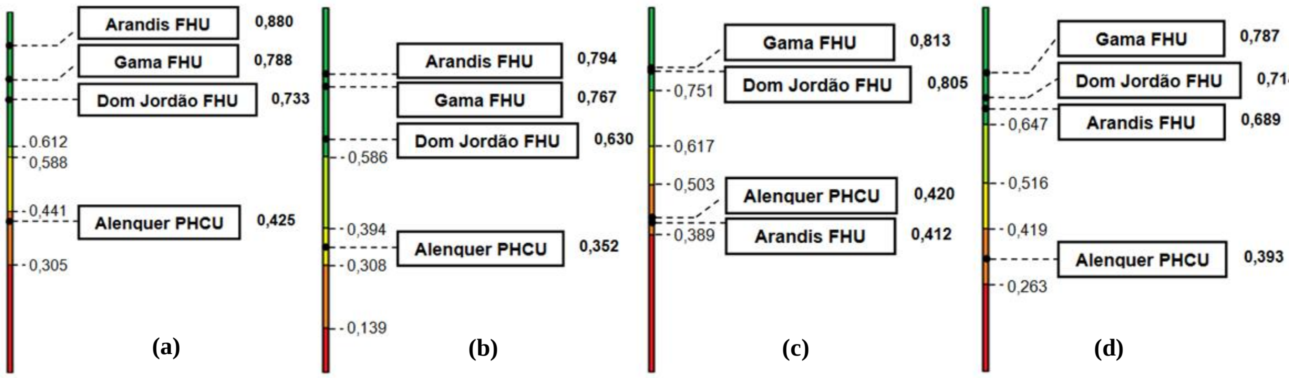
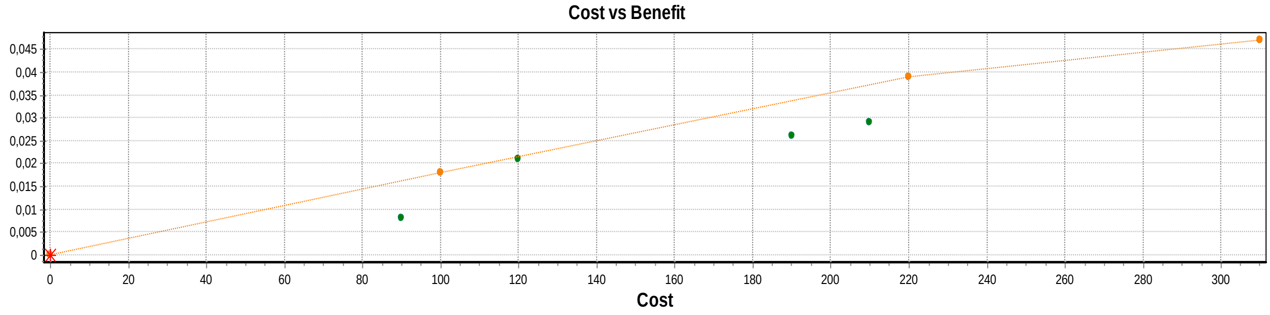


Table 3: States of the four FsU under analysis, in each subarea from the *Performance Care*. Note that the red color represents a very weak state; orange, a weak state; yellow, a acceptable state; light green, a high state; and dark green, a very high state.

	Subarea	Alenquer PHCU	Arandis FHU	Dom Jordão FHU	Gama FHU
PERFORMACE OF CARE	*Diabetes Surveillance*	0,244	0,760	0,684	0,709
	Hypertension Surveillance	0,204	0,683	0,514	0,526
	Cardiovascular disease Surveillance	0,224	0,722	0,599	0,618
	Cancer Surveillance	0,188	0,752	0,631	0,756
	Adult Health	0,208	0,735	0,613	0,678
	Maternal Health	0,487	0,818	0,746	0,780
	Child Health	0,432	0,852	0,563	0,875

Figure 4: Efficient portfolios obtained by the three hypothetical actions.



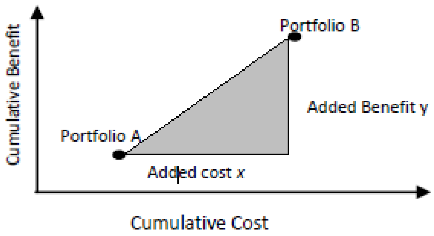
5. Results

Due to time constraints, the facilitator met only with the ED to define the performance profiles corresponding to each FU. These profiles are based on values observed in December 2010, since there is no later available information. The value scores obtained by the FsU in each subarea are shown in table 3 and in each area and in overall terms are shown in figure 3, with each score assigned to a category of accomplishment. As can be seen, the FHsU are in an overall very high state and the PHCU is in an overall acceptable state. It is worthwhile noting that the analysis made at the level of the areas and subareas, results in the same conclusion. In fact the FHsU's performances values are more attractive than the PHCU's performances values in the most levels of specification. This is in line with expectations, in the sense that the FHsU have higher dynamic in their professional teams and a higher degree of organization autonomy than the PHCsU. Other aspect that could be related is the fact that the FHsU was the first FU to be on contracting processes. In addition, PHCsU are not rewarded as a result of good practices, which may arise the demotivation for good results.

