

This is a PDF file of an unedited manuscript that has been accepted for publication in OR Spectrum. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

An application of Value-Based DEA to identify the best practices in primary health care

MC Gouveia

IPC-ISCAC, Quinta Agrícola, Bencanta, 3040-316 Coimbra
INESC Coimbra, Rua Antero de Quental, 199; 3000-033 Coimbra
mgouveia@iscac.pt

LC Dias

Faculdade de Economia, Universidade de Coimbra, Av. Dias da Silva 165, 3004-512 Coimbra
INESC Coimbra, Rua Antero de Quental, 199; 3000-033 Coimbra
lmcdias@fe.uc.pt

CH Antunes

DEEC-FCT Universidade de Coimbra, Polo II, 3030 Coimbra
INESC Coimbra, Rua Antero de Quental, 199; 3000-033 Coimbra
cantunes@inescc.pt

MA Mota

CEO

ACES Baixo Mondego I, Rua Luís de Camões, nº 150; 3000-251 Coimbra
mariaaugustamota@gmail.com

EM Duarte

Member of the Clinical Council

ACES Baixo Mondego I, Rua Luís de Camões, nº 150; 3000-251 Coimbra
emrduarte@gmail.com

EM Tenreiro

Coordinator of the Management Support Unit

ACES Baixo Mondego I, Rua Luís de Camões, nº 150; 3000-251 Coimbra
eduardo.tenreiro@gmail.com

ABSTRACT. In 2005 a new phase for Primary Health Care (PHC) in Portugal began, leading to the voluntary and independent creation of Family Health Units. The impact of this reform is now being witnessed, which justifies the need to evaluate its results as far as best practices in PHC are concerned. This work uses a Value-Based Data Envelopment Analysis (DEA) method, which combines an additive DEA model with Multi-Criteria Decision Analysis (MCDA) concepts, to allow incorporating managerial preferences in the performance assessment of 12 health centres in a Portuguese region. Two models are presented for evaluating access to PHC, which resulted from discussions with a group of decision makers. This analysis helped the decision makers in the identification of best practices, sources of inefficiency and gaps relatively to the best practices, thus leading to shape opportunities for improvement.

KEYWORDS. Data Envelopment Analysis; Multi-Criteria Analysis; Primary Health Care; Uncertainty.

1. Introduction

The Primary Health Care (PHC) system in Portugal is considered of crucial importance in the National Health Service (NHS). A major concern of policy makers is the effective delivery of health care services [1], pursuing objectives of equity, access, efficiency and quality of care.

Several studies of efficiency in provision of PHC exist, especially in the United Kingdom. Most of these studies tend to rely on the use of performance indicators (PIs) to measure health service delivery and quality of care [2-7], but the different PIs are analysed individually. Thanassoulis *et al.* [8] and Giuffrida *et al.* [9] pointed out the disadvantages of using PIs in performance evaluation and proposed alternative techniques for this purpose, such as DEA. Originally developed by Charnes *et al.* [10], DEA is a method for evaluating the relative efficiency of Decision Making Units (DMUs). It emerged with the objective of evaluating performance in the public sector from the standpoint of productive efficiency, in which there was neither a "market" to pick the most efficient organizational units, nor a regulatory model that would guarantee increased efficiency [11]. Each DMU uses multiple inputs to produce multiple outputs. DEA models seek to determine which of n DMUs form the *efficient frontier* (or *envelopment surface*) in the Pareto-Koopmans sense: a DMU lies on the efficient frontier if and only if no other observed DMU improves any input or output without worsening some of other inputs or outputs [12]. By identifying the DMUs with the best practices, DEA models construct an empirical production frontier. Since mid-1980s this methodology is regularly used to evaluate the efficiency of health care services (see [13-15]). Some studies of PHC efficiency use DEA techniques (e.g. [16-22]), but relatively few studies ([15, 23, 24, 25, 26]) include measures for quality of care/service in the efficiency assessment.

Several studies have been undertaken to evaluate efficiency in different areas using DEA, but only a few incorporate managerial preferences in the analysis (see [27]). The mathematical structure of classical DEA models allows flexibility in the choice of inputs and outputs weights. Therefore, the weights chosen in the solution may be null to important variables, and logical relationships between the variables can be disregarded in a way incompatible with the decision maker's (DM) judgment. Hence, the results may not be credible and the DM's confidence in the methodology may be compromised.

DEA can be combined with Multiple Criteria Decision Making approaches as a means of incorporating managerial preferences in the analysis. For instance, Golany [28] combined Multiple Objective Linear Programming and DEA by developing target-setting models in an interactive way with DMs; Joro *et al.* [29] developed a detailed analysis of structural relationships between the two methodologies and Halme *et al.* [30] showed that Multiple Objective Linear Programming models can be used to incorporate DM's preferences into DEA when looking for his/her most preferred solution; Mavrotas and Trifillis [31] proposed a method combining DEA and Multi-Attribute Value Theory, in which each alternative uses its most favourable criteria weights and value functions.

This work exploits the links between DEA and MCDA, applied in a real-world situation in which the DMs sought to evaluate DMUs taking into account their managerial preferences. It uses the Value-Based DEA method [32], which has proven useful for this type of evaluation in previous studies [33, 34]. This

method builds on the additive DEA model with oriented projections [35], using concepts developed in the field of MCDA under imprecise information [36, 37]. Input and output factors are converted into utility functions according to the preference information provided by a group of decision makers (GDM). Additive utility functions are used to aggregate the utilities associated with each alternative, based on Multi-Attribute Utility Theory (MAUT) [38]. DMUs choose the weights associated with these functions, subject to restrictions reflecting GDM's preferences. In Value-Based DEA, the efficiency measure assigned to each DMU has an intuitive meaning: the "min-max regret" (loss of utility) [32].

The purpose of the present work is to evaluate the efficiency of PHC provided by the 12 units monitored by "Group of Health Centres" (*Agrupamentos de Centros de Saúde*, ACES) Baixo Mondego I, in the central region of Portugal. The study was developed in 2011 for the ACES Baixo Mondego I with 2010 data, concerning 203,657 registered users. Among these users, 19.5% are over 65 years old and 22.8% are less than 25 years old [39]. The worth of this work to ACES Baixo Mondego I lies on the ability to use multiple factors in an integrated model and the prospect of identifying best practices, sources of inefficiency and gaps to best practices, as well as contributing to design opportunities for improvement. This approach has revealed to be useful not only for the introduction of corrective measures, but also to inform decisions about the goals of PIs in the future.

In Portugal, there are few studies about performance evaluation of PHC using DEA. Amado and Santos [40] provide an empirical analysis of efficiency, equity and quality of the PHC in Portugal with the involvement of DMs. They discuss the results with health care professionals and their conviction is that engaging stakeholders in the evaluation process contributes to performance improvement in practice. More recently, Ferreira *et al.* [41] also evaluate the performance of PHC providers of a Portuguese region, taking into account both efficiency, equity and quality. In this study the DMUs are the 22 ACES belonging to Lisbon and Tagus Valley region and the authors show that by establishing a learning network the efficiency of ACES is likely to improve, with a better usage of the available resources while keeping quality and equity.

We suggest two models for performance assessment in PHC. These models are aimed at evaluating, in particular, the accessibility to PHC. From the perspective of the GDM it is important to investigate the percentage of patients who have a designated general practitioner and use the medical care they are entitled to. The factors included in the models resulted from discussions with the GDM, who are members of the ACES Baixo Mondego I, based on the Performance Plan [39] and data from 2010.

This introduction provided the interest and motivation of the study. Section 2 reviews the Value-Based DEA method [32] with the modifications to include the super-efficiency concept [42]. Section 3 briefly describes the Portuguese health care system. Section 4 presents the data, the models discussed with the GDM and the elicitation protocols used to obtain the GDM's preferences. In section 5 the model results are presented and analysed. Section 6 highlights the improvement proposals for the inefficient units and the new developments of the method. Concluding remarks are presented in section 7.

2. The Value-Based DEA method

The Value-Based DEA method [32] is a MAUT-based variant of the additive DEA model with oriented projections [35], able to incorporate imprecise information about the DM's preferences. MAUT is used to evaluate the utility of each alternative, taking into account the multiple criteria. First, the DM focuses his/her attention on one criterion at a time, defining a marginal (partial) utility function for each one. This requires that the original input and output scales are converted into utility scales. There are several techniques for questioning the DM in order to construct the utility functions compatible with his/her answers (see [43, 44]). Then, these functions are aggregated to yield a global utility. For this purpose, the Value-Based DEA method uses an additive model.

