ORIGINAL PAPER

The effect of quality and socio-demographic variables on efficiency measures in primary health care

José Manuel Cordero Ferrera · Eva Crespo Cebada ·
Luis R. Murillo Zamorano

Abstract This paper aims to extend the literature on measuring efficiency in primary health care by considering the influence of quality indicators and environmental variables conjointly in a case study. In particular, environmental variables are represented by patients' characteristics and quality indicators are based on technical aspects. In order to deal with both aspects, different extensions of data envelopment analysis (DEA) methodology are applied. Specifically, we use weight restrictions to ensure that the efficiency scores assigned to the evaluated units take quality data into account, and a four-stage model to identify which exogenous variables have impact on performance as well as to compute efficiency scores that incorporate this information explicitly. The results provide evidence in support of the importance of including information about both aspects in the analysis so that the efficiency measures obtained can be interpreted as an accurate reflection of performance.

Keywords Primary health care · Data envelopment analysis · Efficiency · Quality management

JEL Classifications I12 · C14

Introduction

The control of efficiency in the provision of health care is of constant concern to policy-makers worldwide, especially given the high level of expenditure reached in most developed countries. Therefore, it is not surprising that, beginning with the seminal study on nursing services by Nunamaker [1], there has been great interest in evaluating the performance of health care organisations. By far the majority of this work has concentrated on hospitals, with primary health care services having received much less attention.[1] The present study attempts to redress this imbalance with an evaluation of technical efficiency for a sample of primary health care centres. For this purpose, we use data envelopment analysis (DEA) [2]—a nonparametric technique that has been employed widely in the health sector because it can easily handle multiple dimensions of performance and is less vulnerable to the misspecification problems that affect econometric models.[2]

The aim of this paper was to contribute to extending the literature on measuring economic efficiency in primary health care by considering jointly two particularly relevant aspects of health care provision: the quality of care and the effect of environmental factors. Both aspects have been identified in the literature as key factors with a significant influence on results. Accordingly, recent research on this subject has considered quality [3] and environmental factors [4] when measuring economic efficiency in primary health care, although both aspects have been considered separately so far in previous literature. This paper constitutes a first approach to combining both issues within the same framework. The methodological approaches

J. M. Cordero Ferrera (✉) · E. C. Cebada ·
L. R. Murillo Zamorano
Department of Economics, University of Extremadura,
Av. Elvas s/n, 06071 Badajoz, Spain
e-mail: jmcordero@unex.es

---

[1] An exception is Steinmann et al. [5], who provide a comprehensive review of efficiency studies in both hospitals and primary care services.

[2] See [6–8] for an overview of studies on the measurement of efficiency in the health sector.

employed in this study to deal with both aspects help us minimise potential bias that could affect the results.

We measure quality in health care services through technical indicators that reflect the capacity of medical staff to diagnose and treat medical problems. These indicators are based on control programs established by health care authorities in order to improve the provision of services, and can be understood as particular qualitative outputs of the production process. When resources are constrained, there is an inevitable trade-off between (quality) non-quantitative output and (activity) quantitative output [9]. Therefore, any model specification should take into account the implications of such a trade-off and including both activity and quality indicators in the output of the process. We ensure this by incorporating weight restrictions into the DEA models evaluated in this study.

The performance of a health centre can also be influenced by factors beyond its control, especially by differences in the socio-demographic characteristics of the patients that each unit serves [10]. Despite the acknowledged importance of this effect, the few empirical studies that have attempted to incorporate this information performed only second-stage analysis in order to identify potential explanatory factors of inefficient behaviour [11, 12]. In our view, however, the population plays an active part in the process that takes place in health care centres, so this information must be incorporated unequivocally into the construction of efficiency scores. The present study addresses this problem by explicitly including this type of exogenous information using a four-stage model developed by Fried et al. [13] and subsequently enhanced by Cordero et al. [14]. This methodology not only allows the most influential variables to be identified, but also distinguishes different effects for each variable, which could be of great help in interpreting the results.

The rest of the paper is organised as follows. The following section on "Methodology" explains the methodological approaches employed to incorporate the effect of qualitative indicators and environmental factors into the efficiency scores calculated with DEA. "Dataset and variables" describes the characteristics of the dataset and the variables used in our study, while "Results" presents and discusses the main results. Finally, "Conclusions" are drawn in the last section.

Methodology

Although the importance of quality is widely assumed in the health care sector, there have been few attempts to incorporate quality indicators into the performance measurement of primary health care units by means of DEA models. Moreover, the approaches used so far have significant drawbacks. Some studies employ a two-stage estimation procedure by which efficiency scores are estimated from data on inputs and outputs in the first stage, and then those scores are regressed on a set of explanatory variables that include some proxies for quality [15, 16]. This formulation assumes that quality indicators influence the efficiency with which inputs generate outputs, but that they do not influence the transformation process itself, which is difficult to maintain given that the level of quality attained depends crucially on the quantity and form of the resources employed [17]. Another option to include quality indicators in a DEA evaluation is provided by Shimshak and Leonard [18] and Shimshak et al. [19], who suggest the estimation of two different DEA models, one for output indicators based on activities and another for quality outputs. After obtaining initial results from the two separate models, efficient units in a model that have low scores in the other are eliminated. The main shortcoming of this method is that the evaluation of a set of units is incomplete, since some of them are removed according to the criterion selected.

An alternative approach to overcome those limitations is to include quality measures in DEA models as an additional output [20–23]. The problem that arises with this approach is that the number of efficient units increases artificially due to the loss of discriminating power that DEA undergoes when the total number of variables increases and the number of observations remains constant. Moreover, the inaccuracy of this technique in assigning weights to different output indicators can lead to unreasonable results given that the weights provided are frequently inconsistent with prior knowledge of the production process involved. Thus, some units will achieve high scores or even will be identified as efficient in spite of performing poorly in terms of quality simply because this component is ignored by having a zero weight assigned to it.

In our opinion, it is possible to strike a balance between rigidity and excessive flexibility by including weight restrictions in DEA models. This approach, which has been employed previously in some empirical studies measuring efficiency in the health sector [24, 25], consists of restricting the multipliers by directly imposing a lower bound on the weights for both quantity and quality outputs considered in the analysis.[3] This procedure eliminates the possibility that units with a poor performance in terms of quality (or quantity) might be considered as fully efficient, and hence the reference units must have a demonstrated good performance on both indicators. By doing so, we attempt to diminish the shortcomings associated to preceding approaches developed in the literature, although the

---

[3] See [26, 27] for details.

use of this method implies the assumption of strong value judgments about the importance of each dimension.[4]

If appropriate specification of the model is important in terms of properly considering the importance of quality, then the development of models that take into account factors that are not under the control of the health care provider and that can be a potential source of inefficiency are no less so. An evaluation of health care facilities should explicitly include this information in the analysis to ensure that the efficiency score finally assigned to the unit truly reflects the portion of the production process for which that centre is itself responsible [28]. Particularly, in making an efficiency assessment of primary health care units these variables are represented mainly by the characteristics of the population demanding care [29]. Examples of these factors are poverty rate, crude birth rate, mortality rate, and minority populations. In this latter case, for instance, minority populations are more likely to be poor and to endure poverty-related conditions such as chronic illnesses, inadequate health-related behaviour and stress [30].

