HEALTH ECONOMICS
*Health Econ.* **7**: 263–277 (1998)

HOSPITAL PERFORMANCE

# TECHNICAL EFFICIENCY IN THE CLINICAL MANAGEMENT OF CRITICALLY ILL PATIENTS

JAUME PUIG-JUNOY*
*Department of Economics and Business, Pompeu Fabra University (UPF), Barcelona, Spain*

SUMMARY

The purpose of this paper is to obtain empirical measures of performance in the management of critical patients treated in intensive care units (ICUs) and to evaluate the factors associated with performance, in a two stage approach. In the first stage, this paper uses an extended version of Data Envelopment Analysis (non-discretionary and categorical variables, and weight constraints under consideration) to obtain measures of technical efficiency in the treatment of 993 critical care patients in intensive care units in Catalonia (Spain) in 1991–92. The model incorporates accurate individual measures of illness severity from Mortality Probability Models (MPM $II_0$) and quality outcome measures in the input–output set to obtain non-biased efficiency measures. In the second stage, a loglinear regression model is applied to test a number of hypothesis about the role of different environmental factors—such as ownership, market structure, dimension, internal organization, diagnostic, mortality risk, etc.—to explain differences in the efficiency scores. © 1998 John Wiley & Sons, Ltd.

KEY WORDS — data envelopment analysis; intensive care units performance; loglinear regression model; technical efficiency

## INTRODUCTION

The purpose of this paper is to obtain empirical measures of performance, rooted in the principles of production economics, in the management of critically ill patients and to evaluate the factors that are contributing to hospital performance in treating these patients. The method is applied to the individual clinical decisions relative to 993 critical patients in different intensive care units in Catalonia (Spain) in 1991/1992. We identify patients who have been treated in an intensive care unit (ICU) as critically ill.

Critical care is being closely scrutinized given the important contribution of these health care services to growing health care expenditures. Intensive care units were deemed to account for 1% of the Gross Domestic Product (GDP) [1] and 28% of hospitals costs in the US [2]. ICU performance measurement is obviously relevant from the policy point of view.

Measures of the Farrell [3] and Debreu [4] definition of technical efficiency are used as a proxy for ICU performance in this paper. These are partial but theoretically rooted indicators of ICU performance. An intensive care unit is technically inefficient in treating a patient if it does not minimize its inputs given its outputs. Technical efficiency has been advocated as an adequate measure to compare performance of firms having different ownership regimes or legal status, especially to evaluate and compare public sector and not-for-profit activities performance, which are predominant in the hospital sector. Data envelopment analysis (DEA) has proved especially valuable in hospitals and in many institutional settings

* Correspondence to: Department of Economics and Business, Pompeu Fabra University, Trias Fargas 25–27, 08005 Barcelona, Spain. E-mail: puig_jaume@econ.upf.es

CCC 1057–9230/98/030263–15$17.50

*Received 10 September 1996*
*Accepted 6 October 1997*

where non-marketed multiple output are considered and the correct weighting of outputs cannot be defined.

Empirical measurement of inefficiency ranges from two main alternative methodologies: stochastic parametric regression-based methods to non-stochastic non-parametric mathematical programming methods. DEA is the more commonly used family of linear programming models. Parametric methodology obtains efficiency measures computed in terms of the distance that lies between the observation and the estimated function. Thus, scores may differ according to the chosen functional specification. DEA, in contrast, assumes no measurement error or random fluctuations in input–output measures, being a completely deterministic method. Recent attemps to develop a non-parametric stochastic method (Chance constrained DEA) are Land, Lovell and Thore [5] and Olesen and Petersen [6].

Simulation studies comparing DEA with competing forms of statistical regressions indicate the relation between them depends on the choice of the functional form [7]. Comparisons of both methods show that some observations misclassified as efficient by DEA may be *corner* observations: those not appearing in the envelope of any inefficient observation. Recent research comparing the two approaches suggested that econometric and linear programming results do not differ dramatically, when based on the same data and conceptual framework [8].

An increasing number of researchers have recently applied DEA to institutional health care providers (hospitals, nursing homes, primary health care centers, pharmacies) to measure efficiency [9–15]. Institutions comprise many different decision levels, thus there are difficulties in attributing responsibilities for inefficiency in the organization. A more limited number of studies analysed physician efficiency [16], the level at which clinical decisions to allocate resources are made. At both the institutional and physician levels some patients may not be treated efficiently, however, they can not be identified. A unit may be efficient in treating some specific type of patients but not in others.

Health care may be interpreted as a very heterogenous production process given the presence of patients as an input which requires decisions about resource allocation to be specific for each of them. Most DEA studies do not consider previous patient characteristics as an input or consider them in a very rough aggregate form (i.e. the number of admissions). In this case, obtained efficiency scores may be strongly influenced by missing or erroneous measurements in individual data. Hospitals produce a wide range of heterogeneous outputs in differing proportions. Output dimensions have proved to be very difficult to measure. A way to alleviate this measurement problem in efficiency analysis is to use more homogenous and less aggregated units by observing specific services in the hospital, such as the intensive care units. Finkler and Wirtschaffer [17] analyzed obstetric services in nine hospitals.

Some recent attempts to measure and explain differences in ICU *performance* are Knaus *et al.* [18], Rapoport *et al.* [19] and Shortell *et al.* [20]. Knaus *et al.* [18] estimated a multiple logistic regression to explain differences in mortality and length of stay using patient and institutional characteristics as explaining factors. They derived two ratios of *performance*: the first is defined as the ratio between actual and predicted death rate at hospital discharge; and the second is defined as the same ratio for length of stay. Rapoport *et al.* [19] proceeded in a very similar way as Knaus *et al.*, but they used the Mortality Probability Model to predict death and length of stay. These authors conclude that no trade-off was observed between *clinical performance* and *resource use performance*. Clinical performance was measured as the difference between the actual achieved survival rate of patients treated in an ICU and the survival rate expected by the model. The resource use performance for each hospital was computed as the expected mean weighted hospital days minus actual observed mean weighted hospital days. The study conducted by Shortell *et al.* [20] using data from 42 ICUs with 200 or more beds was intended to shed light on 'more efficiently managing ICUs and reducing the variation in patient outcomes'. Measures of *performance* in this study are: risk-adjusted mortality, risk-adjusted length of stay, evaluated technical quality of care, evaluated ability to meet family member needs and nurse turnover. Ordinary least square results showed that factors explaining risk-adjusted mortality were technological availability and diagnostic diversity, the first with a negative effect and the second with a positive one; caregiver interaction appeared negatively related to risk-adjusted ICU length of stay. Technological availability was measured by how many of 39 recommended items were available in the unit.

Past literature presents two main limitations to measuring and explaining ICU performance. First, no homogenous and theoretically rooted concept of efficiency is used describing a clear relationship between inputs and outputs. Usually a broad set of non-related performance indicators is presented ranging from input (length of stay) to output (mortality) variables which have limited value in measuring productive efficiency. Thus, reviewed studies do not permit measurement and explanation of productive efficiency. Second, ratios between observed and expected values may only reflect average functions, but do not permit *a best practice pattern* of the input–output relationship to be estimated, a matter of considerable importance in an industry where incentives to cost minimization are scarce.

