

**EVALUATING HOSPITAL
EFFICIENCY ADJUSTING FOR
QUALITY INDICATORS: AN
APPLICATION TO PORTUGUESE
NHS HOSPITALS**

ÁLVARO ALMEIDA ^{1 2}
JOSÉ PEDRO FIQUE ¹

¹ FACULDADE DE ECONOMIA, UNIVERSIDADE DO PORTO
² CEFUP

Evaluating Hospital Efficiency Adjusting for Quality Indicators: an Application to Portuguese NHS Hospitals

Álvaro Almeida

CEF.UP and Faculdade de Economia, Universidade do Porto †

José Pedro Figue

Faculdade de Economia, Universidade do Porto and LIAAD, INESC-Porto *

October 10, 2011

† Email: almeida@fep.up.pt - Adress: Rua Dr. Roberto Frias, 4200-464 Porto. CEF.UP, Centre for Economics and Finance at University of Porto is supported by the Foundation for Science and Technology (FCT), Portugal.

* Email: 090421010@fep.up.pt – Adress: Rua Dr. Roberto Frias, 4200-464 Porto. Faculdade de Economia - Economics PhD Program, Portugal. The author acknowledges support from the Foundation for Science and Technology (FCT) Ph.D. scholarship SFRH/BD/62309/2009.

Abstract

The objective of this paper is to develop a methodology to incorporate measures of hospital quality in efficiency analysis, applied to Portuguese NHS hospitals, in order to assess whether there is a trade-off between efficiency and quality in Portuguese hospitals. We develop and compare two methodologies to compute DEA technical efficiency scores adjusted for output quality, for a sample of Portuguese NHS hospitals in 2009.

When DEA efficiency scores are adjusted for output quality, the decision making units that lie on the technical efficiency frontier remain largely unaltered, even if a great weight is given to quality indicators over quantity indicators of output. Nevertheless, we find that outside of the frontier adjusting for quality does have an impact in efficiency scores.

We conclude that the empirical evidence is not sufficient to identify a clear trade-off between efficiency and quality in the hospitals under review, implying the possibility that efficiency gains may be achieved without a significant sacrifice of service quality. Nevertheless, there is enough evidence to conclude that analyzing hospital efficiency without consideration of differences in quality of service will generate biased results. When perceived quality is brought to the analysis, the gap between efficient and inefficient units tends to widen.

Keywords: Hospital efficiency, Hospital quality, Data Envelopment Analysis

JEL Classification: I - I-1 - I-18

1 Introduction

In an era of resource constraints, there is an increasing interest of hospital managers and health administration authorities in designing methods to evaluate hospital performance. The ability to rank efficient hospitals over their inefficient counterparts provides a benchmark for hospital managers to discover and reduce potential inefficiencies, and provides health administration authorities with measures that may be used to reward good managers. With this increasing interest in hospital performance, a vast academic literature has emerged on measures and comparisons of hospital efficiency, where quantitative measures of inputs are compared with quantitative measures of outputs, usually using Data Envelopment Analysis (DEA) or Stochastic Frontier Analysis (SFA).

However, many have expressed concerns that this increased focus on efficiency may induce managers to neglect service quality. Since health outcomes depend on the quantity, but also on the quality, of the healthcare services provided, it is possible that a hospital may increase its efficiency ranking with a deterioration of health outcomes, if there is an efficiency/quality trade-off.

The objective of this paper is to develop a methodology to incorporate measures of hospital quality in efficiency analysis, applied to Portuguese NHS hospitals, in order to assess whether there is a trade-off between efficiency and quality in Portuguese hospitals.

We develop and compare two methodologies to compute DEA technical efficiency scores adjusted for output quality, for a sample of Portuguese NHS hospitals in 2009. The quality of Portuguese hospital healthcare services is measured by two types of indicators. The first set of indicators is based on data from a 2009 survey of patients, designed by the ACSS (Administração Central do Sistema de Saúde) and Universidade Nova, whose main goal is to provide an independent system of regular evaluation of patient satisfaction and of hospital quality, as perceived by users of Portuguese NHS hospitals, the “Sistema de Avaliação da Qualidade Apercebida e da Satisfação dos Utentes dos Hospitais EPE e SPA 2009”. The second set of indicators is related to the timeliness of care, as measured by the waiting lists for elective surgery.