Given that it is not feasible to observe the health status of a centre's target population prior to serving them, these factors are beyond the organisation's control, and must therefore be given specific treatment in the efficiency analysis. Moreover, our interest in this study is not just to identify the variables that can affect the output of primary care centres, but also to explicitly include them in the calculation of efficiency scores so that those scores can be an appropriate measure of how the units are performing. We would stress that it is important to bear this purpose in mind, since it determines to a large extent the methodological approach used to deal with those exogenous variables, as will be explained in the rest of this section.

Recent years have seen the development of different models to incorporate the effect of so-called exogenous factors, uncontrollable inputs, or environmental variables in the production process on estimating efficiency scores through non-parametric approaches as DEA [31]. The approach most widely used in the health context is the two-stage approach [32], although it only allows for identifying factors affecting outcomes such as funding issues or geographic location [33]. The main drawback of this approach is that it does not correct adequately for the efficiency scores incorporating the effect of the exogenous variables, because it assumes wrongly that the influence of non-controllable inputs is the same over all inputs and outputs of the production process. Therefore, variables that explain

the overall inefficiency of a decision making unit (DMU) can be inconsistent or incongruent with those that cause its individual input inefficiencies [34].

An alternative method to calculate corrected efficiency scores was proposed by Fried et al. [35]. This work claims that inefficiency is not neutral among the variables included in the initial evaluation because there can be significant *slacks* (non-radial inefficiency) in some of them.[5] From this idea, Fried et al. [13] proposed a four-stage model based on the use of the total slacks obtained from the first DEA stage, which distinguishes the technical inefficiency and the effect of exogenous factors for each variable. This model also uses regression as a mechanism to adjust the inefficiency values obtained with DEA, but in this case applying a system of n Tobit regressions (one for each variable) in the second stage:

$$T\hat{S}_j = f\left(Z_j, \beta_j\right) + u_j \quad j = 1, 2, \ldots, n$$

where $T\hat{S}_j$ represent the total slacks (radial and non-radial) computed in the first stage, Z_j are the exogenous variables, β_j the parameters to be estimated, and u_j the error term, which is distributed normally, $u_j \approx N(0, \sigma^2)$. This decomposition can be done through the estimated parameters $\left(\hat{\beta}_j\right)$, which can be used to predict new slacks for each variable taking into account the real endowment of exogenous variables of each unit. These coefficients provide information about the negative or positive effect of the uncontrollable inputs on each slack. This effect may vary from one slack to another, and it is possible that some exogenous factors influence some of the slacks but not all. Once the decomposition has been done, in the third stage the original values of the variables are corrected using the predictions obtained in the second stage in order to discount the effect of exogenous factors. According to Fried et al. [13], when an input orientation is considered, the adjustments are calculated by adding to the original value of each input the difference between the maximum predicted slack and the predicted slack: $X_{\mathrm{adj}} = X + \left\lfloor Max\{TS_{j\,\mathrm{pred}}\} - TS_{j\,\mathrm{pred}}\right\rfloor$. Finally, by running a DEA with the corrected values of the variables, one obtains new scores that establish exclusively the efficiency level at which each producer operates.

This four-stage procedure constitutes an attractive alternative to calculate technical efficiency scores in the presence of exogenous factors, since the scores can be interpreted readily as production targets. In fact, it has been used previously in the literature to obtain a measure of managerial inefficiency that controls for the effect of

---

[4] Our purpose is limited to provide practitioners and decision makers with a useful tool to perform an efficiency analysis of primary health care centres considering quality and quantitative indicators jointly. The setting of the exact value of weight restrictions should therefore be undertaken by highly qualified experts in order to preserve equity and fairness in the evaluation.

[5] Slacks are potential improvements for inefficient units that are not accounted for in the DEA (radial) score. See Toffallis [36] for an illustrated definition of *slacks*.

exogenous factors for a set of dialysis facilities in Greece [37]. The main shortcoming of this model is that, in its second-stage, the estimated parameters can be biased because the total slacks are computed by taking into account the information for the whole sample and are thus correlated with each other [38]. However, this problem can be overcome by using the enhanced method developed by Cordero et al. [14], which provides unbiased estimations for the parameters using a bootstrap approach in the second stage. To the best of our knowledge, this work represents the first empirical study using this enhanced version of the four-stage model in the context of primary health care.

Dataset and variables

The database used in this application compiles information about the primary health care sector in the Spanish region of Extremadura in the year 2006. Due to the extension of the territory (41,634 square kilometres in area) and its low population density (26.18 inhabitants per square kilometre), this sector is structured into two fundamental territorial administrative levels of aggregation: Health Areas and Health Zones. The former represent the reference entities for establishing health care objectives and funding needs, and their limits and size are established by the regional government according to organisational criteria. At that moment, there were eight Health Areas with one public hospital in each of them. The latter are all organised around a primary care centre (PCC) as the main provider of primary health care services. Those PCCs are the units evaluated in this study. Specifically, our dataset is referred to 94 centres, after eliminating some of them with missing data and others identified as outliers.[6]

We retrieved data from APEX06 [39]—an integrated information system for primary care that provides detailed information for each one of the above PCCs on multiple variables, including the population covered, human resources, activities, costs, accessibility indicators and socioeconomic indicators. Table 1 reports the descriptive statistics of the all the variables used in the analysis together with their role in the productive process of a PCC.

As output indicator we use two sets of variables. The first is the most common in primary health care studies: the number of visits or consultations per primary health care professional (activity-output variables). For each PCC evaluated in the study, the variables FREQUENCYGP, FREQUENCYP, FREQUENCYN, and FREQUENCYU indicate the number of visits or consultations per capita with general practitioners (GPs), paediatricians, nurses, and emergency units, respectively.

The second set of variables is represented by information about quality aspects, which can be defined according to multiple criteria [41]. For example, health managers tend to focus on professional standards and health outcomes, while patients often relate quality to an understanding attitude, communication skills and the comfort of health care facilities. The combination of these three quality dimensions in a single indicator is extremely complex, thus most authors attempting to measure quality usually focus on a single component. In our case, we dismissed the use of variables based on patient satisfaction because the underlying scale used by different customers depends considerably on their prior expectations and therefore is not identical for all units evaluated [42]. Furthermore, it can be argued that those expectations and, consequently, their opinions about facilities and the interpersonal aspects of quality, also depend on their socio-economic characteristics [43], which could imply the existence of a potential bias due to the presence of correlation between output and environmental variables [44]. Therefore, in our study, the measurement of quality involves the use of three (quality-output) variables capturing technical aspects of the production process related to staff characteristics and the accomplishment of certain objectives. In particular, we use information derived directly from available evidence, expert opinions and clinical guidelines [45]. The variable EXPERIENCE is a proxy for the experience of GPs and paediatricians that work in each PCC, measured in days of work during the previous 15 years. HEALTHTARGET captures the extent to which the PCC is able to fulfil some specific health targets. Specifically, it represents an average of the coverage ratios of each of the programs implemented within the PCC portfolio services.[7] Finally, QUESTIONS indicates the number of affirmative answers to a ten-question questionnaire distributed to the managers of the PCCs. This questionnaire is based on the standards considered in the model of total quality elaborated for the Public Health Service of Extremadura and reports information on the three different categories of qualitative

---

[6] The total number of PCCs in 2006 was 103, but 6 of them were excluded due to having missing data, while 3 were considered as outliers by the method developed by Johnson and McGinnis [40]. The main advantage of this method is that it fits better to the multi-stage approach adopted in our empirical analysis, since it avoids potential distortions that could arise if there are outliers among the units exhibiting particularly poor performance.