In this paper we propose to explore the usefulness of DEA to measure technical efficiency at the patient level. Decision making units are defined as the ICUs taking resource allocation decisions in an individual production process, that is, a patient. Defining this level of analysis allows us to consider in detail patient characteristics which constitute necessary dimensions of the input and output set. To assess the impact of health care providers on health outcomes it is necessary to use measures of inputs and outputs among individuals. The shortage of individual health data probably explains the exclusive use of aggregated data to assess the efficiency of providers. Production frontier and efficiency scores are computed through the comparison of homogenous patients treated in the same or different hospitals. In the econometric frontier approach an exception is provided by Bosmans and Fecher [21] who estimate a resource function defined as the relationship between medical fees incurred in the treatment of individual patients and the patient's pathology.

In the rest of the text we use the term ICU performance as being equivalent to performance in clinical management of critically ill patients. As treated in the paper, and as it is in reality, management of critically ill patients is a broader subject than ICU management, strictly speaking.

This paper makes a contribution to the existing literature on DEA and performance of critically ill patient management in three areas. First, it applies an extended DEA model (non-discretionary and categorical variables, and weight constraints under consideration) to the measurement of technical efficiency of ICUs at the patient level. Second, it incorporates severity of illness and quality measures in the input–output set obtained from Mortality Probability Models (MPM $\mathrm{II}_0$). And, third, it presents results from a loglinear regression model of environmental factors explaining differences in efficiency scores.

The paper is organized as follows. Section 2 lays out the general framework for the application of Data Envelopment Analysis to the measurement of technical efficiency. Variable definition and description is presented in Section 3. Section 4 presents DEA technical inefficiency results. Regression analysis of DEA efficiency scores is presented in Section 5. Section 6 concludes.

## THE EFFICIENCY EVALUATION METHODOLOGY

### *Efficiency definition*

Debreu [4] and Farrell [3] introduced a measure of input oriented technical efficiency defined as one minus the maximum equiproportionate reduction in all inputs that still allows continued production of given outputs.

To characterize the production technology relative to which efficiency is measured, ICUs use variable inputs $x=(x_1,\ldots, x_N)\in\mathbb{R}^N_+$ to produce variable outputs $y=(y_1,\ldots, y_M)\in\mathbb{R}^M_+$. Inputs are transformed into outputs using a technology that can be described by the graph: $\mathrm{GR}=\{(x, y)$: $x$ can produce $y\}$. Corresponding to the graph is a family of input sets: $L(y)=\{x\colon (x, y)\in\mathrm{GR}\}$, $y\in\mathbb{R}^M_+$. Input and output sets satisfy the properties of convexity and strong disposability of inputs. Inputs sets contain their isoquants: Isoq $L(y)=\{x\colon x\in L(y),\ \theta x\notin L(y),\ \theta\in[0, 1)\}$, $y\in\mathbb{R}^M_+$; which in turn contain their efficient subsets: Eff $L(y)=\{x\colon x\in L(y),\ x'\notin L(y),\ x'\leq x\}$, $y\in\mathbb{R}^M_+$.

Then, a radial measure of the technical efficiency of the input vector $x$ in the production of output vector $y$ is given by $\mathrm{TE_I}\ (x, y)=\min\{q\colon \theta x\in L(y)\}$, with $\theta=1$ indicating radial technical efficiency and $\theta<1$ indicating the degree of radial technical inefficiency. $\mathrm{TE_I}\ (x, y)=1$ is necessary, but not sufficient, for $x\in\mathrm{Eff}\ L(y)$. Sufficiency fails because $\theta$ is an equi-proportionate (radial) measure that leaves non-proportional efficiency undetected. A score of unity indicates technical efficiency because no equi-proportionate input reduction is feasible, and a score less than unity indicates the severity of technical inefficiency.

*Efficiency measurement*

Assume the ICU being evaluated as having data $(x^0, y^0)$, and consider the problem, where $x^i \in \mathbb{R}^N_+$ and $y^i \in \mathbb{R}^M_+$, and $i = 1..., I$, with I indicating the number of ICUs in the sample:

$$\min_{\mu,v} v^T x^0 / \mu^T y^0$$

subject to

$$v^T x_i / \mu^T y^i \geq 1 \quad i = 1,\ldots,0,\ldots,I$$
$$\mu, v \geq 0$$

The minimization problem seeks a set of non-negative weights $(v, \mu)$ which, when applied to the inputs and outputs of the ICU being evaluated, minimizes the ratio of weighted input to weighted output, subject to the normalizing constraint that no ICU in the sample, including the ICU being evaluated, have a ratio less than unity when weights of the ICU being evaluated are applied.

In order to provide a linear programming representation of the input-oriented radial efficiency measure previously defined, this non-linear ratio model can be converted to a linear programming multiplier problem which dual is the input oriented CCR [22] DEA model:

$$\mathrm{TE_I}\ (x^0, y^0) = \min_{\theta,\lambda} \theta$$

subject to

$$\theta x^0 - X\lambda \geq 0$$
$$-y^0 + Y\lambda \geq 0$$
$$\lambda \geq 0$$

where $X$ is an $NxI$ input matrix with columns $x^i$, $Y$ is an $MxI$ output matrix with columns $y^i$, and where $\lambda$ is an $Ix1$ intensity vector.

Thus this problem provides a linear programming representation of the input-oriented radial efficiency measure given in the definition of Debreu–Farrell. The optimal value of $\theta$ provides a technical efficiency measure of the ICU being evaluated. Input-oriented radial efficiency requires $u^T y^0 = \theta = 1$. A ICU is judged to be technically inefficient if at optimum $\theta < 1$, and technically efficient if at optimum $\theta = 1$.

Input-oriented CCR DEA model incorporates the assumption of constant returns to scale in production. Banker *et al.* [23] (BCC) generalized the CCR formulation to allow for variable returns to scale by the inclusion of the convexity constraint $e^T\lambda = 1$, where $e^T$ is an $Ix1$ row vector of ones. Only convex combinations of ICUs are allowed to be created informing the production frontier.

The input and output-oriented CCR DEA models (constant returns to scale—CRS) measure overall technical efficiency (OTE). Input and output-oriented BCC DEA models (variable returns to scale—VRS) measure exclusively pure technical efficiency (PTE). Then, following Banker *et al.* [23], the ratio between the two measures of efficiency in CCR and BCC DEA models is a measure of scale efficiency (SE).

*Non-discretionary variables*

Some inputs or some outputs may be exogenously fixed or non-discretionary. Severity of illness of admitted patients is an example for ICUs. Banker and Morey [24] show how to incorporate non-discretionary variables into a DEA analysis. This model modifies the original BCC DEA model by inclusion of constraints on the fixed inputs: these constraints differ from constraints on the discretionary inputs with the exclusion of the efficiency term.

*Multiplier restrictions*

DEA analysis identifies relatively efficient ICUs assigning them a score equal to one. Some discrimination between efficient ICUs may be necessary in order to ensure that relative efficiency is not simply a consequence of an unacceptable weighting structure given the absolute weight flexibility of the method. As pointed out by Thanassoulis *et al.* [10], total flexibility is not wholly appropriate when output quality measures are present in the model. Thus, it is possible to incorporate some prior structure of the relative importance of inputs and outputs to avoid spurious efficiencies by restricting weight flexibility.