Let $u_c(a_j)$ denote the utility of an alternative a_j according to the c^{th} criterion ($c = 1, \dots, q$). The additive utility model aggregates these functions using q scale coefficients w_1, \dots, w_q as follows (see [38]):

$$U(a_j) = \sum_{c=1}^q w_c u_c(a_j).$$

The quantity $u_c(a_j)$ can be interpreted as a measure of alternative a_j 's performance on criterion c , and w_c the weight assigned to reflect the importance of the function u_c ($c = 1, \dots, q$). Because utility functions are unique up to positive affine transformations, it is usually assumed that both the global utility functions and marginal utility functions are scaled to lie between 0 and 1. The scale coefficients (weights) are also normalized such that $w_c \geq 0$, $\forall c = 1, \dots, q$ and $\sum_{c=1}^q w_c = 1$. Additional restrictions may be specified by the DM whenever he/she wants to prevent combinations of values for these coefficients that are not aligned with the organization's goals.

The Value-Based DEA model avoids the scale-dependence problem of the additive DEA model [32], since all the input and output measures are translated into utility units. Moreover, the weights used in the aggregation gain a specific meaning: they are the scale coefficients of the utility functions. The weights are variables to be set by each DMU in order to become the best DMU (if possible) or else to minimize the utility difference to the best DMU, rather than being fixed beforehand as in [35] or [45].

The classical decision theory offers some rules, such as min-max and min-max regret, which can be applied to situations where utility functions themselves are uncertain. The Value-Based DEA method uses the min-max regret rule to capture the utility loss associated with choosing an alternative instead of another. The idea is to determine for each alternative the potential regret associated with its choice, noting how much better could another alternative be. For a given weighting vector, the utility loss of choosing one alternative a_k instead of a different alternative a_j is given by: $R(a_k, a_j) = \max_{j \neq k} U(a_j) - U(a_k)$.

The maximum regret associated with a given alternative a_k , denoted by $R_{\max}(a_k)$, is the maximum loss of opportunity associated with choosing that alternative in an environment where w is interpreted as

$$\text{varying over some suitably specified feasible set: } R_{\max}(a_k) = \max_w \left[\max_{j \neq k} U(a_j) - U(a_k) \right].$$

The min-max regret optimal decision minimizes the worst-case loss with respect to possible realizations of the utility function, that is, alternatives with the minimum R_{\max} are preferred.

Considering that the alternatives are the DMUs to be evaluated according to q (with $q = m+p$) criteria, q utility functions must be defined such that the worst level has value 0 and the best level has value 1. Utility functions are monotonic: non-decreasing for outputs and non-increasing for inputs. Hence, after being converted into utility values all factors are treated as outputs to be maximized. For each alternative (DMU), according to the additive MAUT model, the utility obtained is $U(DMU_j) = \sum_{c=1}^q w_c u_c(DMU_j)$, where $w_c \geq 0$, $\forall c = 1, \dots, q$ and $\sum_{c=1}^q w_c = 1$. The scale coefficients w_1, \dots, w_q are the weights of the utility functions and reflect the utility trade-offs of the DM.

The Value-Based DEA method [32] can be extended to consider the super-efficiency concept introduced in DEA by Andersen and Petersen [46] in order to further discriminate efficient units [42]. For that purpose problem (1) is solved (Phase 1):

$$\begin{aligned}
& \min_{d_k, w} d_k \\
& \text{st.} \quad \sum_{c=1}^q w_c u_c(DMU_j) - \sum_{c=1}^q w_c u_c(DMU_k) \leq d_k, \quad j = 1, \dots, n, j \neq k \\
& \quad \sum_{c=1}^q w_c = 1, \\
& \quad w_c \geq 0, \quad c = 1, \dots, q
\end{aligned} \tag{1}$$

In this model, the DMU being evaluated (DMU k) is allowed to choose the weights vector $w = (w_1, \dots, w_q)$ that minimizes the score d_k^* . This score has a “min-max regret” (utility loss) intuitive meaning (the less, the better). It corresponds to the distance defined by the utility difference to the best of all alternatives (note that the best alternative will also depends on w), excluding itself from the reference set. If the optimal value d_k^* of the objective function in (1) is not positive, then the DMU k under evaluation is efficient (there is a weights vector for which it has the best utility, possibly with ties, among all DMUs); otherwise it is inefficient (no matter which weights vector is chosen, there is always a DMU with better utility). Using d_k^* we can assess the extent to which an efficient DMU may worsen its utility while remaining efficient. In case the DMU is inefficient, Phase 2 can be used to find an efficient target.

The details of this process are as follows:

Phase 1: Convert inputs and outputs into utility scales. Compute the efficiency measure, d_k^* , for each DMU, $k = 1, \dots, n$, and the corresponding weighting vector.

Phase 2: If $d_k^* \geq 0$ then solve the “weighted additive” problem (2), using the optimal weighting vector resulting from Phase 1, w_k^* , and determine the corresponding projected point of the DMU under evaluation.

$$\begin{aligned}
\min_{\lambda, s} \quad & z_k = -\sum_{c=1}^q \lambda_c^* s_c \\
\text{st.} \quad & \sum_{j=1, j \neq k}^n \lambda_j u_c(DMU_j) - s_c = u_c(DMU_k), \quad c=1, \dots, q \\
& \sum_{j=1}^n \lambda_j = 1, \\
& \lambda_j, s_c \geq 0, \quad j=1, \dots, k-1, k+1, \dots, n; c=1, \dots, q
\end{aligned} \tag{2}$$

The variables $\lambda_j, j=1, \dots, k-1, k+1, \dots, n$ define a convex combination of utility score vectors associated with the $n-1$ DMUs. The set of efficient DMUs (possibly a single one) defining the convex combination with $\lambda_j > 0$ are called the “peers” of DMU k under evaluation. The convex combination corresponds to a point on the efficient frontier that is better than DMU k by a difference of value of s_c (slack) on each criterion c .

3. The Portuguese Health Care Service – Primary Health Care

PHC units are main form of contact of the population with the health system. In 1979 the NHS has been created to provide national, universal, comprehensive and free access to health care, which should be guaranteed by the State. The NHS comprises five health administrations, distributed by the different regions of Portugal (Regional Health Authorities); they develop strategic guidelines; coordinate all aspects of health care provision; supervise management of hospitals and PHC; establish agreements and protocols with private bodies; and develop a long-term care network [47]. In the last 40 years the PHC has undergone successive reform attempts. The 1971-2004 period was problematical, with the passage of laws adopted in successive governments that limited the management autonomy of these structures. In 2005 a new phase in the PHC sector began with the creation of a taskforce to lead the voluntary and independent creation of Family Health Units, with annual individual and group work plans and performance targets [48].

The constitution of the ACES establishes its organization into functional units, namely in Family Health Units, Personalized Health Care Units (both join general practitioners, nurses and administrative staff), community care units, public health units and units of shared care resources. The ACES also provides for the existence of support units, such as the CEO (a general practitioner), the Community Council, the Clinical Council (includes a general practitioner as president, a public health physician, a nurse, and another health professional), the Management Support Unit and the Citizen Office. Although Family Health Units and Personalized Health Care Units are similar in terms of size, they are different because the first are regulated by specific legislation that includes rewards associated with achieving the goals contracted, whereas the second are bound by rules approved by the Clinical Council. Note that the administrative autonomy of ACES is closely related to the existence of a Contract-Program to be established between the ACES and the Regional Health Authorities.

Starting in 2010, CEOs/Clinical Councils with the support of Contract Departments of Regional

Health Authorities triggered the internal contracting with Family Health Units, which had begun earlier, and the same model was adapted to Personalized Health Care Units. Two planning instruments (the Performance Plan and the Contract-Program) were defined and implemented on pilot basis describing indicators and the internal contracting goals for all the Family Health Units and Personalized Health Care Units, as well as external contracting with ACES. The objectives and goals expressed in the Contract-Program do not necessarily correspond to all performance indicators presented in the Performance Plan. The contracts between the ACES and the Family Health Units and between ACES and Personalized Health Care Units (internal contracting) have in common 15 of the indicators from the Performance Plan. These common indicators are distributed as: 4 access indicators; 8 health care performance indicators; 1 indicator of perceived quality and 2 indicators of economic performance. In 2011, reference values were defined for each indicator. These must be updated annually in agreement with the contracting departments and Regional Health Authorities [39].