[7] The health targets include the following items: population aged 0–14 entered in the register of vaccinations and correctly vaccinated; population aged over 15 entered in the register of adult vaccinations and correctly vaccinated against tetanus; population belonging to a risk group registered and correctly vaccinated against hepatitis B; population aged over 64 vaccinated against influenza; population aged 0–14 included in the dental health program; and female population included in the health program for early diagnosis of breast cancer.

Table 1 Main variables descriptive statistics

Variable	Role	Mean	SD	Maximum	Minimum
FREQUENCYGP	Activity-output	10.78	3.43	18.93	0.79
FREQUENCYP	Activity-output	6.55	2.78	17.08	0.95
FREQUENCYN	Activity-output	7.79	3.05	19.14	2.55
FREQUENCYU	Activity-output	1.63	0.68	3.27	0.19
EXPERIENCE	Quality-output	5,239.45	698.97	5,741.00	2,216.00
HEALTHTARGET	Quality-output	57.09	15.18	89.81	24.47
QUESTIONS	Quality-output	5.51	1.34	10.00	2.00
HLAB1	Input	28.55	15.36	89.59	6.16
HLAB2	Input	16.80	12.52	63.90	0.00
NHLAB	Input	1.24	0.78	4.49	0.20
PHARMA	Input	21.61	4.41	32.52	11.60
INDACT	Activity-output index	50.43	28.99	98.67	3.63
INDQUA	Quality_output index	50.67	30.46	95.64	5.63
CBR	Exogenous variable	8.01	2.97	22.42	1.79
DR	Exogenous variable	59.24	13.76	126.00	37.44
ER	Exogenous variable	23.48	7.32	49.32	9.18
RR	Exogenous variable	132.04	30.17	261.80	61.21
DENSITY	Exogenous variable	33.84	32.46	187.44	3.54
AGRIEMP	Exogenous variable	33.71	18.46	67.37	1.77

aspects, continuous learning of medical personnel and health management skills.[8]

Unfortunately, we cannot use all the available output variables in DEA because the technique would lose its discrimination power. Moreover, activity-output variables were (unsurprisingly) found to be correlated strongly among them. The same was detected for the quality-output variables, although they were uncorrelated with activity-output variables. We therefore decided to use principal component analysis (PCA) in order to condense the original set of variables into two single components.[9] This method decomposes original data with correlated values into a new set of uncorrelated (i.e. orthogonal) variables. Depending on the context, these variables are known as principal components, factors, eigenvectors, singular vectors or loadings. Each factor or principal component is a linear combination of the standardised values of the original variables used for the definition of the index. The weight given to each of these variables corresponds to its statistical correlation with the latent dimension that the synthetic index attempts to measure.[10] The number of factors to be retrieved depends on the correlation of the initial variables. If they are correlated strongly with each other, one factor will be sufficient to explain most of their variance. However, if the correlation is weak, several factors are required in order to explain a significant percentage of their variance. In this case, a set of intermediate indicators could be achieved, as many as there were common factors, and the final synthetic index will be calculated as their weighted sum. The proportion of the total variance explained defines the relevance of each factor.

Following this methodological approach, we performed two separate analyses and calculated two synthetic indices for each PCC in the sample. Figure 1 shows the relationships among the initial variables and the new components calculated for each PCC in the sample, the levels of correlation (given in square brackets), and the common factors (F) involved in the definitions of those indices. INDACT is an activity-output index that synthesizes information on quantitative performance of each PCC by combining the number of visits or consultations with each of the types of primary care professional (GPs, paediatricians, nurses and

---

[8] Specifically, the questionnaire contains ten questions inquiring about the following items: the existence of a training plan for doctors, a stock list of medicines, an organisation chart in the centre, a monitoring plan for economic imbalances, a maintenance plan for the physical facilities, an informative guide about the services provided, accessibility for people with disabilities, a survey questionnaire designed to obtain information about the level of membership of medical staff in scientific associations and regular attendance of medical staff at conferences.

[9] Adler and Yazhemsky [46] use a Monte Carlo simulation to compare the performance of the PCA approach with the variable reduction and demonstrate that the former provides more accurate results.

[10] See [47, 48] for a detailed explanation of the procedure.
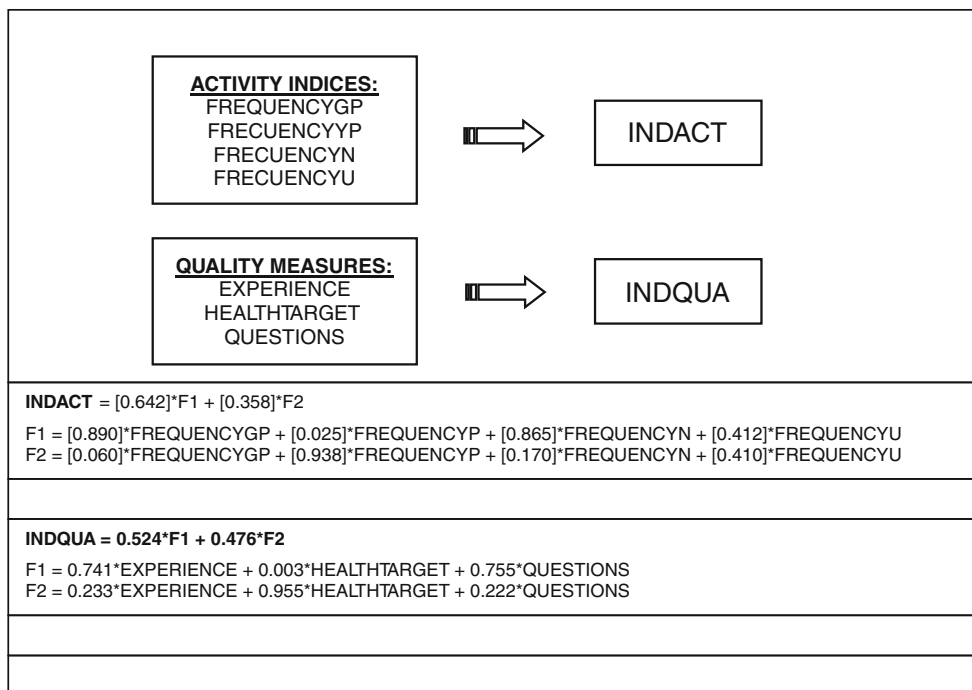
Fig. 1 Synthetic indices and principal component analysis (PCA). Levels of correlation are given in square brackets. *F* indicates the common factors involved in the definitions of the indices. *INDACT* Activity-output index that synthesizes information on quantitative performance of each primary care centre (PCC) by combining the number of visits or consultations with each of the types of primary care professional (GPs, paediatricians, nurses and emergency units). *INDQUA* Compiles information about the experience of the medical personnel, the coverage ratios of the portfolio of services and the affirmative answers to the questionnaire completed by PCC managers

emergency units), while INDQUA compiles information about the experience of the medical personnel, the coverage ratios of the portfolio of services and the affirmative answers to the questionnaire completed by PCC managers.