*Categorical variables*

An input (or an output) may be represented as a binary factor reflecting the presence or absence of a particular characteristic. As an example to apply to our study, consider a critically ill patient with a very low or a high survival probability. We would like to ensure that the management of this patient is compared with patients in the same category. Patients in category *j* are compared to the *j* production frontier [25]. A linear programming algorithm allowing multiple categorical variables is presented in Charnes *et al.* [26].

## DATA AND VARIABLES

The data set used in this paper is based on 993 patients from sixteen ICUs in Catalonia (Spain) that participated in the European and North American Study of Severity of illness of ICU patients (ENAS), whose objective was to update the MPM and the Simplified Acute Physiology Score (SAPS) systems. Data were collected on adult consecutive admissions excluding coronary care, burn and cardiac surgery patients, between October 1991 and February 1992. Methodology details appeared in Lemeshow *et al.* [27] and Le Gall *et al.* [28]. Complementary data relative to resource utilization in the ICUs are used from an inquiry about UCIs in Europe conducted in 1992. An annual survey of Spanish hospitals was used for additional data about complete hospital services.

Catalan Intensive Care Units (UCIs) included in this study have a mean of 14.4 beds, ranging from five to 35 beds. These units represent, on the average, 3% of the total beds in their respective hospitals, with upper and lower ranges of 4.5 and 1.3%, respectively. The number of nurses per patient is close to two, and is fairly homogenous. In comparison with the ICUs included in the ENAS study, the mean ICU stay in Catalonia is longer. The patients admitted to the Catalan ICUs are similar to the rest of ICUs in the study with regard to severity, as measured by the SAPS II, as well as with regard to diagnostic distribution. Hospital mortality, however, is higher. The type of patients admitted is similar to that of the rest of the ICUs in the study, although there are less surgical patients.

A major problem in efficiency analysis of health care providers is the difficulty of appropriately measuring the presence of the patient in the input (severity of illness) and the output set (improved health status). This problem represents that productive efficiency literature usually restricts performance measures to the production of intermediate outputs (activity). Measurement of efficiency in health services is biased by the way that the quality dimension of output is measured. To the extent that outputs are measured with some error, inefficiency could simply represent difficulties in measuring output and adjusting for quality.

The present survey measures severity of illness by the probability of hospital mortality at admission to the ICU (MPM $II_0$). MPM value is calculated from a set of fifteen clinical variables. Variables considered by MPM II at admission are: cirrhosis, metastatic cancer, chronic renal insufficiency, heart rate, systolic blood pressure, presence of coma, acute renal failure, cardiac dysrhytmia, cerebrovascular incident, gastrointestinal bleeding, intracranial mass effect, age, type of admission, cardiopulmonary resuscitation prior to ICU admission and mechanical ventilation. The relatively small number of variables minimizes the burden of data collection and the potential for error. The number of variables is fewer than for either of the other two alternative systems (APACHE and SAPS), and there is less reliance on physiological assessments that require laboratory tests results. Outcome of critical patients after hospital treatment is measured by survival status at discharge.

Severity measures have tended to focus exclusively upon the risk of one particular outcome, death. Existing scoring systems equate the severity of illness with the risk of mortality in ICU patients. The most validated of these scoring systems for ICU patients are the Acute Physiology and Chronic Health Evaluation (APACHE), SAPS II and MPM II. The admission model MPM $II_0$ calibrated well in developmental and validation samples (goodness-of-fit tests: $p = 0.623$ and $p = 0.327$, respectively, where a high $p$-value represents goodness-of-fit between observed and expected values) and discriminated well (area under the receiver operating characteristic curve = 0.837 and 0.824, respectively). All scoring systems are based on rigorous research and reported performance is good. However, we favoured MPM as a measure to be included in our input data set as the most appropriate to test technical inefficiency because the $MPM_0$ is the only available model for use at ICU admission. ICU efficiency measurement requires a measure of severity independent of ICU treatment. The MPM $II_0$ is the only model currently available for assessing a patient's severity of illness at the time the patient enters the ICU and is independent of ICU management.

 

The performance of the severity systems APACHE III, SAPS II and MPM II has been compared using the same subset of patients [29]. All three systems perform well as reflected by large receiver operating characteristic areas and good calibration.

A vector of seven inputs has been specified for every patient in the analysed sample: survival probability at admission ($X_1$), mortality risk level ($X_2$), weighted ICU days ($X_3$), non-ICU hospital days ($X_4$), available nurse days per patient ($X_5$), available physician days per patient ($X_6$), and technological availability ($X_7$). The input set chosen in this analysis includes two different input types: patient illness characteristics and clinical practice characteristics.

The first group might measure patient health status at the hospital admission, which is beyond the influence of clinical management. We consider the hypothesis that the ICU can not select illness severity of admitted patients, since severity is an exogenous factor to the ICU management beyond the ICUs ability to determine. The two selected variables to reflect patient health status prior to hospitalization are computed from the MPM II at hospital admission. The first one, survival probability at admission ($X_1$), is defined for every patient as $1 - \text{MPM II}_0$ score, where the latter is considered as a non-discretionary variable which serves as a proxy for severity of illness. The second one, the mortality risk level at admission ($X_2$), defines 3 patient categories: patients with low risk ($\text{MPM II}_0 \leq 0.2$), patients with moderate risk ($0.2 < \text{MPM II}_0 \leq 0.5$) and patients with high risk ($\text{MPM II}_0 > 0.5$). Information added by this latter input is essentially not different from $X_1$, although it is considered as a categorical variable in the DEA model in order to allow comparisons of patients only within the same risk level category. Thus, differences in the efficiency scores which could be exogenously attributed to treating patients belonging to different risk categories are controlled by the model.

A second group of inputs which measures clinical practice characteristics is defined. These might include physician and hospital resources available for every patient. Nevertheless only rough measures of these resources are available. We use two types of variables as proxies for these resources: a direct type of measures, as weighted patient days in ICU and hospital; and an indirect type of measures defined by the availability of resources (labour, technology) during the hospital stay. That is, weighted ICU and hospital stays are used as discretionary inputs in the DEA model as proxies for individual consumption of resources without considering possible differences in the input intensity for every patient day between hospitals (differences in the intensity of inputs for every day of stay for the same patient are considered under the weighting structure). Input intensity in every patient day may be proxied by the availability of the different inputs (nurses, physicians, technology). These variables are considered as non-discretionary variables in the DEA model, which indicates that they can not be modified when taking decisions to treat a patient. Thus, the efficiency measure obtained in this study will be a measure of short run efficiency: Given the level of labour (nurses and physicians) and technology available in the hospital, which is the efficiency level of clinical management in the treatment of a critically ill patient?

Weighted ICU and hospital days ($X_3$ and $X_4$) are a length of stay measure that distinguishes between ICU days and days of hospitalization after the ICU discharge as well as between medical and surgical patients. For medical patients, the first ICU day is weighted as 3, each subsequent ICU day is weighted as 2 and each post-ICU day is weighted 1. For surgical patients, the first ICU day is weighted as 4, the second ICU day is weighted as 3, each subsequent ICU day is weighted as 2 and each post-ICU day weights 1. These weights have been used in previous studies about resource consumption in ICUs [19].