The contracting has instilled a greater sense of demand and accountability of providers and therefore has clear implications for health administration, that is, it has to be able to identify health needs, understand the differences between units, evaluate, hold responsible and differentiate good practices. Each ACES has a mandate for monitoring and improvement of efficiency of PHC services in order to make the best possible use of resources. They have only partial autonomy, since financial resources are under the control of the Regional Health Authorities, which are still responsible for planning and resource allocation. However, the ACES have the responsibility to develop their own organizational structure and management tools that better serve to define the clinical and technical governance systems [49].

4. Data and models developed including the DM's preferences

This section describes the performance assessment of 12 health units monitored by ACES Baixo Mondego I, in which there was interaction with a GDM: the CEO of the ACES, a member of the Clinical Council and the coordinator of the Management Support Unit. The units to be evaluated are Family Health Units and Personalized Health Care Units, since all of them are in the Contract-Program established between the ACES Baixo Mondego I and Centro Regional Health Authorities. The performance assessment of units is based on the Performance Plan, without distinguishing between Family Health Units and Personalized Health Care Units. Therefore, according to the GDM, the units under evaluation are considered an homogeneous set of DMUs.

The monitoring and follow-up of health units was based on a set of 15 indicators. The units could not select the indicators, but the goals to be achieved on these indicators were negotiated between the health unit and the CEO/Clinical Council, given the health objectives set by Regional Health Authorities and ACES. The data analysed are from 2010 and were extracted from various information systems in the medical office and in the accounting and management department.

The GDM needed a method with the possibility of using multiple factors in an integrated model, with an easy and meaningful process to express managerial preferences and the prospect of identifying

best practices, sources of inefficiency, gaps relatively to the best practices, and opportunities for improvement. The Value-Based DEA method was proposed to the GDM, which represents the ACES Baixo Mondego I, to assess the performance of the 12 DMUs (Family Health Units and Personalized Health Care Units) as providing a solution to these requirements. The GDM made clear in the first meeting that it did not intend to obtain a measure of overall performance. The purpose was to construct different models in accordance with the diverse causes of concerns, leading to a deeper understanding of the reasons of inefficiency of some units, and what corrective actions may needed to be undertaken.

The ACES Baixo Mondego I anticipated that Value-Based DEA would offer decision support that could contribute to design interventions adapted to the limitations of resources available and the specificities of the community that they serve, including managerial expertise into the analysis.

4.1. Definition of models

After the objectives of this study have been established, we began to define the inputs and outputs for each model taking into account the indicators of the main areas of the Performance Plan. We constructed four different models given the various possible insights resulting from the different perspectives of evaluation used. For brevity we present the four models but we only describe and discuss the results for models 1 and 2 in detail (see [50] for detailed results for the other models).

The GDM chose four factors to build model 1: number of doctors (x_{DOC}), number of nurses (x_{NUR}), and number of administrative, technical and other support professionals (x_{ADM}) as inputs, and number of medical consultations for registered patients (y_{CONS}) as an output. These factors stem from indicators defined in the Performance Plan to evaluate access: percentage of medical consultations to the user, overall utilization rate of consultations, medical home visits rate per 1000 subscribers, and nurse home visits rate per 1000 registered. The difference between the first and second indicator defined in the Performance Plan [39] is that the numerator of the first one is the full-face consultation by the own doctor and the numerator of the second one is the number of registered users with at least one medical consultation; the denominator of both indicators is the number of users registered in the health unit in the period. The GDM agreed that model 1 would address human resources needed for the attendance of registered patients in each unit, i.e., the access to medical consultation for the registered patients in 2010.

Since model 1 only takes into account the number of medical and non-medical staff, the CEO considered it would be interesting to build a new model where the inputs would be costs: total cost billed to the NHS with complementary means of diagnosis and treatment (x_{CMDT}); total medicine costs billed to the NHS (x_{MED}); total cost of human resources (x_{HR}) and medicine costs not billed to the NHS, clinical consumables and other costs (x_{OC}) and the only output would be y_{CONS} . The CEO wanted to know the impact on the results if in addition to total cost of human resources (included in x_{HR}) other costs associated with the unit were also considered. This is named henceforward model 2. It is noteworthy that x_{CMDT} and x_{MED} are the two indicators of economic performance of the Contract-Program. To these the

GDM added x_{HR} and x_{OC} .

To complete the study for the ACES Baixo Mondego I, two other models were built to examine perspectives other than access. In a third model the inputs reflect the costs of providing primary health care (the same of model 2), but the outputs are: number of family planning medical consultations; number of maternity health consultations; number of child (aged 0-18) health consultations; number of adult (aged 19-64) health consultation; number of elderly (aged more 65) health consultations and satisfaction with the quality of service. A fourth model links the four areas mentioned in the Performance Plan resulting in nine factors. The inputs are economic indicators of the Performance Plan (the same of model 2) and the outputs are: number of home visits by the doctor; number of doctor consultations in the health unit; number of home visits by the nurse; number of nurse consultations in the health unit and satisfaction with the quality of service. The latter model combines the multiple axes of evaluation, and also reflect the entire medical, nursing and administrative work (for details see [50]).

These inputs and outputs have been widely discussed with the GDM, in accordance with the indicators established in the Performance Plan and are often considered in this type of DEA studies [40].

4.2. Elicitation of utility functions

There are several techniques to elicit preference information from a DM/GDM to build utility functions consistent with their answers (see [43, 44]). In the Value-Based DEA method the purpose of factor conversion into a utility scale (nonlinear utility functions) is to reflect the preferences of the DM/GDM, taking into account the generalization of the DEA methodology presented by Cook and Zhu [51] that incorporates piecewise linear functions of factors. As referred to in section 2, the utility functions are defined for each criterion by converting its range of original performances into the range of utilities [0,1]. Utility levels 0 and 1 are anchor points defined by the worst and best performances on each criterion and the numbers on the scale do not have an absolute meaning. To construct the utility functions we elicit the difference in the DMU's value that corresponds to improvements in inputs and outputs, rather than the utility of having these inputs available or outputs produced. In this way we do not speak of absolute utilities, but rather relative utilities.

The elicitation protocol can be based on comparing the value of increasing an output (or decreasing an input) from a to b versus increasing the same output (or decreasing the same input) from a' to b' , all other performance levels being equal, and asking the DM/GDM to adjust one of these four figures such that the increase of value would be approximately equal. This is always possible assuming the functions are continuous and monotonic. For example, considering x_{DOC} the question posed to GDM was: "Is it more valuable to decrease the number of doctors from 20 to 10 or from 10 to 4, all the other performances being equal?" The answer was: "it is more valuable to decrease from 20 to 10." Then an adjustment has been made and the question was reformulated as: "Is it more valuable to decrease the number of doctors from 20 to 9 or from 9 to 4, all the other performances being equal?" The answer was: "the value is the same". This means that $u_{DOC}(4)-u_{DOC}(9)=u_{DOC}(9)-u_{DOC}(20)$, i.e., $u_{DOC}(9)=(u_{DOC}(4)+u_{DOC}(20))/2$. The same

procedure was used to dichotomize the intervals of value [4,9] and [9,20].

The GDM was available to answer questions about the differences of value between the performance levels in each factor. We used a piecewise linear approximation to represent the utility functions for most of the factors, and we used other functions (such as logarithmic functions) when the GDM's answers could be adjusted to predefined curves.