As input variables, we use data on personnel (labour) and prescriptions, represented by the following four variables: HLAB1, the total number of medical staff, including GPs, paediatricians, nurses, nursing assistants, emergency physicians and emergency nursing assistants; HLAB2, the total number of other technicians such as physiotherapists, dentists and X-ray specialists; NHLAB, the total number of non-medical staff, i.e. administrative staff, porters, veterinarians and social workers; and PHARMA, the number of prescriptions per capita, since these also form part of the budget assigned to each health centre.

Finally, in order to include information about exogenous variables that might affect the activities of the PCCs, we collect a large volume of data on the characteristics of population covered by each unit. This dataset includes demographic, geographic and economic variables.

The compilation of these data is a difficult task, since some information is available only for municipalities and not for health zones. In addition, other desirable indicators such as educational and occupation levels of population or patient health status and wealth simply were not accessible. However, in our opinion, the six variables used in this study provide us with very helpful information (differentiated for each health zone) that can be interpreted as acceptable proxies for those unavailable data. Those variables are the crude birth rate (CBR), the elderly ratio (ER), dependency rate (DR), replacement rate (RR), population density (DENSITY) and the percentage of population employed in agriculture (AGRI-EMP). Table 2 lists the Pearson correlation coefficients between these variables and the two synthetic output indices (INDACT and INDQUA). One observes that all six exogenous variables are correlated significantly with both indicators.

Results

The empirical study consisted of two phases. In the first, initial efficiency scores were estimated without including the effect of the exogenous variables but explicitly distinguishing the implications between specifying an activity-oriented model or a quality-oriented one. Then, in the second phase, the efficiency scores were computed after incorporating the set of exogenous variables and correcting the scores obtained in the first stage.

Table 2 Pearson correlation coefficients between exogenous and indexed output variables

Exogenous variables	INDACT	INDQUA
CBR	−0.566**	−0.421**
DR	0.640**	0.496**
ER	0.695**	0.563**
RR	−0.619**	−0.491**
DENSITY	−0.581**	−0.423**
AGRIEMP	0.420**	0.459**

** Significant at 99 % level

DEA without exogenous variables

Since one of the principal objectives of this study was to evaluate the impact of including quality in the measurement of technical efficiency in the primary health care sector, in this first stage of the analysis two separate DEA models were calculated, one based on activity indicators and one on qualitative indicators. Both models include the four input variables described in the previous section. Nevertheless, the synthetic activity index INDACT is the only output included in the first model while the synthetic quality index INDQUA is the only output included in the second. In both cases, the DEA model has an input orientation and assumes variable returns to scale (VRS). Adopting VRS models [49] allows us to accommodate scale effects in the analysis in order to avoid the potential inefficiencies that may arise if the units were forced to assume a non-optimal scale of production.[11] The input orientation seemed to be the most appropriate option given that the demand for health services cannot be controlled, and regional administration managers can determine only those resources attributed to each PCC to provide those services adequately.

Columns 2 and 3 of Table 3 report the efficiency scores obtained with those two specifications.[12] The two models identify a similar number of efficient units (20 and 18, respectively) and present similar mean values. Nevertheless, there are some noteworthy divergences in the two rankings of the units, as reflected in the value of the Spearman correlation coefficient between them (0.549).[13] Indeed, further examination of the values shown in Table 3

highlights the difficulties that arise when activity and qualitative indicators are included in an efficiency analysis independently, since some units identified as efficient using the activity-output index present some of the lowest levels of efficiency when the quality-output index is selected (e.g. units 30, 33, 35, 63). The opposite situation also happens. For example, the units 14, 55 and 85 should not serve as a benchmark despite being fully efficient in the qualitative model, because they present a lower level of efficiency in terms of the activity model specification.

This problem is even more evident if one focuses on the units used as references in each model (Table 4). According to the values given in the first two columns, it can be observed that some of the units among the main one referenced in the model based on activity indicators have less importance in the quality-output model (unit 17) or are even considered to be inefficient in that model (unit 29). This also happens for the quality-output model, where one of the main referenced unit (55) is considered inefficient in the activity-output model.

These divergences would seem to be sufficiently important to merit the attention of researchers and policy-makers regarding the need to take both quantitative and qualitative aspects into account. In this way, it can be ensured that measures of the performance of health units in the primary health care sector are reliable and non-biased. The simplest way of including the two dimensions in the analysis is to run a new DEA model with the two previous synthetic indices, INDACT and INDQUA, used now as a dual output within a unique model (combined model). Column 4 of Table 3 reports the efficiency scores obtained with this alternative model specification.

According to the Spearman correlation coefficient reported in Table 5, the efficiency scores calculated with this new DEA model are very similar to those obtained with the previous activity model (0.921). However, some units undergo a slight improvement with the inclusion of the quality-output index in the combined model and even become efficient (units 14, 32, 53, 55 and 85). Likewise, those units that were efficient in terms of the activity model, but inefficient units in terms of the quality model are still efficient (4, 14, 55, 81 and 85). The same happens to those that were qualitatively efficient but quantitatively inefficient (units 30, 33, 35 and 63). The reason for this is that units that present clearly better results in one dimension are usually assigned weights with a zero value to the other. This is a common shortcoming in DEA models, which essentially ignores one of the dimensions in these cases.

In order to overcome this limitation, we propose incorporating weight restrictions on the output measures in the combined efficiency model initially calculated. This approach allows one to ensure that both dimensions

---

[11] Following Banker [50], we tested the data for scale effects, finding that there are units operating at different scales. Moreover, the VRS assumption facilitates comparison with the constrained model, for which the status of returns to scale may undergo a change when weight restrictions are added [51].

[12] Efficiency scores have been calculated using DEAP 2.1 software [52], which also estimates radial and non-radial slacks for each variable using a multi-stage process.

[13] Table 5 reports the Spearman correlation coefficients among all the models specified and calculated in this study.