Labour resources availability measures are defined as the average number of nurse ($X_5$) and physician ($X_6$) equivalent full days which a patient has had available during his ICU and hospital stay. Lack of individual input use data has been replaced by the use of available inputs computed as the number of whole time equivalent employees days divided by the number of weighted ICU days. Technological availability ($X_7$) is measured as the proportion of 33 technological items available in the ICU expressed as a percentage. The 33 items of equipment are given equal weight in calculating $X_7$ given the absence of criteria or data allowing aggregation.

The output set might measure observed health status after discharge, considering both increased quantity and quality of life dimensions. Nevertheless, only rough measures of quantity and quality dimensions of outcome are available as proxies for adequate output measures in the model. Two different discretionary outputs are defined for every critically ill patient in the model: the number of days surviving in the hospital ($Y_1$), and the surviving discharge status ($Y_2$). Survival status is defined as a binary variable with two possibilities: death or surviving at hospital discharge. For a patient not surviving at discharge, the only output considered is the number of days the patient has survived in the hospital.

$Y_1$ may also be viewed as a measure of hospital activity (length of stay), an output measure being traditionally considered in DEA models measuring hospital efficiency. When the patient has survived hospital discharge, we additionally consider the surviving status as a second output, and probably the most important. Both output variables reflect differences in quality of life for in-hospital and after discharge days. Death or survival may be a reasonable proxy for quality of care when we are considering the management of severely ill patients.

Thanassolulis *et al.* [10] have illustrated the need to restrict weight flexibility in DEA models when introducing quality variables in the measurement of perinatal care efficiency. In our case, flexible weights may clearly result in unacceptable efficiency scores. First, given the input–output set definition, all cases may be deemed highly efficiently managed by assigning a high weight to $Y_1$ and a null weight to $Y_2$, i.e. considering activity only as an output (patient days or simply discharge) and omitting activity outcomes. Second, social output values or preferences must be considered, before total weight flexibility, when assigning weights to the importance of a survival in contrast to merely surviving as an inpatient, or death discharge. Social values would express preferences over output variables when there are no observable prices for outputs.

We assign greater preference or social value to survival at hospital discharge in relation to surviving inpatient days (as given in relation to dead status) by defining a relative output constraint: $\mu_{Y_2}/\mu_{Y_1} > \delta$. The problem of weight flexibility is even more important when data is restricted to death or survival at discharge given that the number of days a patient survives after discharge is unknown, as in this study. However, if the number of days a patient is expected to survive after discharge were known, there would also be need to restrict flexibility to weight pre- and post-discharge weights. The problem is the value to assign to $\delta$, in absence of reasonably rooted criteria. We arbitrarily define values between 1 and 100 to test for differences in the scores. As the value of $\delta$ increases, the greater value assigned to a survival discharge relative to an inpatient day.

## THE EFFICIENCY SCORES

DEA analysis identifies relatively efficient ICUs, assigning them a unitary score. In this study various DEA models were analyzed reflecting different output weight structures with values of $\delta$ from 1 to 100. Descriptive statistics for DEA overall technical efficiency (OTE) scores obtained from different values of $\delta$ are presented in Table 1. Two obvious considerations arise from this results. First, inefficiency level is extremely sensitive to $\delta$ value, which means that this level depends, in an important way, on the social relative weight attached to survival status at discharge in comparison to death at discharge. Second, efficiency scores are a decreasing function of preference or social weight attached to survival status at discharge. If efficiency level is a function of the unknown relative output weight when weight flexibility is unacceptable, DEA scores would be interpreted more as a rank order than as an absolute value classification.

We evaluate the nature of the efficiency density distributions of the efficiency measures obtained from different values of $\delta$ using two non-parametric tests statistics. First, a Friedman two-way analysis of variance based on ranks was computed to test whether the efficiency measures are all drawn from the same population (or at least from populations with the same median). Second, a Wilcoxon test signed rank test does the same for all pairs of efficiency measures. The Friedman-two way test suggests that the efficiency measures are drawn from a single population. The

Table 1. Overall technical efficiency scores from different weight structures

| Model | $\mu_{Y2}/\mu_{Y1}$ | Mean | S.D. | Min | Max |
|---|---|---|---|---|---|
| 1 | 1 | 0.974 | 0.059 | 0.600 | 1.000 |
| 2 | 10 | 0.837 | 0.225 | 0.179 | 1.000 |
| 3 | 20 | 0.751 | 0.283 | 0.104 | 1.000 |
| 4 | 50 | 0.623 | 0.329 | 0.050 | 1.000 |
| 5 | 100 | 0.533 | 0.332 | 0.024 | 1.000 |

 

Table 2. Average DEA efficiency scores ($\delta = 10$)

| Efficiency measure | Mean | S.D. | Min | Max |
|---|---|---|---|---|
| Overall technical efficiency (OTE) | 0.837 | 0.225 | 0.179 | 1.000 |
| Pure technical efficiency (PTE) | 0.925 | 0.160 | 0.299 | 1.000 |
| Scale efficiency (SE) | 0.898 | 0.164 | 0.179 | 1.000 |

Wilcoxon matched-pairs signed-ranks test indicates that all pairs of efficiency measures shares a common distribution.

Table 1 shows that overall technical efficiency decreases rapidly as more relative weight is attached to patient survival at discharge, being only 0.533 when $\delta$ equals 100. We present DEA efficiency scores obtained from $\delta = 10$ in Table 2 as an illustration of the observations that could be inferred using this method with clinical patient data.

Table 2 summarizes the efficiency scores for overall technical efficiency, pure technical efficiency and scale efficiency. Results from DEA model 2 show an average overall technical efficiency of 0.837. That is, ICUs and hospitals use on average 16.3% more inputs than would be necessary if all of them were operating in the efficiency frontier. The average overall efficiency scores range from 0.225 to 1. Pure technical inefficiency scores show a lower level of inefficiency, with 7.5% as the average. Average scale inefficiency is 10.2%. Scale inefficiency may be interpreted, in this context, as inefficiency attributed to decreasing returns to scale which appear as the level of resources devoted to a patient is increased.

For overall technical efficiency the percentage of patients managed efficiently (on the production frontier) is 15.4%. The average efficiency score for non-frontier hospitals is 0.807, implying that non-efficiently managed patients received on average 19.3% more inputs per unit of output than efficiently managed ones. However, according to pure technical efficiency criterion, 76.7% of the patients were managed efficiently, with an average efficiency score of 0.678 for non-frontier managed patients.

Analysis of overall technical inefficiency permits the estimation of possible discretionary input savings. Possible savings represent a complementary measure of technical inefficiency. If it were possible for the inefficiently managed patient be managed like those in the best-practicing facilities, a savings of 13.4% in the number of weighted ICU days and of 6.8% in the number of hospital days would be possible.

Assuming that efficiency in the management of critically ill patients is the ICU responsibility, ICU efficiency can be obtained as an average of the efficiency with which the patients have been treated. An ICU will be considered as efficient only if all patients treated in this hospital are considered efficiently managed when compared with patients in this or in other hospitals in the same mortality risk level.