The rest of this paper is organized as follows. In section 2 we provide a brief review of the relevant literature. Section 3 describes the data used in this paper, while the methodology is

presented in section 4. Our results are in section 5. Section 6 presents is dedicated to a robustness analysis. Section 7 concludes.

2. Literature review

There is a vast literature on measurement of efficiency and productivity of health care organizations. Hollingsworth (2008) provides a review of 317 published papers on frontier efficiency measurement in healthcare, concluding that even though there is an increasing use of parametric techniques, such as stochastic frontier analysis, around three quarters of the papers use nonparametric data envelopment analysis. Studies of hospital efficiency are the most common (52% of the papers reviewed in Hollingsworth, 2008).

Hollingsworth (2008) does not cover any paper studying the efficiency of Portuguese hospitals, even though it includes 3 papers on the efficiency of OECD health systems, where Portugal is one of the countries mentioned (Afonso and St Aubyn 2005, Bhat 2005, Retzlaff-Roberts et. al 2004). Nevertheless, there are several studies on the efficiency and productivity of Portuguese hospitals.

Barros et. al (2007) use the directional distance function and the Luenberger productivity indicator to generate a productivity indicator that is decomposed into the usual constituents of productivity growth: technological change and efficiency change. They conclude that Portuguese hospitals experienced very weak productivity growth and low incidence of technological change in the period from 1997 to 2004. Harfouche (2008) used DEA to evaluate the impact of changes in the hospital management model in the technical efficiency level, having concluded that the new public-enterprise hospitals were more efficient than the traditionally managed hospitals. Rego et. al (2010) use DEA to investigate the efficiency of a set of public Portuguese hospitals. They compare the performance of 21 state-owned hospital enterprises with 38 traditional public administration sector hospitals, to conclude that the introduction of market processes and changes in organizational structure – such as managerial autonomy and corporatization of public hospitals – have had a positive impact on Portuguese public hospitals. Simões and Marques (2011) assess the performance of Portuguese hospitals and particularly the contribution of the congestion effect, using DEA and a double-bootstrap

procedure to take into account the influence of operational environment on efficiency, to conclude that there were significant levels of inefficiency in 68 major Portuguese hospitals for the year 2005 and more than half of them were found to be congested.

However, none of these papers take into consideration that output quality may vary across hospitals, and that traditional quantitative measures of output do not capture all the relevant dimensions of efficiency. The importance of incorporating quality measures in DEA hospital efficiency analyzes has been recognized in a few recent papers. Clement et al. (2008) include multiple quality indicators as outputs in a DEA efficiency study of 667 American hospitals, concluding that lower technical efficiency is associated with poorer risk-adjusted quality outcomes in the study hospitals. Nayar and Ozcan (2008), use DEA and a sample of Virginia hospitals to examine performance measures of quality and relate them to technical efficiency, having concluded that some of the technically efficient hospitals were performing well as far as quality measures were concerned. Valdmanis et al. (2008) apply DEA to input and output data from 1377 urban hospitals, include nurse-sensitive measures of quality, and conclude that higher quality in some dimensions of care need not be achieved as a result of higher costs or through reduced access to healthcare. More recently, Navarro-Espigares and Torres (2011), analyse the evolution of efficiency and quality in Andalusian Hospitals during the years 1997–2004 and rule out the existence of an efficiency–quality trade-off.

The international evidence seems to link poor quality outcomes to higher cost, which is confirmed by the Portuguese evidence presented in Gouveia et al (2006) using mortality as proxy for quality. However, there is no study where the quality proxies have a direct relationship with the quantitative output measures. This is the main contribution of this paper. Our paper is the first one, to our knowledge, to apply quality-adjusted DEA efficiency measures to Portuguese hospitals, allowing us to examine whether an efficiency / quality trade-off exists in the Portuguese hospital sector. Furthermore, our analysis allows us to investigate whether abstracting from a qualitatively oriented approach produces biased results.

3. Data

3.1 Data Sources

Since the objective of this paper is to analyze whether there is a trade-off between efficiency

and perceived quality/patient satisfaction the needed data is extracted from different sources.