Table 3 Efficiency scores of primary care centres (PCCs) under alternative model specifications

PCC	Separate models		Combined model	Constrained combined model with weight restrictions ($\lambda = 20 \%$)
	Activity-output (INDACT)	Quality-output (INDQUAT)		
1	0.749	0.553	0.749	0.727
2	0.707	0.543	0.737	0.737
3	0.955	0.955	0.955	0.955
4	0.956	1.000	1.000	0.956
5	0.900	0.891	0.900	0.899
6	1.000	1.000	1.000	1.000
7	0.763	0.654	0.773	0.773
8	0.794	0.595	0.794	0.726
9	0.851	0.672	0.851	0.834
10	0.749	0.539	0.749	0.714
11	0.792	0.732	0.841	0.841
12	0.603	0.584	0.603	0.603
13	0.625	0.602	0.625	0.624
14	0.822	1.000	1.000	1.000
15	0.796	0.809	0.874	0.874
16	0.831	0.831	0.831	0.831
17	1.000	1.000	1.000	1.000
18	0.817	0.808	0.817	0.814
19	0.842	0.800	0.841	0.837
20	1.000	1.000	1.000	1.000
21	0.871	0.584	0.871	0.711
22	0.737	0.578	0.737	0.650
23	0.744	0.649	0.750	0.750
24	1.000	1.000	1.000	1.000
25	0.830	0.815	0.830	0.825
26	0.927	0.621	0.927	0.745
27	1.000	1.000	1.000	1.000
28	0.860	0.725	0.967	0.967
29	1.000	0.756	1.000	1.000
30	1.000	0.640	1.000	1.000
31	0.830	0.826	0.830	0.830
32	0.951	0.758	1.000	1.000
33	1.000	0.669	1.000	0.994
34	0.596	0.586	0.596	0.586
35	1.000	0.555	1.000	0.744
36	0.543	0.603	0.603	0.561
37	0.754	0.747	0.754	0.754
38	0.589	0.825	0.825	0.600
39	0.903	0.544	0.903	0.794
40	0.985	0.516	0.985	0.953
41	0.742	0.519	0.742	0.686
42	0.606	0.553	0.606	0.585
43	0.908	0.719	0.908	0.862
44	0.722	0.584	0.722	0.651

Table 3 continued

PCC	Separate models		Combined model	Constrained combined model with weight restrictions ($\lambda = 20 \%$)
	Activity-output (INDACT)	Quality-output (INDQUAT)		
45	0.753	0.690	0.753	0.690
46	0.760	0.772	0.793	0.793
47	0.764	0.777	0.779	0.773
48	0.683	0.556	0.687	0.686
49	1.000	1.000	1.000	1.000
50	0.688	0.687	0.794	0.794
51	0.809	0.598	0.809	0.699
52	1.000	1.000	1.000	1.000
53	0.850	0.991	1.000	0.990
54	1.000	1.000	1.000	1.000
55	0.869	1.000	1.000	1.000
56	0.830	0.562	0.830	0.780
57	0.731	0.575	0.769	0.769
58	0.669	0.582	0.669	0.668
59	0.646	0.824	0.824	0.786
60	1.000	1.000	1.000	1.000
61	0.777	0.384	0.777	0.396
62	1.000	1.000	1.000	1.000
63	1.000	0.604	1.000	0.960
64	0.709	0.518	0.727	0.727
65	1.000	1.000	1.000	1.000
66	0.599	0.647	0.700	0.701
67	0.564	0.508	0.594	0.594
68	0.655	0.773	0.773	0.754
69	0.818	0.796	0.818	0.813
70	0.885	0.958	0.982	0.988
71	0.628	0.624	0.653	0.653
72	1.000	0.836	1.000	1.000
73	0.822	0.971	0.971	0.965
74	0.666	0.735	0.779	0.779
75	0.964	0.932	1.000	1.000
76	0.979	0.721	1.000	1.000
77	0.663	0.659	0.675	0.675
78	0.969	0.763	0.982	0.982
79	0.928	0.737	0.927	0.836
80	0.634	0.759	0.759	0.753
81	0.927	1.000	1.000	1.000
82	0.742	0.769	0.812	0.812
83	0.916	0.668	0.916	0.909
84	1.000	0.963	1.000	1.000
85	0.809	1.000	1.000	1.000
86	0.530	0.602	0.668	0.672
87	1.000	1.000	1.000	1.000
88	0.814	0.515	0.827	0.825
89	0.705	0.519	0.705	0.653

Table 3 continued

PCC	Separate models		Combined model	Constrained combined model with weight restrictions ($\lambda = 20\%$)
	Activity-output (INDACT)	Quality-output (INDQUAT)		
90	1.000	1.000	1.000	1.000
91	0.927	0.927	0.927	0.927
92	0.911	0.990	0.996	0.960
93	0.935	0.603	0.935	0.788
94	0.730	0.557	0.765	0.765
Total average score	0.828	0.751	0.860	0.833

Table 4 Units used as a reference in more cases under alternative model specifications

Activity-output (INDACT)		Quality-output (INDQUAT)		Combined model		Constrained combined model with weight restrictions ($\lambda = 20\%$)	
Unit	Refs	Unit	Refs	Unit	Refs	Unit	Refs
84	58	24	55	84	41	84	46
24	48	55	41	24	35	24	41
29	39	54	29	29	31	29	34
17	37	6	23	17	28	17	31
65	12	87	22	65	20	65	22
54	9	17	12	55	19	87	18
30	8	65	11	81	14	81	17
6	6	81	11	87	14	55	16
63	5	4	10	30	7	6	8
72	3	14	6	6	6	32	5

(activity and quality) are considered in the DEA results. Moreover, it ensures that units with a low value in one of those indicators cannot be placed in the boundary. In the present case, we acknowledge that such weight restrictions reflect the perceived relative importance of the two indicators, thus representing a value judgement. Health managers concerned about the results of the evaluation should therefore decide these restriction values. The contribution of the present work is to show how the establishment of such restrictions may affect the results. To this end, we tested three potential lower bounds on the ratios of the weights (10, 20 and 30 %, respectively), finding that they led to similar results.[14] Obviously, the average efficiency is

lower when the constraint is higher. For the sake of simplicity, we present only the scores obtained using the intermediate value (20 %), which involves a sufficient level of restriction without to any great extent hampering the chance of any given unit to be considered efficient if it reaches a certain level of performance in one of the two dimensions.

A close examination of the efficiency scores calculated with this constrained combined model (Column 5 in Table 3) shows that a total of four units (4, 33, 35 and 63) that were fully efficient in the combined model (Column 4) become non-fully efficient when the weights are restricted. The consideration of the two dimensions also affects the identification of reference units. In particular, the constrained model identifies units 24 and 84 as the main references for the rest of the units (final column in Table 4), coinciding with the top-ranked reference units in the activity-output model (84) and the quality-output model (24).

In view of the result presented in this section, it seems clear that ignoring qualitative aspects in measuring technical efficiency of primary health care centres might lead to biassed and inappropriate results. One potential option to correct this misspecification problem within a non-parametric context could be the implementation of a constrained combined model that considers both quantitative and qualitiative output indicators.

DEA with exogenous variables

In this section, we consider the second stage of the empirical study where information about exogenous variables affecting the performance of units is assessed. In the context of our study, those factors are represented by the socio-demographic characteristics of the population served by the health centres. Under the reasoning just outlined, we estimate the efficiencies levels of each one of the PCCs of our study by means of incorporating the set of exogenous variables into the specification of a weight-constrained combined model. Specifically, we consider total slacks (radial and non-radial components) of the four input variables included in the model, and regress them on the six exogenous variables selected as a representation of the patient characteristics (CBR, ER, DR, RR, DENSITY, AGRIEMP). In order to avoid bias in the estimates that result from the censored normal Tobit regressions in the original four-stage model, we apply the enhanced method based on a bootstrap procedure developed in [14].[15] Table 6 presents the results.