The DEA model defined in this paper makes it possible to obtain efficiency scores for patients grouped according to their mortality risk at admission for the overall sample and for every ICU. Tables 3 and 4 present OTE scores for low, moderate and high risk patients. Results in these tables show that efficiency scores are not homogenously distributed among the three risk levels. A higher proportion of patients are managed efficiently at low and moderate risk patients in comparison with patients at high risk levels. Econometric tests have been applied to verify that there is no relation between efficiency scores and risk of mortality inside every one of the three risk groups of patients A total of 87.8% of low risk patients are treated with an efficient score equal to or greater than 0.8, but in the case of high risk patients this proportion is only 39.8%. Less efficiently managed patients (efficiency scores less than 0.5) account for 4.4% of low risk patients and for 17.9% of moderate risk patients, but they account for 40.4% of high risk patients. As efficiency scores are obtained from comparing patient management in the same risk group, these results may indicate that higher risks are treated less efficiently, i.e. increasing resources without obtaining a survival discharge. Differences in efficiency between risk groups widen as $\delta$ value increases. This observation may indicate that higher risk patients may have been treated without taking into consideration the survival probability. Or, in the context of our input–output data set, valuing more in-hospital surviving days than surviving at hospital discharge. This could be the case of units applying only palliative measures.

 

Table 3. Overall technical efficiency by mortality risk level ($\delta = 10$)

| Score (%) | Low risk (%) | Moderate risk (%) | High risk (%) |
|---|---|---|---|
| <0.5 | 4.4 | 17.9 | 40.4 |
| 0.500–0.799 | 7.8 | 15.3 | 19.8 |
| 0.800–0.999 | 71.4 | 50.7 | 29.2 |
| 1.000 | 16.4 | 16.1 | 10.6 |

Diagnostic case-mix may be influential and its effect not perfectly captured in this context. However, case-mix effects are controlled in this study in several ways. First, MPM $II_0$ scores explicitly take into account diagnostic information. Second, there is no significant correlation between diagnostic category and MPM $II_0$ scores. And third, differences in efficiency scores are not explained by diagnostic categories (see Section 5).

These differences between risk groups were tested for statistical significance with the non-parametric Mann–Whitney test and Kruskal–Wallis test. The null hypothesis that the distribution of efficiency measures are the same for the three risk groups (and for all pairs of risk groups) is rejected at the 99% level of confidence. Differences between for-profit and not-for-profit hospitals, teaching and non-teaching hospital status, and surgical and medical admission were also tested for statistical significance using the Mann–Whitney test; results show that in all three comparisons, the null hypothesis is not rejected.

In Table 5 we present results from computing DEA overall efficiency scores for every risk patient group and ICU. Average efficiency for management of low risk patients is 0.906, ranging from 0.833 and 0.973 between ICUs. In the case of moderate risk patients, average is 0.793 ranging from 0.544 and 0.899. And in the case of high risk patients, average is 0.637, ranging from 0.418 to 0.874 between hospitals. The range variation is greater as risk level increases. Note that ICU number 5 has the higher OTE mean score in Table 5, but it obtains this score managing low risk patients in a relatively efficient way and not treating patients at high risk.

Table 4. Descriptive statistics of overall technical efficiency scores by risk group

| Risk group | Mean | S.D. | Min | Max | $N$ |
|---|---|---|---|---|---|
| Low | 0.906 | 0.14 | 0.238 | 1.000 | 609 |
| Moderate | 0.793 | 0.25 | 0.179 | 1.000 | 223 |
| High | 0.637 | 0.29 | 0.179 | 1.000 | 161 |

## AN ECONOMETRIC ANALYSIS OF EFFICIENCY

What causes an ICU to manage critically ill patients using more than the minimum quantity of inputs for a specific vector of outputs? What are the factors associated with inefficiency?

In order to determine the influence of environmental variables on efficiency, we adopt a two-stage approach. Let $z_j \in \mathbb{R}_+$, $j = 1,..., I$ be a discrete or continuous environmental variable. $z_j$ represents variables over which the ICU has no control during the time period under consideration; $z_j$ and input variables should be uncorrelated. In the first stage inefficiencies are calculated using a DEA model in which the environmental variables are ignored. In the second stage, variation in calculated efficiencies is attributed to variation in operating environments by means of a

Table 5. Average overall technical efficiency scores by ICU and mortality risk level ($\delta = 10$)

| ICU | Low risk | Moderate risk | High risk | Mean |
|---|---|---|---|---|
| 1 | 0.929 | 0.739 | 0.636 | 0.775 |
| 2 | 0.903 | 0.800 | 0.534 | 0.819 |
| 3 | 0.861 | 0.770 | 0.546 | 0.812 |
| 4 | 0.942 | 0.891 | 0.725 | 0.906 |
| 5 | 0.973 | 0.819 | — | 0.959 |
| 6 | 0.928 | 0.733 | 0.628 | 0.826 |
| 7 | 0.916 | 0.843 | 0.706 | 0.870 |
| 8 | 0.924 | 0.766 | 0.418 | 0.841 |
| 9 | 0.961 | 0.899 | 0.578 | 0.850 |
| 10 | 0.855 | 0.871 | 0.876 | 0.879 |
| 11 | 0.916 | 0.778 | 0.523 | 0.791 |
| 12 | 0.963 | 0.753 | 0.749 | 0.857 |
| 13 | 0.881 | 0.705 | 0.712 | 0.840 |
| 14 | 0.873 | 0.766 | 0.809 | 0.818 |
| 15 | 0.886 | 0.871 | 0.610 | 0.835 |
| 16 | 0.833 | 0.544 | 0.639 | 0.736 |
| Average | 0.906 | 0.793 | 0.637 | 0.837 |

regression model. It is hypothesized that the DEA efficiency measures to be some function of the vector $z$ and a random disturbance $\epsilon^{i}$:

$$\theta_i = f(z_{ij})\epsilon_i$$

Some authors, such as Rosko *et al.* [30], Chilingerian [16] and Kooreman [9], have conceptualized DEA efficiency scores as presenting a censored normal distribution, i.e. the values of the dependent variable in the regression model above a threshold are measured by a concentration of observations at a single value. Therefore, ordinary least squares, as used by Chirikos and Sear [31], was not an appropriate method. Then, some authors [9,16,30] conclude that a censored Tobit model is appropriate in order to avoid biased estimates from ordinary least squares. The Tobit model is based on normally distributed latent variables. However, DEA scores do not fit the theory of sampling censoring that gives rise to Tobit models: the accumulation of sample observations at the highest level of efficiency is intrinsic to the model [16,14]. Tobit, and also Probit, estimates are inconsistent in the case of non-normality of the errors terms and/or error term heterogeneity. To overcome this problem, different alternatives have been proposed in the literature: Luoma *et al.* [32] used a test statistic sensitive for linearity of the fit, the type of heteroscedasticity which is related to the fit and excess skewness of the error term; González and Barber [14] estimated the model assuming different distributions of probability of disturbance, and check whether the differences in estimates are substantial; and Burgess and Wilson [33] remove the censoring problem directly through the addition of information on the distance of each observation from the technology represented by the other observations in the sample.