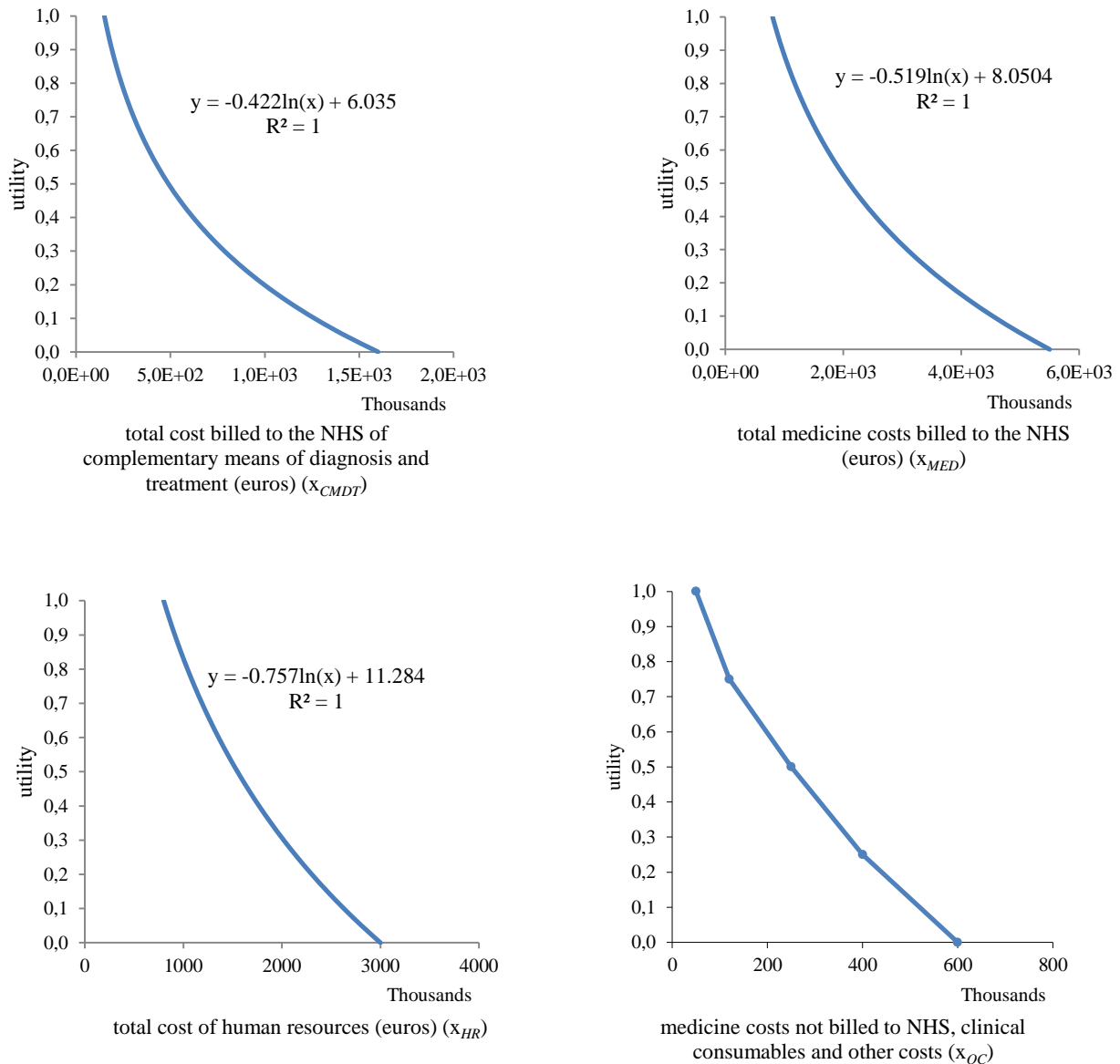


Figure 1. The utility functions of costs elicited.

The elicited ranges were chosen to include the observed performance ranges plus or minus the highest tolerance value considered (in this case $\delta = 20\%$ to account for uncertainty).

Table 1 summarizes the performance levels corresponding to value 0.25, 0.5, and 0.75 (resulting from this dialogue), such that an improvement from level 0 to level 0.25 corresponds to the same value as an improvement from level 0.25 to 0.5, etc. In Figure 1, the utility functions for the factors x_{CMDT} , x_{MED} ,

x_{HR} and x_{OC} are displayed. For example, the medicine costs not billed to the NHS, clinical consumables and other costs (x_{OC}) is a piecewise linear function where: $u_{OC}(600000) - u_{OC}(400000) = u_{OC}(400000) - u_{OC}(250000) = u_{OC}(250000) - u_{OC}(120000) = u_{OC}(120000) - u_{OC}(50000)$, all other performance levels being equal. For x_{CMDT} , x_{MED} and x_{HR} , the utility functions were obtained by fitting a logarithmic function to match as well as possible the answers of the GDM. The other inputs and the output utility functions of model 1 (x_{DOC} , x_{NUR} , x_{ADM} and y_{CONS}) are also piecewise linear utility functions.

Table 1. Summary of the performance corresponding to different (and equally spaced) utility levels for the factors in model 1.

Utility	Inputs			Output
	x_{DOC}	x_{NUR}	x_{ADM}	y_{CONS}
0	20	20	15	4000
0.25	12	12	10	12000
0.5	9	8	7	18000
0.75	6	5	5	22000
1	4	4	3	25000

4.3. Elicitation of weight restrictions

In the Value-Based DEA method the DMU under evaluation is free to choose the scale coefficients (weights) of the marginal utility functions, so as to become the best DMU (if possible) or else to minimize the utility difference to the best DMU, that is, to obtain the best possible efficiency score. In additive MAUT models all utility functions are added (adjusted by scaling coefficients) regardless of their nature (costs or benefits, monetary or non-monetary). These models add up utility value units (the translation of inputs or outputs into utility values); they do not add up the input and output scales [38, 43]. However, some factors may be disregarded from the assessment because DMUs can assign zero weights to some factors, in a way incompatible with the GDM's preferences. So, it is necessary to consider weight restrictions, once these reflect better the organizational objectives and therefore ensure meaningful results closer to what the GDM considers to be the best practices. Several approaches exist to define appropriate weight restrictions, which can be a challenging task [52, 53].

In the Value-Based DEA method [32] the scale coefficients (weights) reflect possible trade-offs of utility between different factors (see [38]). The assignment of values to the scale coefficients requires a series of judgments by the DM. Direct rating techniques should be avoided, because the value of these coefficients does not reflect the intuitive notion that the DM has of the importance of each criterion. Rather, they are strongly dependent on the performances chosen to represent the levels 0 and 1 on the utility scale. In MCDA there are several valid protocols to elicit the weight restrictions derived from DM's preferences [43, 44]. In this case the swings technique is simple and clear to the DM.

The swing method begins by constructing two extreme hypotheses, P_0 and P_1 , the first one displaying the worst performance (having utility 0) in all criteria scales and the second one the corresponding best performance (having utility 1). The preference elicitation protocol consists in querying the DM to look at the potential gains from moving from P_0 to P_1 on each criterion and then to decide which one of the

criteria he/she prefers to shift to hypothesis P_1 . Suppose that the transition from P_0 to P_1 on that criterion is worth 100 units on a hypothetical scale. Then, the DM is asked to give a value (<100) to the second criterion moved to P_1 , then to the third criterion and so on, until the last criterion is moved to P_1 .

The procedure used in this work was first to obtain a ranking of weights and then to establish a limit to the ratio between the weights ranked first and last to avoid null weights. Let W denote the set of weight vectors compatible with the elicited ranking and limit.

To include the weight restrictions, Phase 1 (see section 2) has the same formulation as (3) but adding the restriction $(w_1, \dots, w_q) \in W$. This change in Phase 1 implies changing the formulation of the problem to be solved in Phase 2. It is now necessary to allow slacks to have negative values; otherwise it might not be possible to keep the optimal utility difference d_k^* derived from formulation (3) including the weight restrictions. Note that if slacks are imposed to be positive with the introduction of weight restrictions, then the projection of inefficient units has to be made out of the production possibility set.

5. Results

As previously presented, the GDM proposed two models (1 and 2) to evaluate efficiency in terms of access. Table 2 and Table 3 indicate the original data and the corresponding utilities.

Table 4 gives the results for model 1 using formulation (3) which allows to rank the DMUs in terms of optimal utility loss d^* : DMU 7 \succ DMU 5 \succ DMU 4 \succ DMU 3 \succ DMU 9 \succ DMU 11 \succ DMU 8 \succ DMU 6 \succ DMU 1 \succ DMU 10 \succ DMU 2 \succ DMU 12. The less the value of d^* the better, and if d^* is negative then the DMU under analysis is efficient (DMUs 7, 5, 4 and 3). The remaining DMUs have $d^* > 0$ and hence are not efficient.

Table 2. Performances of DMUs in original scales and in utility scales for model 1.

DMU	Factors in original scales				Factors in utility scales			
	x_{DOC}	x_{NUR}	x_{ADM}	Y_{CONS}	u_{DOC}	u_{NUR}	u_{ADM}	u_{CONS}
1	8	8	6	8962	0.583	0.500	0.625	0.155
2	9	9	7	8439	0.500	0.438	0.500	0.139
3	7	5	6	6247	0.667	0.750	0.625	0.070
4	5	6	4	7291	0.875	0.667	0.875	0.103
5	4	5	4.5	5631	1.000	0.750	0.813	0.03
6	10	9	8	11942	0.417	0.438	0.417	0.248
7	15	14	12	19356	0.156	0.188	0.150	0.585
8	10	12	10	14011	0.417	0.250	0.250	0.334
9	6	6	5	7230	0.750	0.667	0.750	0.101
10	9.5	10	8	11436	0.458	0.375	0.417	0.232
11	13	13.5	10	16351	0.219	0.203	0.250	0.431
12	14	13	11	14300	0.188	0.219	0.200	0.346

Table 3. Performances of DMUs in original scales and in utility scales for model 2.