The analysis of the estimated parameters allows some preliminary inferences to be drawn about the influence of

---

[14] The correlation coefficients between all the pairs of these models are greater than 0.93 in every case (see Table 5).

[15] We performed 1,000 iterations.

Table 5 Spearman correlation coefficients among different model specifications

	Activity-output (INDACT)	Quality-output (INDQUAT)	Combined model	Constrained combined model ($\lambda = 10\%$)	Constrained combined model ($\lambda = 20\%$)	Constrained combined model ($\lambda = 30\%$)
Activity-output (INDACT)	1.000					
Quality-output (INDQUAT)	0.549	1.000				
Combined model	0.921	0.690	1.000			
Constrained combined model ($\lambda = 10\%$)	0.876	0.747	0.946	1.000		
Constrained combined model ($\lambda = 20\%$)	0.830	0.775	0.906	0.990	1.000	
Constrained combined model ($\lambda = 30\%$)	0.787	0.800	0.862	0.962	0.988	1.000

Table 6 Estimated parameters of total slacks using a Tobit bootstrap procedure. The standard errors are shown in brackets

Regressor	Dependent variable			
	Slacks HLAB1	Slacks HLAB2	Slacks NHLAB	Slacks PHARMA
Constant	9.341** (1.318)	8.146** (1.972)	8.414** (1.643)	4.743** (0.864)
CBR	−0.313 (0.407)	−0.519 (0.374)	−0.371** (0.120)	−0.118 (0.155)
ER	0.674** (0.174)	0.431* (0.186)	0.168 (0.083)	0.891** (0.217)
DR	0.388** (0.092)	0.245** (0.074)	0.092* (0.042)	0.012 (0.091)
RR	0.167 (0.093)	0.168 (0.125)	−0.023 (0.011)	0.168* (0.066)
Density	−0.184** (0.039)	−0.464** (0.067)	−0.136** (0.018)	−0.018 (0.019)
AGRIEMP	0.164** (0.046)	0.129* (0.048)	0.071** (0.019)	0.027 (0.018)

* Significant at 95 %, ** significant at 99 %

the patient characteristics on efficiency. Firstly, it is noteworthy that the influence of external factors varies across different inputs, although we can identify several significant parameters that enable us to claim that the correction of the initial scores to include the effect of these variables is totally justified. The elderly ratio (ER) has a significant (and positive) effect on the slack of every input variable (with the exception of NHLAB), although its impact is greater on PHARMA and HLAB1. This positive effect means that a higher proportion of elderly population increases the inefficiency. In contrast, the population density (DENSITY) has a significant negative effect on the slack of every input variable (with the exception of PHARMA), which means that health zones with a higher density, usually large cities, perform more efficiently (the slacks are lower). Likewise, it is worth noting that some exogenous variables have a weak impact of almost every input, such as the replacement rate (RR) and the crude birth rate (CBR), while other variables affect only some specific variables, like the dependency rate (DR), which has only a significant (and positive) impact on the variables representing medical and technical staff (HLAB1 and HLAB2). Finally, the proportion of the population employed in agriculture (AGRIEMP), which may be interpreted as a proxy for the low income and low educational level section

of the population, has a negative effect on almost every input (positive effect on slacks), although this influence is smaller than other variables.

We next used the mean values of the parameters estimated by the bootstrap procedure to predict the total input slack for each unit based on the values of the exogenous variables.[16] These predictions adjust the primary output data according to the difference between maximum predicted slack and the predicted slack. The final stage consists of using the adjusted input data to run a new DEA model maintaining the restriction ($\lambda = 20\%$) for the output original values, so that it can be comparable to the previous model. The new efficiency scores allow one to distinguish the inefficiency that is attributable to management once the characteristics of the population being served are taken into account. Table 7 reports the efficiency scores obtained with this constrained four-stage bootstrapping model.

Comparing the values obtained from the constrained four-stage model with those calculated previously, one observes that there are some significant changes when socio-demographic patient characteristics are included in

---

[16] In this procedure, we take into account only those variables that have a significant effect on each slack.

Table 7 Efficiency scores with and without exogenous variables

PCC	Constrained combined model ($\lambda = 20\%$)	Constrained four-stage model ($\lambda = 20\%$)	Basic health zone	Constrained combined model ($\lambda = 20\%$)	Constrained four-stage model ($\lambda = 20\%$)
1	0.727	0.970	49	1.000	0.921
2	0.737	0.837	50	0.794	0.862
3	0.955	0.710	51	0.699	0.839
4	0.956	0.697	52	1.000	0.973
5	0.899	0.687	53	0.990	0.750
6	1.000	0.716	54	1.000	0.879
7	0.773	0.870	55	1.000	0.821
8	0.726	0.770	56	0.780	0.959
9	0.834	0.706	57	0.769	0.866
10	0.714	0.864	58	0.668	0.851
11	0.841	0.860	59	0.786	1.000
12	0.603	0.621	60	1.000	1.000
13	0.624	0.819	61	0.396	0.591
14	1.000	0.965	62	1.000	1.000
15	0.874	0.891	63	0.960	1.000
16	0.831	0.756	64	0.727	0.852
17	1.000	0.879	65	1.000	1.000
18	0.814	0.631	66	0.701	0.961
19	0.837	0.715	67	0.594	0.982
20	1.000	0.855	68	0.754	1.000
21	0.711	0.960	69	0.813	0.861
22	0.650	0.823	70	0.988	0.787
23	0.750	0.724	71	0.653	0.973
24	1.000	0.885	72	1.000	1.000
25	0.825	0.783	73	0.965	0.978
26	0.745	0.799	74	0.779	0.974
27	1.000	0.860	75	1.000	0.900
28	0.967	1.000	76	1.000	0.872
29	1.000	0.933	77	0.675	0.817
30	1.000	0.986	78	0.982	0.825
31	0.830	0.848	79	0.836	0.918
32	1.000	1.000	80	0.753	0.840
33	0.994	0.953	81	1.000	1.000
34	0.586	0.771	82	0.812	0.787
35	0.744	1.000	83	0.909	0.891
36	0.561	0.981	84	1.000	0.880
37	0.754	0.812	85	1.000	0.837
38	0.600	0.737	86	0.672	0.903
39	0.794	0.904	87	1.000	1.000
40	0.953	1.000	88	0.825	1.000
41	0.686	0.854	89	0.653	0.893
42	0.585	0.852	90	1.000	0.784
43	0.862	0.880	91	0.927	0.877
44	0.651	0.979	92	0.960	1.000
45	0.690	0.884	93	0.788	0.906
46	0.793	0.777	94	0.765	1.000
47	0.773	0.846	Total average	0.833	0.873
48	0.686	0.776			

Table 8 Reference units with and without exogenous variables

Constrained combined model with weight restrictions ($\lambda = 20$ %)		Constrained four-stage model with weight restrictions ($\lambda = 20$ %)	
Unit	Refs	Unit	Refs
84	46	59	62
24	41	92	37
29	34	68	31
17	31	63	26
65	22	32	20
87	18	88	19
81	17	60	13
55	16	40	12
6	8	81	10
32	5	72	9

the analysis. In general terms, the average efficiency rises, although the patterns of specific cases are quite diverse. Many units have higher efficiency scores and even some of them become efficient (units 32, 40, 44, 59, 63, 68, 79, 81, 88 and 92). Most of these units are located in zones with low population densities and high elderly ratios, reflecting that, once the evaluation has taken into account that these units are operating in an unfavourable context, they obtain higher efficiency scores.