Banker and Johnson [34] propose a non-maximum-likelihood estimator in an empirical application, but one which was based on the theoretical conditions established by Banker [35] and was a consistent estimator. Given implausibility of the normally distributed latent efficiency required by Tobit models, and according to Banker and Johnson we make the assumption that the inefficiencies are log-normally distributed, and define the following transformations:

$$\bar{\theta}_i = 1/\theta_i - 1 + \omega$$

where $\omega$ is some very small amount. Then inefficiency can be posited to be a multiplicative function of the explanatory variables and random error term:

$$\bar{\theta}_i = \beta_0 \Pi_j z_j^{\beta_j}\, e^{v_i}$$

where $\beta_j$ and $\gamma_j$ are the parameters capturing the relationships between the explanatory variables and input and output inefficiency, respectively; and $\exp\{v_i\}$ is a random error term that is assumed to be independently and identically distributed and lognormally distributed with mean 1.

Three types of variables have been proposed for regressing technical efficiency scores of critically ill patient management. The first group includes variables that may represent factors beyond the control of ICU or hospital management and directly influenced by government regulation, such as ownership, market structure, and payment regulation. The second group of variables includes factors that are related to ICU management decisions in the short and medium terms, such as the number of ICU beds, the degree of specialization on more or less severely ill patients, the use of clinical guidelines in the UCI, the existence of evaluation programs for nurses and physicians, the number of daily visits per patient, etc. The third group of variables represent factors associated with the individual patient being treated in the ICU, which may be interpreted as factors out of control of ICU clinical decisions, and which, in some cases, may be partially interpreted as control variables to test the influence of input and output characteristics omitted or imperfectly measured in the DEA model. Examples include the risk group of the patient at admission, mortality probability variation during the first 24 h in the ICU, age, principal diagnosis leading to ICU admission and teaching status.

Specifically, the following variables have been employed to explain the variation in DEA overall technical efficiency scores of critically ill patients:

| | |
|---|---|
| PROFIT | Equals 1 for for-profit hospitals; 0 otherwise. |
| HHI | Herfindhal–Hirschman index of competition using discharges as output. |
| BEDS | Number of beds in the ICU where the patient was treated. |

| | |
|---|---|
| SPEC | Proportion of patients treated in the ICU that belong to the same risk group as the patient. |
| GUIDEL | Number of inpatient days in an ICU frequently using clinical guidelines. |
| EVAL | Number of inpatient days in an ICU with evaluation program for nurses and physicians. |
| VISITS | Number of daily visits. |
| TEACHING | Equals 1 for teaching hospitals; 0 otherwise. |
| MOD_ RISK | Equals 1 for patients with MPM0 score between 0.2 and 0.5; 0 otherwise. |
| HIGH_ RISK | Equals 1 for patients with MPM0 scores higher than 0.5; 0 otherwise. |
| VAR_24 | MPM 0 score divided by MPM24 score; |
| AGE | Age. |
| RESP | Equals 1 for nonoperative patients with respiratory failure or insufficiency; 0 otherwise. |
| CARD | Equals 1 for nonoperative patients with cardiovascular failure or insufficiency; 0 otherwise. |
| TRAUMA | Equals 1 for nonoperative patients with trauma; 0 otherwise. |
| URGENT | Equals 1 for postoperative patients with urgent admission; 0 otherwise. |
| PROGR | Equals 1 for postoperative patients with programmed admission; 0 otherwise. |

Overall technical efficiency scores were regressed against the explanatory variables. Different statistical tests show that the hypothesis that the inefficiency scores are log-normally distributed can not be rejected. The data were examined for evidence of collinearity and the residuals for evidence of non-linearity, non-normality and heteroscedasticity. Correlations between explanatory variables are not statistically significant and collinearity was not considered a problem. Also, non-linearity was not apparent. Kolmogorov–Smirnov test statistics for the normality of the residuals do no reject normality for the equations. Finally, Breucsh–Pagan test and others did not reject homoscedasticity. Regression results have been checked comparing efficiency scores using $\delta = 20$, 50 and 100, the results showing coincidence in sign and significance.

The key evidence from the regression model in Table 6 may be summarized in six main points. First, for-profit hospitals present a significantly higher technical inefficiency in comparison with public and not-for-profit hospitals. Thus, the profit motive does not appear to have improved efficiency. Evidence from empirical analysis of

Table 6. Factors explaining DEA inefficiency scores

| Factor | Coefficient | S.D. | $t$-Statistic |
|---|---|---|---|
| Constant | −1.461133 | 0.8405 | −1.738 |
| PROFIT | 0.745820 | 0.1720 | 4.335 |
| HHI | 0.062588 | 0.1495 | 0.0418 |
| BEDS | 0.277903 | 0.2413 | 1.151 |
| SPEC | −0.375298 | 0.4158 | −0.903 |
| GUIDEL | −1.025103 | 0.1727 | −5.935 |
| EVAL | −0.184560 | 0.0205 | −8.968 |
| VISITS | 0.312872 | 0.3831 | 0.817 |
| TEACHING | 0.662055 | 0.1365 | 4.849 |
| MOD_RISK | 0.511348 | 0.2176 | 2.350 |
| HIGH_ RISK | 1.138168 | 0.2611 | 4.359 |
| VAR_24 | 1.011843 | 0.1966 | 5.145 |
| AGE | −1.316737 | 0.2992 | −4.401 |
| RESP | 0.117625 | 0.1457 | 0.807 |
| CARD | 0.249975 | 0.1622 | 1.540 |
| TRAUMA | −0.044733 | 0.1902 | −0.235 |
| URGENT | 0.154914 | 0.1432 | 1.082 |
| PROGR | 0.558779 | 0.1412 | 3.957 |
| $R^2$ | 0.19404 | | |
| $F$ | 14.964 | | |
| $n$ | 993 | | |

hospital inefficiency by repeatedly using DEA on the same set of data by Grosskopf and Valdmanis [36] and Valdmanis [37,38] suggests that public hospitals are more technically efficient than non-profit and private ones. Register and Bruning [39], also using DEA, found no difference between non-profit and public hospitals when comparing technical efficiency. Ozcan *et al*. [40] and Ozcan and Luke [41] observed that US government hospitals tend to be more efficient and for-profit hospitals less efficient than other hospitals. Chirikos and Sear [31] conclude that for-profit hospitals are technically less efficient when they perform in less competitive markets. Also, Burgess and Wilson [33] reported average input oriented efficiency scores of for-profit hospitals lower than those of not-for-profit or local government hospitals. However, empirical evidence refers almost exclusively to the North American health care system, which might not be adequate for those European systems that have an effective NHS. Spanish private hospitals tend to be smaller and technologically less well equipped than the public and not-for-profit hospitals, whose major source of funding is the public sector. Average occupancy rates of Spanish private hospitals are lower than the other hospitals and they treat less complicated cases. Results indicate that efficiency in for-profit hospitals may be improved in comparison with not-for-profit hospitals when they treat critically ill patients with the same mortality probability at admission. The estimated coefficients corresponding to HHI and BEDS are insignificant. The Herfindahl–Hirschman index of competition is defined over a range between 0 and 10 000 such that increases correspond to decreases in hospital competition. These results do not show any influence of hospital size and the degree of specialization according to risk groups on efficiency.