DMU	Factors in original scales					Factors in utility scales				
	X_{CMDT}	X_{MED}	X_{HR}	X_{OC}	Y_{CONS}	u_{CMDT}	u_{MED}	u_{HR}	u_{OC}	u_{CONS}
1	362900	1114935	1398033	123382	8962	0.633	0.824	0.572	0.743	0.155
2	463549	1192330	1050558	128606	8439	0.529	0.789	0.788	0.733	0.139
3	466561	1113734	1589818	101750	6247	0.527	0.824	0.475	0.815	0.070
4	277140	1321779	1001297	152974	7291	0.746	0.735	0.825	0.687	0.103
5	232565	1147295	1101200	141864	5631	0.820	0.809	0.753	0.708	0.051
6	613535	2491923	1724028	218890	11942	0.411	0.406	0.413	0.560	0.248
7	1332317	4365605	2368401	360011	19356	0.084	0.115	0.173	0.317	0.585
8	701121	2289334	2225321	491140	14011	0.355	0.450	0.220	0.136	0.334
9	314090	1196353	1054564	65416	7230	0.694	0.787	0.785	0.945	0.101
10	606148	2954054	1348651	203098	11436	0.416	0.318	0.599	0.590	0.232
11	1007909	3350321	1933791	371468	16351	0.202	0.253	0.326	0.298	0.431
12	708936	2793686	2227392	278936	14300	0.350	0.347	0.219	0.452	0.346

Table 4. Results from Phase 1 without weight restrictions for model 1.

DMU	d^*	w^*_{DOC}	w^*_{NUR}	w^*_{ADM}	w^*_{CONS}
1	0.0599	0.000	0.391	0.088	0.521
2	0.1039	0.000	0.391	0.088	0.521
3	-0.0099	0.000	0.487	0.000	0.513
4	-0.0625	0.000	0.000	1.000	0.000
5	-0.1250	1.000	0.000	0.000	0.000
6	0.0529	0.000	0.429	0.053	0.518
7	-0.1535	0.000	0.000	0.000	1.000
8	0.0528	0.387	0.000	0.000	0.613
9	0.0120	0.000	0.391	0.088	0.521
10	0.0869	0.000	0.391	0.088	0.521
11	0.0523	0.000	0.000	0.399	0.601
12	0.1077	0.000	0.429	0.053	0.518

The GDM did not agree with null weights in the results, since this corresponds to disregarding some of the factors used in the evaluation. That led to the introduction of weight restrictions. For example, DMU 5 is efficient but all weights were null except w^*_{DOC} . This DMU has only four doctors and is the best on this criterion; therefore the vector of weights that favoured it most was (1,0,0,0).

In Table 5 the results from Phase 1 for model 2 are presented. For model 2 the ranking of units is (the first seven DMUs are efficient): DMU 9 > DMU 7 > DMU 5 > DMU 4 > DMU 1 > DMU 2 > DMU 3 > DMU 8 > DMU 10 > DMU 11 > DMU 12 > DMU 6.

The GDM was dissatisfied with the continued existence of units that did not consider all factors in their evaluation. In fact, DMU 5 and DMU 7 considered only one of the five factors to be ranked as efficient. Nevertheless, the rank of the units is completely different from model 1, including a higher number of efficient units. It was also explained that the introduction of weights restrictions, as a way of ensuring that all factors are counted in the analysis, would degrade the efficiency of some units, which could move from an efficient status to an inefficient status [54].

Table 5. Results from Phase 1 without weight restrictions for model 2.

DMU	d^*	w^*_{CMDT}	w^*_{MED}	w^*_{HR}	w^*_{OC}	w^*_{CONS}
1	-0.0339	0.112	0.421	0.000	0.000	0.467
2	-0.0226	0.000	0.242	0.186	0.000	0.572
3	-0.0115	0.000	0.846	0.000	0.154	0.000
4	-0.0420	0.205	0.000	0.795	0.000	0.000
5	-0.0740	1.000	0.000	0.000	0.000	0.000
6	0.0504	0.363	0.000	0.000	0.078	0.559
7	-0.1535	0.000	0.000	0.000	0.000	1.000
8	0.0238	0.327	0.096	0.000	0.000	0.577
9	-0.1672	0.000	0.000	0.207	0.793	0.000
10	0.0240	0.052	0.000	0.325	0.064	0.559
11	0.0241	0.029	0.000	0.396	0.000	0.575
12	0.0263	0.380	0.000	0.000	0.055	0.565

Table 6. Extreme performances associated with utility levels 0 and 1 (model 1).

Utility level	x_{DOC}	x_{NUR}	x_{ADM}	y_{CONS}
$u(.)=0$	20	20	15	4 000
$u(.)=1$	4	4	3	25 000

According to the procedure explained in section 4.3, weight restrictions were elicited by asking the GDM to compare the “swings” from utility 0 to 1 as depicted in Table 6, for model 1.

The GDM was asked to consider a unit with the performance level 0 for all factors and the question was: "If you could improve one and only one factor to the level 1, which one would it be?". The GDM answer was unanimous: y_{CONS} . This allows the inference that w_{CONS} is the highest scaling coefficient. By repeating this question successively for the remaining factors, we obtained the following ranking of the coefficients of scale (all the answers converged): $w_{CONS} \geq w_{DOC} \geq w_{NUR} \geq w_{ADM}$.

Null weights were avoided by asking an indifference judgment allowing to limit the ratio of the weights ranked first (w_{CONS}) and last (w_{ADM}): “What is the highest number h that would allow a DMU with 3 administrative, technical and other support professionals and 4000 medical consultations for registered patients to be considered as having more value than a unit with 15 administrative, technical and other support professionals and h medical consultations for registered patients?”. The answer allows to build the following inequality: $w_{ADM} u_{ADM}(3) + w_{CONS} u_{CONS}(4\ 000) \geq w_{ADM} u_{ADM}(15) + w_{CONS} u_{CONS}(h)$. The elicited bound was $h = 10000$. Increasing this amount the GDM would start hesitating whether the inequality would still hold. Substituting h in the previous expression yields: $w_{CONS} \leq 5.32 w_{ADM}$.

The same methodology was used for model 2. Table 7 shows the utility levels 0 and 1 for the extreme performance values. Using the same protocols to elicit the weight restrictions as before, the ranking obtained for the scale coefficients was: $w_{MED} \geq w_{CMDT} \geq w_{HR} \geq w_{OC} \geq w_{CONS}$. The answer to the question “What is the lowest amount h that would allow a unit with 25000 medical consultations for registered patients and total medicine costs billed to the NHS of 5.5 million euros to be considered as having more value than a unit with 4000 medical consultations for registered patients and total medicine costs billed to the NHS of h ?” was $h = 2.5$ million euros. This answer is translated as: $w_{CONS} u_{CONS}(25000)$

+ $w_{MED} u_{MED}(5,500,000) \geq w_{CONS} u_{CONS}(4000) + w_{MED} u_{MED}(h)$. Substituting h yields: $w_{MED} \leq 2.47 w_{CONS}$.

The results for Phase 1 and Phase 2 with the addition of weight restrictions and free slacks are presented in Table 8 for model 1. There are no longer null weights and being the best DMU in some of the factors no longer guarantees being considered efficient.

Table 7. Extreme performances associated with utility levels 0 and 1 (model 2).

Utility level	X_{CMDT}	X_{MED}	X_{HR}	X_{OC}	Y_{CONS}
$u(.)=0$	1 600 000	5 500 000	3 000 000	600 000	4 000
$u(.)=1$	150 000	800 000	800 000	50 000	25 000

Table 8. Phase 1 and Phase 2 results under weight restrictions and free slacks (model 1).