We use cross-sectional data from a sample of 37 Portuguese hospitals for 2009. Although data were available also for oncology and psychiatric hospitals, they were not included given the highly specialized nature of these units. The choice of this particular sample rested on data availability for physical inputs and outputs. The sample is quite representative, since the hospitals in it were responsible for a very high share of the total production of Portuguese hospitals in the previous year (e.g., the hospitals in the sample were responsible for 80% of the inpatient visits, of emergency episodes and of outpatient visits and 95% of the surgery interventions).

To measure the impact of perceived quality/patient satisfaction in the DEA efficiency scores we resort to the report produced by the Assessment System of the Perceived Quality and Patient Satisfaction (Sistema de Avaliação da Qualidade Apercebida e da Satisfação do Utente dos Hospitais EPE e SPA - 2009)¹. This enquiry, conducted by ISEGI-UNL, is based on data collected from phone interviews from July 2009 until March 2010 on a sample of 28669 individuals. As a result four satisfaction indices were created covering inpatient visits, outpatient visits, emergency episodes and ambulatory surgery interventions. The methodology employed is compatible with the American Customer Satisfaction Index (ACSI).

Quality is also measured by the timeliness of care, as measured by the waiting lists for elective surgery. The variable used is the number of patients, LIC, registered in the waiting list of each hospital. The data is extracted from the report of the scheduled surgical activity (Relatório da Actividade em Cirurgia Programada ²).

3.2 DEA input measures

Following several papers on hospital efficiency (see Hollingsworth 2008), we use physical

¹ More information at Administração Central do Sistema de Saúde, Inquéritos de satisfação <http://www.acss.min-saude.pt/Direc%C3%A7%C3%B5eseUnidades/Gest%C3%A3odeRiscoeAuditoriaSNS/Inqu%C3%A9ritosdeSatisfa%C3%A7%C3%A3o/tabid/455/language/pt-PT/Default.aspx> (accessed in May 5, 2011)

² Portuguese Health Ministry, “Relatório da Actividade em Cirurgia Programada, 2009”, <http://www.min-saude.pt/NR/rdonlyres/0123250E-B4DE-4AC7-BB25-082701E40F3D/0/UCGICTL20100713RelatorioTP2009VE1.pdf> (accessed in May 5, 2011)

inputs data as a proxy for the labor and capital factor. The labor proxies are: (i) Doctors – number of doctors, (ii) Nurses – number of nurses and, (iii) OtherStaff – remaining staff at the units’ service. As a proxy for capital we use the number of beds (Beds). Alternatively, in the robustness analysis, we also look at total costs, in million euros, as described in the hospitals’ financial accounts.

3.3 DEA output measures

The output measures are comprised by inpatient visits, outpatient visits, emergency episodes, and surgery interventions (ambulatory plus non-ambulatory).

The interesting element of using this set of physical outputs is that we can get a direct correspondence of most of the outputs with the perceived quality/patient satisfaction indices. Table 1 provides the data descriptive statistics.

Inpatient visits are very heterogeneous; therefore, we adjust this variable by weighting it with a case-mix index (CMI). Since the latest CMI published by the System of Patient Classification in DRG is for 2006,³ we use a length of stay base case-mix index for 2009. To construct this index, we follow Herr (2008, pp. 1062-1063) by using the length of each inpatient visit as a proxy for the complexity of each case.

This index uses the length of stay as a proxy to the resource use associated with each diagnosis. The underlying idea is that hospitals with a higher incidence of diagnosis with high treatment duration have a higher case-mix index. To do so, the mean length of stay (los) of each main diagnosis $m = 1, \dots, M$ over all Portuguese hospitals $i = 1, \dots, N$ is computed:

$$\text{los}_m = \frac{1}{N} \sum_{i=1}^N \frac{\text{days}_{mi}}{\text{cases}_{mi}}. \quad (1)$$

Then, using equation (1), the *los* of each diagnosis is compared with the overall average *los* of all diagnoses:

$$\pi_m = \frac{\text{los}_m}{\text{los}_p}, \quad (2)$$

³ Administração Central do Sistema de Saúde, “Informação de Retorno do Sistema de Classificação de Doentes em Grupos de Diagnósticos Homogéneos (GDH) 2006” http://www.acss.min-saude.pt/Portals/0/Nacional_RR2006.pdf.

where $\text{los}_p = \frac{1}{M} \sum_{m=1}^M \text{los}_m$ with $\frac{1}{M} \sum_{m=1}^M \pi_m = 1$. Therefore, π_m can be used as a proxy for the complexity of diagnosis m , and with it an alternative case-mix index can be calculated.