Second, hospitals with program evaluation for nurses and physicians present a significantly higher level of efficiency. Our data report a hospital having a program evaluation for nurses and physicians when explicitly written criteria of performance are observed and when routine evaluation is reported by the ICU. This result suggests that quality of labor services improves efficiency of management in the ICUS. The results suggest that hospitals may be able to improve ICU efficiency by establishing accurate evaluation programs for nurse and physician activities.

Third, the estimated coefficient of TEACHING is positive and significant indicating a positive contribution to inefficiency. The results suggest that ICUs of teaching hospitals are less efficient than those of non-teaching hospitals, controlling for other factors. The input data set used in the DEA model does not include any measure of teaching activity, given the lack of reliable data. Then, regression results confirm that teaching activity is an omitted output in our efficiency measures, and that, in line with international findings of other researchers [42], teaching units imply an increased cost for the majority of hospital activities, including ICUs.

Fourth, the estimated coefficients corresponding to MOD_RISK and HIG_RISK are significant and positive, implying that patients with a higher mortality probability at ICU admission are associated with greater technical inefficiency. This result supports previous results in this paper showing significant differences in DEA efficiency scores between risk groups. The positive coefficient of both variables suggests that probability of efficient treatment decreases as risk increases. The HIG_RISK coefficient is higher than the MOD_RISK, indicating that inefficiency is greater in the first group. The estimated coefficient of VAR_24 is significant, with the sign indicating that decreasing (increasing) mortality probability during the first 24 h in the ICU lead to increased (decreased) efficiency. This result is in line with the effect of MOD_RISK and HIG_RISK variables. An increase in the risk of death during the first 24 h, which can not be completely exogenous to ICU management, results in a higher level of inefficiency. Alternatively, if the risk of death decreases during the first 24 h in the ICU, the patient is managed more efficiently. These are interesting results since they probably indicate that clinical decisions about allocation of resources are not at all related with the survival probability: more resources are allocated despite the low marginal probability of a contribution to health improvement when the admitted patient presents a high mortality probability. It seems that this behaviour is one of the more important determinants in explaining differences in efficiency between ICUs, since differences are more important when comparing patient management in moderate and high risk groups. When they treat low risk patients, efficiency is higher and differences between ICUs are reduced. These observations suggest that hospitals treating a higher proportion of less severely ill patients may be erroneously classified as being more efficient than those treating more high risk patients when severity is not adequately measured in efficiency measurement.

Fifth, the estimated coefficient corresponding to AGE is significant and positive. This result indicates that increasing patient age, given the probability of mortality at admission and other factors, leads to increasing efficiency. The result suggests a curious development about the pattern of clinical decisions in ICUs: patients with a higher mortality probability are managed less efficiently than patients with a lower probability; however, if they treat older persons with the same risk at admission, these patients are treated more efficiently. These results may be interpreted in the sense that ICU management assigns more resources to young patients than to aged patients for the same risk of death. Thus, for a given MPM 0, inefficiency is lower when the age of the patient is higher. Results may indicate that clinical decisions assign a higher value to prolonging the life of young patients.

Sixth, diagnostic variables are not significant, after controlling for other factors, except for the PROGR variable. The estimated coefficients of RESP, CARD, TRAUMA and URGENT do not show any diagnostic contribution to explaining efficiency variations. However, post-operative patients with programmed admission are treated less efficiently than post-operative patients with urgent admission. This is a curious result indicating that ICUs with a higher proportion of programmed admissions are less efficient than those with a lower proportion. The result suggests that, for a given probability of mortality, a greater volume of resources is allocated to programmed patients than to urgent ones, with little effect on the output.

## CONCLUSIONS

In this paper technical efficiency in the management of critically ill patients has been analyzed by means of an extended version of DEA (non-discretionary and categorical variables, and weight constraints under consideration). The paper has proposed a limited set of seven inputs and two outputs for use in measuring the relative efficiency, including mortality risk level at admission and survival status at discharge. Efficiency measured in this context may be interpreted as a short-term measure, given that only some inputs are considered as discretionary ones. Usefulness of DEA in the measurement of technical efficiency of clinical decisions at the patient level has been illustrated. In this context, an ICU is only considered efficient if all patient treatments are considered efficient. And patient management is considered efficient if it compares favorably with an established comparison group which includes all similar patients in the same and in other hospitals.

According to our results, three main findings may be emphasized. First, the introduction of output weight structures reflecting social values or preferences over the outputs in the DEA model results in an efficiency level being a function of the unknown relative weighting structure. In this sense, scores obtained under weight constraint would represent a 'mix' of technical and allocative efficiency. A clear limitation of the method appears in this context: DEA scores should be interpreted more as a rank order than as an absolute value. In this study, efficiency scores are a decreasing function of the weight attached to the survival of patients in relation to activity measures such as number of patient days or discharges. Weight flexibility in the estimation of health production in DEA models, as used in the majority of previous studies, may provide efficiency scores reflecting unacceptable or implausible relative shadow prices. Our results illustrate the need to restrict output weights in DEA models to obtain meaningful scores, and they also show that the results appear to vary markedly with the weighting structure. The latter fact emphasizes the interest to develop theoretical and empirical bases for the weighting structures and their implications.

Second, efficiency scores are not distributed homogenously in the three mortality risk levels. Higher risk patients are systematically managed less efficiently than lower risk ones. This evidence is in line with past literature observing that a large amount of resources was being devoted to more severe patients who died. Regression results confirm that inefficiency significantly increases when ICUs treat patients with higher levels of risk. These results indicate that changes in clinical decisions may improve efficiency, given that the present resource allocation decisions do not seem to be closely related to the expected outcome.

Third, our results prove the importance of factors related to clinical decisions in the efficiency of critically ill patient management. Some explanatory factors represent the importance of changes at the clinical level in the explanation of inefficiency. Results indicate that clinical decisions about allocation of resources are not at all related to the survival probability, but when an ICU treats older patients, controlling for the risk level and other factors, they are treated more efficiently. Efficiency of clinical management may be improved by the frequent use of clinical guidelines. Also, efficiency may be improved by accurately revising care received by patients with programmed admission.

Additional limitations of DEA frontier estimation are related to the existence of omitted outputs or inputs not measured; and to the assumption of no measurement error or no random fluctuations in the input–output set. Although these problems have been managed in this paper through the two stage approach, this research might be extended in several ways. Input and output variables might be improved, especially output measures to more accurately take into account the quality dimension of health. It is essential that the relationship between quality and input volume be accounted for in the measurement of inefficiency. The method employed in this paper might be used to estimate allocative and cost efficiency rather than only limiting attention to technical efficiency. Although we focus on ICU efficiency, we do not pay attention to whether or not patients receive an appropriate amount of care.

## ACKNOWLEDGEMENTS

This study has been financed by The BBV Foundation. I would like to thank M. Rué, from the Department of Epidemiology, Consorci Hospitalari Parc Taulí, Sabadell (Spain), for providing and assessing the use of clinical patient data. I thank the anonymous referees for useful comments.