DMU	Phase 1					Phase 2			
	d^*	w^*_{DOC}	w^*_{NUR}	w^*_{ADM}	w^*_{CONS}	s^*_{DOC}	s^*_{NUR}	s^*_{ADM}	s^*_{CONS}
1	0.0715	0.143	0.143	0.143	0.571	0.292	0.167	0.250	-0.052
2	0.1184	0.163	0.163	0.106	0.567	0.375	0.229	0.375	-0.036
3	0.0655	0.170	0.170	0.105	0.556	0.208	-0.083	0.250	0.033
4	-0.0080	0.145	0.145	0.145	0.564				
5	-0.0454	0.314	0.314	0.059	0.314				
6	0.0788	0.163	0.163	0.106	0.567	0.458	0.229	0.458	-0.145
7	-0.0767	0.120	0.120	0.120	0.639				
8	0.0738	0.205	0.109	0.109	0.578	0.458	0.417	0.625	-0.231
9	0.0348	0.163	0.163	0.106	0.567	0.125	-0.000	0.125	0.002
10	0.0912	0.163	0.163	0.106	0.567	0.417	0.292	0.458	-0.130
11	0.0621	0.143	0.143	0.143	0.571	0.656	0.464	0.625	-0.328
12	0.1198	0.163	0.163	0.106	0.567	0.688	0.448	0.675	-0.243

The ranking induced by d^* , for model 1 is: DMU 7 > DMU 5 > DMU 4 > DMU 9 > DMU 11 > DMU 3 > DMU 1 > DMU 8 > DMU 6 > DMU 10 > DMU 2 > DMU 12. The first three are efficient.

Table 9 shows the results obtained in Phase 1 and Phase 2 considering the weight restrictions and free slacks for model 2. In model 2, the d^* value obtained led to the following ranking: DMU 9 > DMU 5 > DMU 4 > DMU 2 > DMU 1 > DMU 3 > DMU 10 > DMU 6 > DMU 12 > DMU 8 > DMU 11 > DMU 7, and only the first two DMUs are efficient. Making the comparison of model 1 results without weight restrictions and with weight restrictions, DMU 3 loses the efficient status but the remaining efficient DMUs maintain the same ranking order in terms of efficiency. The last three inefficient units in both rankings remain the same. Regarding model 2, which had seven efficient units, only DMU 9 and DMU 5 are now efficient, which is better suited to reality according to the GDM's opinion. The position of the units in the two rankings is very different. In fact, the introduction of weight restrictions in model 2 led to major changes in the evaluation of DMUs.

Considering now model 1 and model 2 with weight restrictions, the GDM attached significance to the fact that DMU 5 remained efficient in both models, since their aim is to assess users' access, given the resources consumed in each unit. This result is consistent with the perception the GDM had of the operation of this unit in particular. DMU 9 is the first of the efficient units in model 2, but the first

inefficient one in model 1, which makes sense because of the importance that the GDM gave to costs (more weight) in model 2 to the detriment of y_{CONS} . Note that DMU 9 has more human resources than DMU 5; however, it has less total cost of human resources and less than half in medicine costs not billed to the NHS, clinical consumables and other costs. A major turnaround happens with DMU 7, which is efficient in model 1 with and without restrictions (the best of all units) and also in model 2 without weight restrictions, but falls to the last position in the ranking in model 2 with weight restrictions. The explanation is that although DMU 7 has more y_{CONS} than any other unit, it spends a lot of money in medicines and x_{CMDT} and these two factors have more weight for the GDM. This was welcomed by the GDM, knowing that the unit in question provides health services to a very elderly population, which in turn consumes more in medicines and x_{CMDT} .

Table 9. Phase 1 and Phase 2 results under weight restrictions and free slacks (model 2).

DMU	Phase 1						Phase 2				
	d^*	w^*_{CMDT}	w^*_{MED}	w^*_{HR}	w^*_{OC}	w^*_{CONS}	s^*_{CMDT}	s^*_{MED}	s^*_{HR}	s^*_{OC}	s^*_{CONS}
1	0.0512	0.155	0.382	0.155	0.155	0.155	0.061	-0.037	0.213	0.201	-0.054
2	0.0511	0.155	0.382	0.155	0.155	0.155	0.164	-0.002	-0.003	0.211	-0.038
3	0.0845	0.155	0.382	0.155	0.155	0.155	0.167	-0.037	0.311	0.130	0.031
4	0.0167	0.262	0.262	0.262	0.106	0.106	-0.053	0.052	-0.039	0.258	-0.002
5	-0.0059	0.311	0.311	0.126	0.126	0.126					
6	0.2541	0.200	0.200	0.200	0.200	0.200	0.283	0.381	0.372	0.385	-0.147
7	0.4085	0.200	0.200	0.200	0.200	0.200	0.610	0.672	0.612	0.628	-0.484
8	0.3547	0.277	0.326	0.132	0.132	0.132	0.339	0.337	0.565	0.809	-0.233
9	-0.0342	0.200	0.200	0.200	0.200	0.200					
10	0.2312	0.200	0.200	0.200	0.200	0.200	0.277	0.469	0.186	0.355	-0.131
11	0.3605	0.200	0.200	0.200	0.200	0.200	0.492	0.534	0.459	0.647	-0.330
12	0.3196	0.200	0.200	0.200	0.200	0.200	0.344	0.440	0.566	0.493	-0.245

6. Improvements of inefficient units and further developments

Phase 2 uses the optimal weighting vector $w^* \in W$, to solve the LP problem (2) with free slacks. The solution is a proposal of an efficiency target (projection) for each inefficient DMU. To attain an efficient status these inefficient DMUs must change their utility in each factor by the amount indicated by s^* , considering the two models. Note that not all the proposed changes correspond to improvements in all the factors since some of the slacks are negative. For an inefficient unit to "emulate" its peers on the efficient frontier, if it has a negative slack for an input factor this corresponds to an increase of the input, and if it has a negative slack for an output factor this corresponds to a decrease of the input. This is an interesting result since for model 1 it can be shown that some units (e.g., DMU 3) have an excessive number of doctors and few nurses and to be efficient those units should decrease the number of doctors, even if at the same time they increase the number of nurses. As we have anticipated, the GDM did not fully understand the reduction of y_{CONS} proposed for most of the inefficient units. Additionally, the increase in costs and the y_{CONS} reduction proposed in the results of Phase 2 of model 2 were unacceptable to the

GDM. It was necessary to propose another solution, but these preliminary results were instrumental in helping the GDM discuss what improvements should be sought for inefficient DMUs.

Let us consider that u_k^* is the utility of the best DMU using the optimal weighting vector $w^* \in W$ resulting from (1) with the weight restrictions added, i.e., the utility value that an inefficient DMU k ought to achieve:

$$u_k^* = \sum_{c=1}^q w_k^* u_c(DMU_k) + d_k^* \quad (3)$$

The LP problem solved in Phase 2 of the Value-Based DEA method admits alternative optimal targets, each of which corresponds to a different way of closing the value gap d_k^* . For this case, the GDM intends to constrain the efficiency targets that are proposed to achieve u_k^* , not only to avoid those targets that imply an increase of inputs or a decrease of outputs, but also to choose which factors can be changed, given the characteristics of each unit.

Let $S_< = \{c \in \{1, \dots, q\}: s_c^* < 0 \text{ in Phase 2}\}$ denote the negative slacks in the optimal solution obtained for (2) with free slacks; these slacks will now become null constants. Let $S_> = \{c \in \{1, \dots, q\}: s_c^* \geq 0 \text{ in Phase 2}\}$ denote the remaining slacks, which will be considered as non-negative variables. Therefore, a formulation that yields an alternative target that maintains or improves the performance of all factors can be obtained by solving the LP model (4) in which the maximum slack (in terms of value) shall be minimized to achieve the global utility value target. No negative slacks are allowed, but the target will no longer be a convex combination of the observed DMUs (for details see [33]).