Let us denote by $LCMI_i$ the length based case mix index of DMU i , such that using equation (2) we have:

$$LCMI_i = \sum_{m=1}^M \pi_m \frac{\text{cases}_{mi}}{\text{cases}_i}, \quad (3)$$

where $\text{cases}_i = \sum_{m=1}^M \text{cases}_{mi}$. Additionally, we normalize this index by dividing equation (3) by the average value of the index:

$$NLCMI_i = \frac{LCMI_i}{\frac{1}{N} \sum_{i=1}^N LCMI_i}. \quad (4)$$

The data for inpatient visits came from the official registry of inpatient episodes and outpatient surgeries in Portuguese NHS hospitals, held by the Central Administration of the Health System (ACSS), for 2009. These records are coded according to the international classification of diseases ICD-9-CM and also classified through the Portuguese DRG system, which is based on the version 21.0 of the AP-DRG grouper. The dataset included 930878 episodes, which were classified under 26 different diagnoses.

In the construction of the 2009 LCMI, to overcome the issue of missing values, it was assumed that each hospital that is in the composition of the hospital centre is representative of it. Given the specificity of certain units, such as oncology centers, they were removed for the sample. Once the index obtained in equation (4) is calculated, the risk-adjusted inpatient admissions can be calculated by its product with the effective production.

Table 1 - Data descriptive statistics

| | Mean | Std. Dev. | Min | Max |
|---------------|-------------|------------------|------------|------------|
| Inputs | | | | |
| Doctors | 405,94 | 317,41 | 103 | 1316 |
| Nurses | 778,28 | 521,90 | 191 | 2131 |
| Otherstaff | 1139,47 | 811,70 | 251 | 3548 |
| Beds | 534,36 | 317,94 | 124 | 1456 |
| Total Costs | 132,35 | 98,50 | 28,14 | 417,52 |

| Outputs | | | | |
|---------------------------|-----------|-----------|-------|--------|
| Inpat | 21164,19 | 11331,06 | 5208 | 50128 |
| Outpat | 247313,54 | 178228,78 | 66194 | 787716 |
| Emergency | 148651,57 | 55629,51 | 67161 | 286430 |
| Surgery | | | | |
| Ambulatory | 5772,59 | 4207,09 | 922 | 18966 |
| Non-ambulatory | 8871,03 | 6539,65 | 0 | 27734 |
| Quality Indicators | | | | |
| PQ - Inpat | 80,77 | 3,01 | 72,20 | 87,00 |
| PQ – Outpat | 77,17 | 2,67 | 72,40 | 82,30 |
| PQ – Emergency | 70,86 | 3,57 | 62,10 | 77,20 |
| PQ – Ambulatory | 83,46 | 2,50 | 76,70 | 89,60 |
| LIC | 4005,38 | 2766,85 | 767 | 11799 |

4. Methodology

In this study we follow the majority of the vast academic literature on measures and comparisons of hospital efficiency, and use Data Envelopment Analysis (DEA) to compare quantitative measures of inputs with quantitative measures of outputs (see section 2). DEA is a non-stochastic, nonparametric estimation method that determines a “best-practice” frontier given the available data. Hospital efficiency scores are then computed with respect to this reference.

4.1 Patient Perceived Satisfaction

Here, we focus on radial output-oriented efficiency measures, i.e., given a fixed vector of inputs what is the maximal proportional expansion achievable of the vector of outputs (see Hollingsworth 2008 for a more complete description of distance function based efficiency measures). The lower bound of these efficiency scores is 1.00, which indicates the decision

making unit (DMU) is output-based efficient, and a score greater than the unity points towards efficiency improvement opportunities. Our research interest focus on the emphasis given to the procedures taken by the hospital management towards improving patient satisfaction holding fixed the unit's resources. Therefore, we adopt an output orientation, i.e., what is the maximum expansion that hospitals can achieve holding inputs fixed,