## REFERENCES

1. Russell, L. *Technology in hospitals: medical advances and their diffusion*. Washington: Brookings Institution, 1979.
2. Jacobs, P. and Noseworthy, T.W. National estimates of intensive care utilization and costs: Canada and the United States. *Critical Care Medicine* 1990; **18**: 1282.
3. Farrell, M.J. The measurement of productive efficiency. *Journal of the Royal Statistical Society Series A* 1957: **120**(3): 253–281.
4. Debreu, G. The Coefficient of resource utilization. *Econometrica* 1951; **19**: 273–292.
5. Land, K.C., Lovell, C.A.K. and Thore, S. Chance-constrained data envelopment analysis. *Managerial and Decision Economics* 1993; **14**: 541–554.
6. Olesen, O.B. and Petersen, N.C. Chance constrained efficiency evaluation. *Management Science* 1995; **41**: 442–457.
7. Gong, B.H. and Sickles, R.C. Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data. *Journal of Econometrics* 1992; 259–284.
8. Resti, A. Evaluating the cost-efficiency of the Italian Banking System: What can be learned from the joint application of parametric and non-parametric techniques. *Journal of Banking and Finance* 1997; **21**: 221–250.
9. Kooreman, P. Nursing home care in The Netherlands: a nonparametric efficiency analysis. *Journal of Health Economics* 1994; **31**: 301–316.
10. Thanassoulis, E., Boussofiane, A. and Dysson, R.G. Exploring output quality targets in the provision of perinatal care in England using data envelopment analysis. *European Journal of Operational Research* 1994; **80**: 588–607.
11. Byrnes P. and Valdmanis V. Analyzing technical and allocative efficiency of hospitals. In: Charnes A., Cooper W., Lewin A.W., Seiford L.M. (eds.). *Data envelopment analysis: theory, methodology and application*. Boston: Kluwer, 1994: 129–144.
12. Burgess, J.F., Wilson, P.W. Hospital ownership and technical inefficiency. *Management Science* 1996; **42**: 110–123.
13. Magnussen, J. Efficiency measurement and the operationalization of hospital production. *Health Services Research* 1996; **31**: 21–37.
14. González, B. and Barber, P. Changes in the efficiency of Spanish public hospitals after the introduction of program-contracts. *Investigaciones Económicas* 1996; **XX**: 377–402.
15. Dalmau, E. and Puig-Junoy, J. Market structure and hospital efficiency: evaluating potential effects of deregulation in a national health service. *Review of Industrial Organization* 1997; (forthcoming).
16. Chillingerian, J.A. Evaluating physician efficiency in hospitals: A multivariate analysis of best practices. *European Journal of Operational Research* 1995; **80**: 548–574.
17. Finkler, M.D. and Withschafter, D.D. Cost-effectiveness and data envelopment analysis. *Health Care Management Review* 1993; **18**: 81.
18. Knaus, W.A., Wagner, D.P., Zimmerman, J.E. and Draper, E.A. Variations in mortality and length of stay in intensive care units. *Annals of Internal Medicine* 1993; **118**: 753–761.
19. Rapoport, J., Teres, D., Lemeshow, S. and Gehlbach, S. A method for assessing the clinical performance and cost-effectiveness of intensive care units: A multicenter inception cohort study. *Critical Care Medicine* 1994; **22**(9): 1385–1391.
20. Shortell, S.M. The performance of intensive care units: Does good management make a difference? *Medical Care* 1994; **32**(5): 508–525.

21. Bosmans, N. and Fecher, F. Performance of Belgian Hospitals: A Frontier Approach. *Health Economics* 1995; **4**: 389–397.
22. Charnes, A, Cooper, W.W. and Rhodes, E.L. Measuring the efficiency of decision making units. *European Journal of Operational Research* 1978; **2**: 429–444.
23. Banker, R.D., Charnes, A. and Cooper, W.W. Some models for estimating technical and scale efficiencies in data envelopment analysis. *Management Science* 1984; **30**(9): 1078–1092.
24. Banker, R.D. and Morey, R.C., Efficiency analysis for exogenously fixed inputs and outputs. *Operations Research* 1986; **34**(4): 513–521.
25. Kamakura, W.A. A note on the use of categorical variables in data envelopment analysis. *Management Science* 1988; **34**(10): 1273–1276.
26. Charnes, A., Cooper, W., Lewin, A.Y. and Seiford, L.M. *Data envelopment analysis. Theory, methodology and applications*. Boston: Kluwer, 1994.
27. Lemeshow, S., Teres, D, Klar, J., Spitz, J., Gehlbach, S.H. and Rapoport, J. Mortality Probability Models (MPM II) based on an International Cohort of Intensive Care Unit Patients. *Journal of the American Medical Association* 1993; **270**(20): 2478–2486.
28. Le Gall, J.R., Lemeshow, S. and Saulnier, F. A New Simplified Acute Physiology Score (SAPS II) Based on a European/North American Multicenter Study. *Journal of the American Medical Association* 1993; **270**(24): 2957–2963.
29. Castella, X. *et al*. A Comparison of severity of illness scoring systems for intensive care unit patients: results of a multicenter study. *Critical Care Medicine* 1995; **23**: 1237.
30. Rosko, M.D., Chilingerian, J.A., Zinn, J.S. and Aaronson, W.E. The effects of ownership, operating environment, and strategic choices on nursing home efficiency. *Medical Care* 1995; **33**: 1001.
31. Chirikos, T.N. and Sear, A.M. Technical efficiency and the competitive behavior of hospitals. *Socio-Economic Planning Science* 1994; **28**(4): 219–227.
32. Luoma, K., Jarvio, M-L., Suoniemi, I. and Hjerppe, R.T. Financial incentives and productive efficiency in finnish health centres. *Health Economics* 1996; **5**: 435–445.
33. Burgess, J.F. and Wilson, P.W. Variation in Inefficiency among US hospitals. Unpublished 1996.
34. Banker, R.D. and Johnston, H.H., Evaluating the impacts of operating strategies on efficiency in the US airline industry. In: Charnes A., Cooper W., Lewin A.Y., Seiford L.M. (eds.). *Data envelopment analysis: theory, methodology and application*. Boston: Kluwer 1994: 97–128.
35. Banker, R.D. Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. *Management Science* 1993; **39**: 1265–1273.
36. Grosskopf, S. and Valdmanis, V. Measuring hospital performance: A nonparametric approach. *Journal of Health Economics* 1987; **5**: 107–127.
37. Valdmanis, V. Ownership and technical efficiency of hospitals. *Medical Care* 1990; **28**: 552–561.
38. Valdmanis, V. Sensitivity analysis for DEA model. *Journal of Public Economics* 1992; **48**: 185–205.
39. Register, C.A. and Brunnin, E.R. Profit incentives and technical efficiency in the production of hospital care. *Southern Economic Journal* 1987; **53**: 899–914.
40. Ozcan, Y.A., Luke, R.D., Haksever, C. Ownership and organizational performance. A comparison of technical efficiency accross hospital types. *Medical Care* 1992; **30**(9): 781–791.
41. Ozcan, Y.A. and Luke, R.D. A national study of the efficiency of hospitals in urban markets. *Health Services Research* 1993; **February**: 719–739.
42. Wagstaff, A. and Lopez, G. Hospital costs in Catalonia: a stochastic frontier analysis. *Applied Economics Letters* 1996; **3**: 471–474.