The formulation below differs from the one used by Almeida and Dias [33] in that it introduces a new parameter σ_c (a positive value) for each factor to bound the value that a slack may have. The aim is to avoid setting unrealistic targets, allowing the proposals to be better tailored to the GDM's requirements. The targets will never exceed the value 1 in any factor due to the constraint $u_c(DMU_k) + s_c \leq 1$, $c = 1, \dots, q$. This ensures the utility function does not “spill over” outside the performances ranges elicited.

$$\begin{aligned} \min_{\rho_k, s} \quad & \rho_k \\ \text{s.t.} \quad & \sum_{c \in S_>} w_c^* s_c = d_k^* \\ & u_c(DMU_k) + s_c \leq 1, c = 1, \dots, q \\ & s_c - \sigma_c \leq 0, c \in S_> \\ & \rho_k \geq s_c \geq 0, c = 1, \dots, q \end{aligned} \quad (4)$$

For model 1 all the inefficient units had chosen the DMU 4 as peer and the GDM began by analysing the results in terms of proposals to improve the inefficient units, blocking changes in the factors that had negative slacks in Phase 2 (see proposal 1 in Table 10). Given the characteristics of certain units, the GDM deemed some of the proposed reductions in the number of medical and non-medical staff were not reasonable. Thus, the proposal emerged to maintain the original performance of the input factors for some units, or to limit the reduction of those factors for other units, in model 1. The gap between the unit

under evaluation and DMU 4 is closed by increasing the output factor (see proposal 2 in Table 10). For some units, the increase in the number of medical consultations for registered patients was considered to be achievable. It is well known that not all registered patients attend the health centre regularly, thereby hindering continuity of the relationship established between the patient and the general practitioner (in 2005 approximately 60% of the registered patients did not use the services of the health centre [55]). For other units, the proposed numbers of medical consultations for registered patients would be impossible to achieve. That led to formulation (5) in which the value gap relatively to the efficient peer is distributed in a balanced way on all the factors (see proposal 3 in Table 10).

$$\begin{aligned}
& \min_{r_k, s_c} r_k \\
& \text{st. } \sum_{c=1}^q w_c s_c = d_k^* \\
& u_c(DMU_k) + s_c \leq 1, c=1, \dots, q \\
& s_c - s_c \leq 0, c=1, \dots, q \\
& r_k \geq s_c \geq 0, c=1, \dots, q
\end{aligned} \tag{5}$$

The improvements to all the scenarios referred to above are displayed in Table 10, with the factors that did not change in **bold**. The GDM was very pleased to be able to choose in each case the improvement proposal that they felt was most appropriate knowing the capabilities and context of each unit.

Table 10. Improvement proposals for inefficient units in original scales (model 1).

DMU	Proposal 1				Proposal 2				Proposal 3			
	X _{DOC}	X _{NUR}	X _{ADM}	Y _{CONS}	X _{DOC}	X _{NUR}	X _{ADM}	Y _{CONS}	X _{DOC}	X _{NUR}	X _{ADM}	Y _{CONS}
1	6	6	4.5	8962	8	8	6	12708	7	7	5.5	11248
2	6	5	5	8439	9	9	7	14340	7.5	7.5	6	12156
3	6	5	5.5	8752	7	5	6	10000	6.5	4.5	5.5	8336
6	8	6.5	6	11942	9	9	8	14126	9	8	7	13836
8	8	9	8	14011	10	11	10	16328	9	11	9	15780
9	5.5	6	4.5	8560	6	6	5	9168	6	5.5	4.5	8336
10	7	7	6	11436	9.5	9	8	15620	8.5	8.5	7	13761
11	10.5	10.5	8.5	16351	12	12	10	18632	11.5	12	9.5	17820
12	9.5	8	7.5	14300	12	12	11	17830	11.5	10.5	9.0	17172

The proposed improvements to model 2 are presented in Table 11. Taking into account the model at stake and the units being assessed, the GDM was satisfied with Proposals 1 and 2. In Proposal 1, changes to achieve the same global utility as the peer are made on the factors that have non-negative slacks in Phase 2. In Proposal 2, a balanced distribution of the gap value relative to the peer is made by changing all factors (inputs and outputs), which the GDM felt more adequate to their needs.

Table 11. Improvement proposals for inefficient units in original scales (model 2).

DMUs	Proposal 1					Proposal 2				
	X_{CMDT}	X_{MED}	X_{HR}	X_{OC}	Y_{CONS}	X_{CMDT}	X_{MED}	X_{HR}	X_{OC}	Y_{CONS}
1	279690	1114935	1209770	91020	8962	321660	1010895	1307830	107540	10576
2	313390	1192330	1050558	78420	8439	411540	1081470	983170	110340	10070
3	338080	1113734	1001297	63580	10610	382410	946840	1423210	78140	8975
4	277140	1211670	1001297	129620	7291	266750	1281300	980500	144700	7824
6	289130	1349720	1132450	84300	11942	335680	1526860	1232350	102220	18040
7	399000	1636530	1208170	98860	19356	508190	1991940	1382440	133780	24898
8	266750	1042590	1297510	226860	14011	303160	1156910	1393440	255700	21000
10	306050	1694290	921550	84020	11436	351050	1894610	994930	100260	17124
11	347010	1408170	1067090	121300	16351	429500	1674810	1201810	168100	22498
12	275090	1296200	1314760	91860	14300	332510	1509310	1461310	114260	20632

7. Concluding remarks

This study reports an intervention in which analysts and health experts worked together in an assessment of PHC units, seeking to identify best practices and inform decision-making. The participation of the three elements constituting the GDM was important at different times. The two health professionals had a more active intervention in the construction of the assessment models and utility functions, as well as in restrictions setting, while the coordinator of the Management Support Unit had his main intervention on building the assessment models and on the collection of data and information. The three members performed the discussion of results. Five meetings were carried out with the three DMs together in most of them. As the roles of each DMs are well defined in the ACES, the interaction process was always held in a cordial atmosphere and there was rarely room for disagreement. The Value-Based DEA method was used as a modelling basis and then tailored to the requirements of this application, namely to define realistic targets (by the introduction of a new parameter into the model), to allow ranking the DMUs.

The study to assess the performance of PHC units started with the choice by the GDM of factors to include into the models, namely regarding access. Two models to assess PHC units were created as a results of discussions. The incorporation of the GDM's preferences in those models was made possible by the conversion of input and output factors into utility, which required interpreting the utility functions as devices to compare the change in value for the GDM that corresponds to performance differences. Models became more complete and credible for the GDM only after the introduction of weight restrictions that were elicited using the swings technique to draw a ranking of weights. Moreover, a trade-off question was used to limit the ratio between the weights ranked first and last and consequently avoid null weights. The effort required from the DMs depends on the number of factors. With a high number of factors it can be more difficult for DMs to provide a ranking for the respective weights. However, typically the number of factors used in DEA applications is not large.

A ranking of all the units and benchmarks was identified for each model and it was possible to conclude that the Family Health Units are not worst or better than Personalized Health Care Units. In the access evaluation model with and without restrictions on weights, the results showed that there are

efficient health units of both types. However, in the Model 1 results the efficient units are mostly Personalized Health Care Units, whether considering or not restrictions on weights. The improvement proposals to inefficient health centres emerged taking advantage of the GDM's feedback on the model results and their expert knowledge about the units, in order to obtain more realistic targets. For this purpose a new model was developed.

Our findings helped the GDM in the identification of best practices, sources of inefficiency and gaps to the best practices, as well as possibilities and opportunities for improvement. The results were useful not only for the introduction of corrective measures, but also to inform decisions about the goals of PIs in the future. It was also possible to conclude that there are units that given their characteristics, such as having more elderly registered users who in turn have more chronic diseases, benefited from the inclusion of indicators that reflect that. In the future it would be interesting to include indicators chosen by the health units themselves and evaluate in the same group the ones that chose the same indicators (or better, to allow comparisons, DMUs would choose between a small set of indicator groups).

This application to a real-world problem allowed ascertaining the value of the proposed methodology in practice. The GDM easily understood the methodology and the receptivity increased during the interaction process. The most prized results were the complete ranking of the units and the different proposals for improvement, since the GDM had then the possibility to choose how to improve each unit taking into account its specific characteristics. The answers to the trade-off questions were the methodological aspect that was received with more difficulty by the GDM.

In the future it would be interesting to develop DEA models capable of helping the PHC appraise investments in the production of outcomes.

Acknowledgements

This work has been partially supported by the Portuguese Foundation for Science and Technology under project grant UID/MULTI/00308/2013. Authors are also indebted to the anonymous reviewers and the area editor for their constructive comments and suggestions on earlier versions of this paper.

REFERENCES

- [1] Carter M, Hans EW, Kolisch R (2012) Health care operations management. *OR Spectrum* 34:315-317.
- [2] Donabedian A (1980) *The Definition of Quality and Approaches to its Assessment*. Health Administration Press, Ann Arbor, MI.
- [3] Baker R (1992) General practice in Gloucestershire, Avon and Somerset: Explaining variations in standards. *British Journal of General Practice* 42:415–418.
- [4] van den Hombergh P, Grol R, van den Hoogen HJM, van den Bosch WJHM (1998) Assessment of management in general practice: Validation of a practice visit method. *British Journal of General Practice* 48:1743–1750.
- [5] Campbell SM, Roland MO, Quayle JA, Buetow SA, Shekelle PG (1998) Quality indicators for general practice: Which ones can general practitioners and health authority managers agree are important and how useful are they?. *Journal of Public Health Medicine* 20:414–421.