Figure 4 shows all portfolios that can be formed with the three hypothetical actions mentioned in the previous section to improve the PHCU's performance of Alenquer. The efficient portfolio are shown as green dots, whereas the orange dashed line links only convex efficient portfolios. To aid the DM in comparing the different portfolio it is helpful to use the cost-benefit triangle, as exemplified in figure 5. This figure illustrates that the shift from portfolio A to portfolio B has an added cost, given by the triangle's base, and an added gain, given by the triangle's height. Hence, comparing the size of the triangle's base and height facilitated the DM's choice of shifting from portfolio A to portfolio B or not. The notion of value-for-money of each portfolio is associated with the slope of the triangle. Thus, the cost-benefit triangle is useful to aid the decision makers to compare the several portfolios amongst themselves.

Note that this analysis was made by the use of PROBE. Besides budget constraints, this software also allows the user to add synergies among actions and constraints in the actions.

Figure 5: Cost-benefit triangle [24].



6. Discussion

The methodology constructed intended to support South West PHCG, by adopting a multicriteria methodology, in a rational and transparent way, throughout a rich interactive consulting path. The methodology applied was based on the extension of MACBETH to the 2-additive CI operators that are more general than the weighted mean. In fact, these operators permit the modeling of the interactions between criteria, which dictate the exploratory nature of this work.

The constructed model was applied to a set of four FsU: three FHsU (Arandis, Dom Jordão and Gama) and one PHCU (Alenquer). The FsU's performances values are in line with expectations, since FHsU's performances values are more attractive than the PHCU's performances values in the most levels of specification.

Although we have not implemented the procedure for prioritization of actions in real terms, it was noted the importance of the multicriteria model developed in the previous step to determine the incremental benefit of each action as well as the steps necessary for a successful selection of the efficient portfolio.

Despite the consistent results, it is important to note some of the limitations associated with the constructed model. One of them is related to the adoption of the 2-additive CI procedure. Despite its richness on the interaction modeling, this procedure becomes impractical when applied with more than four criteria. Thus, it was not used in *Access* area and *Child Health* subarea, which might lead to inaccurate results. Other limitation consists in the used swing weights procedure to the area weights determination. In fact, the DM showed some difficulties to elicit quantitative judgments. Once the results of application of the model were consistent with expectations, it is considered that the limitations of the model did not harm the development of the methodology.

However, it is important to note that for a proper implementation of the performance monitoring and evaluation model, it is necessary its periodic update, with regard to descriptors of impacts, value functions, etc. For a proper update, it is suggested its occurrence in decision conferences with all the decision makers. The sharing of knowledge among all the DM is the key for building proper management tools.

For further research, it is important to note the unadvisable application of the 2-additive IC procedure in more than four criteria. To enjoy the best of the potential of this procedure, it is suggested its mathematical formulation review, in order to:

- Be applicable to unlimited bipolar scales, in order to make the process easier and less time consuming, avoiding the changes of scales. Although there is already a mathematical formulation adapted for limited bipolar scales ([-1,1]), this leads to a very complex process, having no examples of application in real cases [11].
- Integrate more semantic judgments between pairs of hypothetical FsU in order to make the model more robust.

Concluding, the results support the PHCG in monitoring and evaluating the performance of the FU as well as the prioritization of the corrective actions. It was also concluded that the extension of MACBETH to the CI operators is still in the first steps to make its application easier and more integrated. Despite being an innovation in the health literature, this methodology represents a requisite model, “in the sense that everything required to solve the problem is either included in the model or can be simulated in it” [25].