In contrast, other PCCs obtain lower scores when these variables are included (units 6, 14, 17, 20, 24, 27, 29, 30, 49, 52, 54, 55, 75, 76, 84, 85 and 90). Those units benefited from the first evaluation because the characteristics of the population they were serving were ignored. Many PCCs experiencing a decrease in their efficiency scores belong to the two relatively large cities of the region, Badajoz and Caceres.[17] Among the units enumerated above, two (24 and 84) stand out because they were the principal references in the constrained model without exogenous variables, but have now become inefficient in this new model (Table 8). Indeed, it can be noted that only units 32 and 81 maintain their position as main referents in the four-stage model.

These results underline the importance of taking into account data about patient characteristics in the calculation of efficiency scores so that the evaluated units can be assigned production targets according to the context in which they are operating. In addition, the method employed allows us to identify which variables have a greater effect on input variables (for instance, the percentage of elderly population on the number of prescriptions) as well as correct the initial values of those variables based on unbiased parameters representing those effects.

---

[17] Here the values of the population density variable (DENSITY) are the greatest of the entire sample and the values of ER, AGRIEMP and DR are notably lower than in rural areas.

Conclusions

The empirical investigation developed in this study contributes to extending the literature on the measurement of economic efficiency in primary health care. We do that by considering jointly two particularly relevant aspects of health care provision: the quality of care, and the effect of environmental factors on the performance of primary care centres. To date, and despite their relevance, these two aspects have not received much attention in the literature on the measurement of economic efficiency in the field of primary health care. Moreover, to our knowledge, this investigation presents the first attempt to combine both issues jointly within the same framework. By doing so, the methodological approach introduced in this study helps to minimize the bias of previous investigations.

As a first departure from previous literature, and using a newly constructed information system for the primary health care sector (APEX06), we were able to construct two different measures of output, accounting for both activity and quality features of health services. Hence, we overcome the criticisms associated with the use of solely quantitative indicators of output.

However, simply using quality measures as additional outputs can result in some serious modelling problems when DEA is the method used to measure efficiency. The main problem arises due to the possibility that this technique may assign weights of zero value to quality indicators and hence allow units to be identified as efficient, even though they actually have low levels of quality in their performance. In the present application, we dealt with this problem by incorporating weight restrictions in a combined DEA model including both quantity and quality indicators. This ensured that both dimensions are properly taken into account in the analysis. The results of our investigation show that the use of weight restrictions modifies the composition of the efficient frontier considerably, as well as which units were identified as references to be used as benchmarks for the inefficient units.

A second aspect in which the present study improves on previous research has to do with the specification procedure employed by modelling exogenous variables related directly to the population served by each health centre but over which it has no control. In this sense, our aim was not limited to identifying which variables may affect the performance of PCCs, but to explicitly include them in the calculation of the efficiency scores and so better measure PCC performance. To this end, we used a recently developed four-stage model that includes a bootstrap procedure to ensure unbiased estimates of the parameters. The semi-parametric structure of this model, in which the effect of each exogenous variable can be tested by using a set of Tobit regressions, allowed us to identify two variables as

the main factors influencing PCC performance: the population density and the elderly ratio. Also, the analysis of specific cases showed units operating in the largest cities to be the most negatively affected by the correction of efficiency scores derived from the inclusion of exogenous variables in the analysis. We therefore believe that it is particularly important to include population socio-demographic characteristics in calculating the efficiency levels of primary health care centres. Otherwise, the results will most likely be affected by a major bias in their statistical and economic meaning.

In sum, observing the changes in the efficiency scores of the different PCCs considered in our study as the model specification was made progressively more sophisticated, we conclude that greater accuracy should always be required in both the specification and the estimation of economic models for the measurement of efficiency in the health sector. Specifically, in the field of primary health care, further research is still necessary in order to better understand how these centres operate. There is also a clear need for additional effort to be made in collecting reliable information about other aspects that may affect the performance of primary health care centres, and that were not considered in our evaluation due to the lack of available data. Some examples are the structure of financing, the methods used to determine payments to physicians, the economic level of municipalities served by the unit, and the proportion of immigrants in its target population. In this sense, the participation and implication of practitioners and decision-makers in retrieving data would be essential. Methodological approaches such as that described in the present work could facilitate work towards this goal and provide satisfactory analytical tools for decision makers in this sector.

References

1. Nunamaker, T.R.: Measuring routine nursing service efficiency: A comparison of cost per patient day and data envelopment analysis models. Health Serv. Res. **18**, 183–205 (1983)
2. Charnes, A., Cooper, W.W., Rhodes, E.: Measuring the efficiency of decision making units. Eur. J. Oper. Res. **2**(6), 429–444 (1978)
3. Murillo-Zamorano, L.R., Petraglia, C.: Technical efficiency in primary health care: Does quality matter? Eur. J. Health Econ. **12**, 115–125 (2011)
4. Cordero-Ferrera, J.M., Crespo-Cebada, E., Murillo-Zamorano, L.R.: Measuring technical efficiency in primary health care: The effect of exogenous variables on results. J. Med. Syst. **35**(4), 545–554 (2011)
5. Steinmann, L., Dittrich, G., Karmann, A., Zweifel, P.: Measuring and comparing the (in)efficiency of German and Swiss hospitals. Eur. J. Health Econ. **5**(3), 216–226 (2004)
6. Hollingsworth, B.: Non-parametric and parametric applications measuring efficiency in health care. Health Care Manag. Sci. **6**(4), 203–218 (2003)
7. Hollingsworth, B.: The measurement of efficiency and productivity of health care delivery. Health Econ. **17**, 1107–1128 (2008)
8. Worthington, A.C.: Frontier efficiency measurement in health care: a review of empirical techniques and selected applications. Medical Care Res. Rev. **61**(2), 135–170 (2004)
9. Newhouse, J.: Toward a theory of non-profit institutions: An economic model of a hospital. Am. Econ. Rev. **60**, 64–74 (1970)
10. Lezzoni, L.: Risk adjustment for measuring health care outcomes, Ann Arbor. Heath Administration Press, Michigan (1994)
11. Kontodimopoulos, N., Moschovakis, G., Aletras, V., Niakas, D.: The effect of environmental factors on technical and scale efficiency of primary health care providers in Greece. Cost Eff. Resour. Alloc. **5**, 14 (2007). doi:10.1186/1478-7547-5-14
12. Ramirez-Valdivia, M., Maturana, S., Salvo-Garrido, S.: A multiple stage approach for performance improvement of primary health care practice. J. Med. Syst. **35**(5), 1015–1028 (2011)
13. Fried, H., Schmidt, S., Yaisawarng, S.: Incorporating the operating environment into a nonparametric measure of technical efficiency. J. Prod. Anal. **12**, 249–267 (1999)
14. Cordero-Ferrera, J.M., Pedraja, F., Santín, D.: Enhancing the inclusion of non-discretionary inputs in DEA. J. Oper. Res. Soc. **61**, 574–584 (2010)
15. Ferrier, G., Valdmanis, V.: Rural hospital performance and its correlates. J. Prod. Anal. **7**, 63–80 (1996)
16. Linna, M., Häkkinen, U., Linnakko, E.: An econometric study of costs of teaching and research in finnish hospitals. Health Econ. **7**(4), 291–305 (1998)
17. Arocena, P., García-Prado, A.: Accounting for quality in the measurement of hospital performance: Evidence from Costa Rica. Health Econ. **16**, 667–685 (2007)
18. Shimshak, D., Lenard, M.: A two-model approach to measuring operating and quality efficiency with DEA. INFOR **45**(3), 143–151 (2007)
19. Shimshak, D., Lenard, M., Klimberg, R.K.: Incorporating quality into data envelopment analysis of nursing home performance: a case study. Omega **37**, 672–685 (2009)
20. Salinas-Jiménez, J., Smith, P.C.: Data envelopment analysis applied to quality in primary health care. Ann. Oper. Res. **67**, 141–161 (1996)
21. García, F., Marcuello, C., Serrano, D., Urbina, O.: Evaluation of efficiency in primary healt care centres: An application of data envelopment analysis. Financial Account. Manag. **15**(1), 67–83 (1999)
22. Wagner, J.M., Shimshak, D., Novak, M.: Advances in physician profiling: the use of DEA. Socio-Econ. Plan. Sci. **37**, 141–163 (2003)
23. Rosenman, R., Friesner, D.: Scope and scale efficiencies in physician practices. Health Econ. **13**, 1091–1116 (2004)
24. Chilingerian, J., Sherman, H.: DEA and primary care physicians report cards: Deriving preferred practice cones from managed care service concepts and operating strategies. Ann. Oper. Res. **73**, 35–66 (1997)
25. Shimshak, D.: Managing nursing home quality using DEA with weight restrictions. In: Lawrence, K.D., Kleinman, G. (eds.) Applications in Multicriteria Decision Making, Data Envelopment Analysis and Finance, Applications of Management Science, 14. Emerald Group Publishing, Bingley, UK (2010)
26. Dyson, R.G., Thanassoulis, E.: Reducing weight flexibility in data envelopment analysis. J. Oper. Res. Soc. **39**(6), 563–576 (1988)