- [6] McColl A, Roderick P, Gabbay J, Smith H, Moore M (1998) Performance indicators for primary care groups: An evidence based approach. *British Medical Journal* 317:1354–1360.
- [7] Proctor S, Campbell J (1999) A developmental performance framework for primary care. *International Journal of Health Quality Assurance* 12: 279–286.
- [8] Thanassoulis E, Boussofiene A, Dyson RG (1996) A comparison of data envelopment analysis and ratio analysis as tools for performance assessment. *Omega* 24:229–244.
- [9] Giuffrida A, Gravelle H, Sutton M (2000) Efficiency and administrative costs in primary care. *Journal of Health Economics* 19:983–1006.
- [10] Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. *European Journal of Operational Research* 2:429-444.
- [11] Sherman HD, Zhu J (2006) *Improving Service Performance using Data Envelopment Analysis*. Springer, New York.
- [12] Cooper WW, Seiford LM, Zhu J (2004) *Handbook on Data Envelopment Analysis*. Boston: Kluwer Academic Publishers, Boston.
- [13] Hollingsworth B, Dawson PJ, Maniadakis N (1999) Efficiency measurement of health care: a review of non-parametric methods and applications. *Health Care Management Science* 2:161-172.
- [14] Sommersguter-Reichmann M (2003) Analysing hospital productivity changes using non-parametric approaches. *OR Spectrum* 25:145-160.
- [15] Amado CAF, Dyson RG (2008) On comparing the performance of primary care providers. *European Journal of Operational Research* 185:915-932.
- [16] Huang YG, McLaughlin CP (1989) Relative efficiency in rural primary health care: an application of data envelopment analysis. *Health Services Research* 24:143-158.
- [17] Giuffrida A (1999) Productivity and efficiency changes in primary care: a Malmquist index approach. *Health Care Management Science* 2:11-26.
- [18] Linna M, Nordblad A, Koivu M (2003) Technical and cost efficiency of oral health care provision in Finnish health centres. *Social Science & Medicine* 56:343-353.
- [19] Kirigia JM, Emrouznejad A, Sambo LG, Munguti N, Liambila W (2004) Using data envelopment analysis to measure the technical efficiency of public health centers in Kenya. *Journal of Medical Systems* 28:155-166.
- [20] Rosenman R, Friesner D (2004) Scope and scale inefficiencies in physician practices. *Health Economics* 13:1091-1116.
- [21] Rouse P, Harrison J, Turner N (2011) Cost and Performance: Complements for Improvement. *Journal of Medical Systems* 35:1063-1074.
- [22] Thanassoulis E, Portela MC, Graveney M (2012) Estimating the scope for savings in referrals and drug prescription costs in the General Practice units of a UK Primary Care Trust. *European Journal of Operational Research* 221:432–444.
- [23] Salinas-Jimenez J, Smith P (1996) Data envelopment analysis applied to quality in primary health care. *Annals of Operational Research* 67:141-161.
- [24] Wagner J, Shimshak D, Novak M (2003) Advances in physician profiling: the use of DEA. *Socio-Economic Planning Sciences* 37:141-163.
- [25] Nuti S, Daraio C, Speroni C, Vainieri M (2011) Relationships between technical efficiency and the quality and costs of health care in Italy. *International Journal for Quality in Health Care* 23:324-330.
- [26] Milliken O, Devlin RA, Barham V, Hogg W, Dahrouge S, Russell G (2011) Comparative Efficiency Assessment of Primary Care Service Delivery Models Using Data Envelopment Analysis. *Canadian Public Policy* 37:85-109.
- [27] Thanassoulis E, Portela MC, Allen R (2004) Incorporating value judgments in DEA. In: Cooper WW, Seiford LM, Zhu J (Ed) *Handbook on Data Envelopment Analysis*, Kluwer, Boston, pp 99-137.
- [28] Golany B (1988) An interactive MOLP procedure for the extension of data envelopment analysis to effectiveness analysis. *Journal of the Operational Research Society* 39:725–734.
- [29] Joro T, Korhonen P, Wallenius J (1998) Structural comparison of Data Envelopment Analysis and Multiple Objective Linear Programming. *Management Science* 44:962-70.
- [30] Halme M, Joro T, Korhonen P, Salo SE, Wallenius J (1999) A value efficiency approach to incorporating preference information in Data Envelopment Analysis. *Management Science* 45:103-15.

- [31] Mavrotas G, Trifillis P (2006) Multicriteria decision analysis with minimum information: combining DEA with MAVT. *Computers & Operations Research* 33:2083-2098.
- [32] Gouveia MC, Dias LC, Antunes CH (2008) Additive DEA based on MCDA with imprecise information. *Journal of the Operational Research Society* 59:54-63.
- [33] Almeida PN, Dias LC (2012) Value-based DEA models: application-driven developments. *Journal of the Operational Research Society* 63:16-27.
- [34] Gouveia MC, Dias LC, Antunes CH, Boucinha J, Inácio CF (2015) Benchmarking of maintenance and outage repair in an electricity distribution company using the Value-Based DEA method. *Omega* 53: 104-114.
- [35] Ali AI, Lerme CS, Seiford L (1995) Components of efficiency evaluation in Data Envelopment Analysis. *European Journal of Operational Research* 80:462-473.
- [36] Athanassopoulos AD, Podinovski VV (1997) Dominance and potential optimality multiple criteria models decision analysis with imprecise information. *Journal of the Operational Research Society*:142-150.
- [37] Dias LC, Clímaco JN (2000) Additive Aggregation with Variable Interdependent Parameters: the VIP Analysis Software. *Journal of the Operational Research Society* 51:1070-1082.
- [38] Keeney RL, Raiffa H (1976) *Decisions with multiple objectives: preferences and value tradeoff*. Wiley, New York.
- [39] PHP (2010) *Metodologia de Contratualização, Coordenação Estratégica dos CSP*. Administração Central do Sistema de Saúde. (In Portuguese).
- [40] Amado CAF, Santos SP (2009) Challenges for performance assessment and improvement in primary health care: The case of the Portuguese health centres. *Health Policy* 91:43-56.
- [41] Ferreira C, Marques RC, Nicola P (2013) On evaluating health centers groups in Lisbon and Tagus Valley: efficiency, equity and quality. *BMC Health Services Research* 13:529.
- [42] Gouveia MC, Dias LC, Antunes CH (2013) Super-efficiency and stability intervals in additive DEA. *Journal of the Operational Research Society* 64:86-96.
- [43] von Winterfeldt D, Edwards W (1986) *Decision analysis behavioral research*. Cambridge University Press, New York.
- [44] Goodwin P, Wright G (1998) *Decision analysis for management judgment*, 2nd edn. Wiley, Chichester.
- [45] Thrall RM (2000) Measures in DEA with an application to the Malmquist index. *Journal of Productivity Analysis* 13:125–137.
- [46] Andersen P, Petersen NC (1993) A procedure for ranking efficient units in data envelopment analysis. *Management Science* 39:1261-1264.
- [47] Barros PP, Machado SR, Simões JA (2011) Portugal health system review. *Health systems in transition* 13:1–156.
- [48] Lapão LV (2008) The role of complexity and innovation in the new primary care governance model in Portugal. *The Innovation Journal—Special Issue on Complexity* 13(3).
- [49] Lapão LV, Dussault G (2012) From policy to reality: clinical managers' views of the organizational challenges of primary care reform in Portugal. *International Journal of Health Planning and Management* 27:295–307.
- [50] Gouveia MC (2012) *Efficiency evaluation with a DEA-MCDA method with incorporation of preferences and data uncertainty*. Ph.D. Thesis, University of Coimbra. <http://hdl.handle.net/10316/21070>.
- [51] Cook WD, Zhu J (2009) Piecewise linear output measures in DEA (third revision). *European Journal of Operational Research* 197:312-319.
- [52] Salo AA, Hämäläinen RP (2001) Preference Ratios in Multiattribute Evaluation (PRIME) - elicitation and decision procedures under incomplete information. *IEEE Transactions on Systems Man and Cybernetics part A* 31:533-545.
- [53] Podinovski VV (2004) Production trade-offs and weight restrictions in data envelopment analysis. *Journal of the Operational Research Society* 55:1311–22.
- [54] Dyson RG, Allen R, Camanho AS, Podinovski VV, Sarrico CS, Shale EA (2001) Pitfalls and protocols in DEA. *European Journal of Operational Research* 132:245-259.
- [55] IGIF (2006) *Estatística do Movimento Assistencial dos Centros de Saúde em 2005*. Instituto de Gestão Informática e Financeira da Saúde. (In Portuguese)