References

1. OMS and UNICEF, *Declaration of Alma-Ata*, in *International Conference on primary Health Care*. 1978: Alma-Ata.
2. Missão para os Cuidados de Saúde Primários. *Cuidados de Saúde Primários - Metodologia de Contratualização*. 2010 [cited Dezembro de 2010]; Available from: http://mcsp.min-saude.pt/.
3. Lourenço J. C. and Bana e Costa C. A., *PROBE - A Multicriteria Decision Support System for Portfolio Robustness Evaluation*, in *Working Paper LSEOR 09.108*. 2009, London School of Economics: London.
4. Smith P., et al., *Performance measurement for health system improvement: experiences, challenges and prospects*. 2008: World Health Organization.
5. Smith P., *Performance measurement in health care: history, challenges and prospects.* Public Money & Management, 2005. **25**(4): p. 213-220.
6. Bourne M., et al., *Designing, implementing and updating performance measurement systems.* International Journal of Operations & Production Management, 2000. **20**(7): p. 754-771.
7. Neely A., Gregory M., and Platts K., *Performance measurement suystem design: A literature review and research agenda.* International Journal of Operations & Production Management, 2005. **25**(12): p. 1228-1263.
8. Santos S., Belton V., and Howick S., *Adding Value to Performance Measurement by Using Systems Dynamics and Multicriteria Analysis.* Management Science: Theory, Method & Practice, 2001.
9. Berrah L. and Clivillé V., *Towards an aggregation performance measurement system model in a supply chain context.* Computers in Industry, 2007. **58**(7): p. 709-719.
10. Berrah L., Clivillé V., and Mauris G., *Quantitative expression and aggregation of performance measurements based on the MACBETH multi-criteria method.* International Journal of Production Economics, 2007. **105**(1): p. 171-189.
11. Berrah L., Mauris G., and Vernadat F., *Industrial performance measurement: an approach based on the aggregation of unipolar or bipolar expression.* International Journal of Production Research, 2006. **44**(18-19): p. 4145-4158.
12. Clivillé V., Mauris G., and Berrah L., *A Quantified Industrial Performance Measurement System Based on a Choquet Fuzzy Integral*, in *IEEE International Conference on Fuzzy Systems*. 2006: Vancouver, BC.
13. Bana e Costa C. A., Fernandes T. G., and Correia P., *Prioritisation of public investments in social infrastructures using multicriteria value analysis and decision conferencing: a case study.* International Transactions in Operational Research, 2006. **13**(4): p. 279-297.
14. Mitton C., *Priority setting for decision makers: using health economics in practice.* European Jornal of Health Economics, 2002. **4**: p. 240-243.
15. Mitton C. and Donaldson C., *Tools of the trade: a comparative analysis of approaches to priority setting in health care.* Health Services Management Research, 2003. **16**(2): p. 96-105.
16. Bana e Costa C. A. and Beinat E., *Model-structuring in public decision-aiding.* Working Paper LSE OR 05.79 - London School of Economics, 2005.
17. Bana e Costa C. A., et al., *Decision Support Systems in action: Integrated application in a multicriteria decision aid process.* European Journal of Operational Research, 1999. **113**(2): p. 315-335.
18. Bana e Costa C. A., et al., *Facilitating bid evaluation in public call for tenders: a socio-technical approach.* The International Journal of Management Science, 2002. **30**(3): p. 227-242.
19. Bana e Costa C. A., Carnero M. C., and Oliveira M. D., *A Multi-Criteria Model for Auditing a Predictive Maintenance Programme.* European Jornal of Operational Research, 10.1016/j.ejor.2011.09.019.
20. Bana e Costa C. A., et al., *Multicriteria approach for strategic town planning: the case of Barcelos*, in *Aiding decisions with Multiple Criteria: Essays in Honour of Bernard Roy, D. Bouyssou, E. Jacquet-Lagrèze, P. Perny, R. Slowinsky, D. Vanderpooten, P. Vincke*. 2002, International Series in Operations Research & Management Science. p. 429-456.
21. Bana e Costa C. A., Ferreira J. A. A., and Vansnick J.-C., *Avaliação Multicritério de Propostas: O Caso de uma Nova Linha do Metropolitano de Lisboa.* Revista de Transportes e Tecnologia, 1995.
22. Grabisch M. and Labreuche C., *A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid.* Annals of Operations Research, 2010. **175**(1): p. 247-290.
23. Goodwin P. and Wright G., *Decision Analysis for Management Judgement*. Jonh Wiley & Sons. 1991.
24. Oliveira M. D., et al. *Prioritizing Health Care Interventions: A Multicriteria Resource Allocation Model to Inform the Choice of Community Care Programmes*. in *Proceedings of the Conference Operational Research applied to Health Services: Operational Research Informing National Helath Policy (Eds.: Harper P., Knight V., Vieira I., Williams J.)*. 2011. Cardiff, United Kingdom.
25. Bana e Costa C. A. and Oliveira R. C., *Assigning priorities for maintenance, repair and refurbishment in managing a municipal housing stock.* European Jornal of Operational Research, 2002. **138**: p. 380-391.