27. Roll, Y., Cook, W.D., Golany, B.: Controlling factor weights in data envelopment analysis. IIE Trans. **23**(1), 2–9 (1991)

28. Muñiz, M.: Separating managerial inefficiency and external conditions in data. Eur. J. Oper. Res. **143**(3), 625–643 (2002)

29. Pestieau, P.: Assessing the performance of the public sector. Ann. Public Cooperative Econ. **80**, 133–161 (2009)

30. Marathe, S., Wan, T., Zhang, J., Sherin, K.: Factors influencing community health centers' efficiency: a latent growth curve modeling approach. J. Med. Syst. **31**, 365–374 (2007)

31. Cordero, J.M., Pedraja, F., Salinas, J.: Measuring efficiency in education: An analysis of different approaches for incorporating non-discretionary inputs. Appl. Econ. **40**(10), 1323–1339 (2008)

32. Hoff, A.: Second stage DEA: Comparison of approaches for modelling the DEA score. Eur. J. Oper. Res. **181**, 425–435 (2007)

33. Ozcan, Y.A.: Physician benchmarking: measuring variation in practice behavior in treatment of otitis media. Health Care Manag. Sci. **1**, 5–17 (1998)

34. Chen, A., Hwang, Y., Shao, B.: Measurement and sources of overall and input inefficiencies: evidences and implications in hospital services. Eur. J. Oper. Res. **161**, 447–468 (2005)

35. Fried, H.O., Lovell, C.A.K., Van Den Eeckaut, P.: Evaluating the performance of US Credit unions. J. Bank. Finance **17**, 251–265 (1993)

36. Tofallis, C.: Combining two approaches to efficiency assessment. J. Oper. Res. Soc. **52**, 1225–1231 (2001)

37. Kontodimopoulos, N., Papathanasiou, N., Tountas, Y., Niakas, D.: Separating managerial inefficiency from influences of the operating environment: an application in dialysis. J. Med. Syst. **34**(3), 397–405 (2010)

38. Simar, L., Wilson, P.W.: Estimation and inference in two-stage, semiparametric models of production processes. J. Econom. **136**, 31–64 (2007)

39. Murillo-Zamorano, L.R., Vega, J., De Miguel, F., Morillo, J. and Sánchez, M.: APEX06. Sistema de Información de la Atención Primaria en la Comunidad Autónoma de Extremadura, 2006. In: Murillo and Vega. (eds.) Badajoz, Spain (2011)

40. Johnson, A., McGinnis, L.F.: Outlier detection in two-stage semiparametric DEA models. Eur. J. Oper. Res. **187**, 629–635 (2008)

41. Donabedian, A.: The Definition of Quality and Approaches to its Assessment, vol. 1. Health Administration Press, Ann Arbor, Michigan (1980)

42. Dyson, R.G., Allen, R., Camanho, A.S., Podinovski, V.V., Sarrico, C.S., Shale, E.A.: Pitfalls and protocols in DEA. Eur. J. Oper. Res. **132**, 245–259 (2001)

43. Thompson, A., Sunol, R.: Expectations as determinants for patient satisfaction: Concept, theory and evidence. Int. J. Qual. Health Care **7**(2), 127–141 (1995)

44. Ruggiero, J.: Performance evaluation when non-discretionary factors correlate with technical efficiency. Eur. J. Oper. Res. **159**, 250–257 (2004)

45. Campbell, S.M., Braspenning, J., Hutchinson, A., Marshall, M.: Research methods used in developing and applying quality indicators in primary care. Qual. Saf. Health Care **11**, 358–364 (2002)

46. Adler, N., Yazhemsky, E.: Improving discrimination in data envelopment analysis: PCA-DEA or variable reduction. Eur. J. Oper. Res. **202**, 273–284 (2010)

47. Jobson, J.D.: "Principal components factors and correspondence analysis". In: Jobson, J.D. (ed.). Springer texts in statistics. Applied multivariate data analysis, Vol. II. Categorical and multivariate methods, pp. 345–482. Springer, New York, (1992)

48. Abdy, H.: Multivariate Analysis. In: Lewis-Beck M., Bryman A., Futing T. (eds.) Encyclopedia for research methods for social sciences. Sage, Thousand Oaks (2003)

49. Banker, R.D., Charnes, A., Cooper, W.W.: Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag. Sci. **30**(9), 1078–1092 (1984)

50. Banker, R.D.: Hypotesis tests using data envelopment analysis. J. Prod. Anal. **7**, 139–159 (1996)

51. Tone, K.: On returns to scale under weight restrictions in data envelopment analysis. J. Prod. Anal. **16**, 31–47 (2001)

52. Coelli, T.: "A Guide to DEAP Version 2.1, A Data envelopment analysis (computer) program. CEPA Working Paper 96/08 (1996)