In the description of the methodology, we follow the notation used by Ferrier et. al (2006). The production function is represented by the correspondence of the vectors of outputs $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_m)$ that can be produced by using the vectors of inputs $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_k)$ as follows:

$$P(\mathbf{x}|V) = \left\{ \mathbf{y}: \mathbf{y} \leq \mathbf{z} \cdot \mathbf{M}, \mathbf{z} \cdot \mathbf{K} \leq \mathbf{x}, \mathbf{z} \in \mathfrak{R}, \sum_{i=1}^n z_i = 1 \right\}, \quad (5)$$

where \mathbf{z} is the vector of weights that generate the convex combinations of inputs (where \mathbf{K} is a n -by- k matrix) and outputs (where \mathbf{M} is a n -by- m matrix). Under this formulation, the technology exhibits variable returns to scale (VRS), as imposed by the restriction on the summation of the elements of \mathbf{z} . Given this characterization of the productive technology, the Farrell output based efficiency measure can be computed by solving the following linear programming problem for each DMU, where the scalar $\theta (\geq 1)$ is the technical efficiency score:

$$\begin{aligned} & \max_{\mathbf{z}, \theta} \theta \\ & s. t. \quad \theta \cdot \mathbf{y} \leq \mathbf{z} \cdot \mathbf{M} \\ & \quad \mathbf{z} \cdot \mathbf{K} \leq \mathbf{x} \\ & \quad \mathbf{z} \in \mathfrak{R} \\ & \quad \sum_{i=1}^n z_i = 1 \end{aligned} \quad (6)$$

Given the purpose of our analysis, we define a perceived quality index (PQI) that allows us to adjust the quantitative output measures. The index is defined as follows:

$$PQI_{i,j} = \frac{PQ_{i,j}}{\overline{PQ}_j}, \quad (7)$$

where $PQI_{i,j}$ is the value of the index for DMU i and for output j , $PQ_{i,j}$ is the perceived quality indicator designed by ACSS and ISEGI-UNL for DMU i and for output j and \overline{PQ}_j is the average value of the indicator for output j in our sample. Clearly, a value greater (lower) than one indicates the DMU i is providing a service with perceived quality standards above (below) the average of the DMUs included in the analysis. Once the index is obtained, it is

used to adjust the quantitative output measures. This adjustment is performed simply by multiplying the PQI by the quantitative output measures, such that those DMU that provide a service with quality standards above average are producing more quality adjusted outputs and thus obtain higher efficiency scores.

The objective is to obtain a set of efficiency scores that does not take into account the perceived quality indicators (baseline - BL scores) and another one that does (perceived quality adjusted – PQA scores). Then, we compare the two sets to determine the effect of the perceived quality adjustment. In order to do so, we conduct a Wilcoxon matched-pairs test. This is a nonparametric test that analyzes the relation between two paired groups. It is used to test whether a treatment has an effect in the population, it does so by looking at the median difference given the set of differences shown by these two groups.

4.2 Timeliness of Care

In addition to the perceived quality indicators, quality is also assessed by timeliness of care, as measured by the waiting lists for elective surgery. Like the analysis done in the previous subsection, the answer we try to answer is whether an efficiency analysis provides a complete picture absent of our chosen qualitative indicators.

Methodologically, this variable is treated as an undesirable output. The fundamental idea is that when producing desirable outputs (such as inpatients visits), the unit also produces waiting lists. Shorter waiting lists could be achieved by shifting resources towards the production of more surgeries, but that would imply producing fewer quantities of some desirable outputs. Thus, this implies a trade-off between producing more of the desirable outputs and less of the undesirable output.

To see whether there is a trade-off between efficiency and quality (as measured by the timeliness of care) it is crucial to find a measure of output congestion, i.e., how much efficiency is reduced by ‘producing’ this undesirable output. To do so we follow the methodology employed by Ferrier et. al (2006) and Clement et al. (2008) that has its roots in Färe et al.(1994). This approach starts by taking all outputs as desirable using the assumption

of strong disposability of all outputs, and then this assumption is replaced by the one of weak disposability of the undesirable output⁴.

Under the strong output disposability assumption, the production technology is described in a similar way as it was in expression (5) with the exception of an added output, as shown in expression (8).

$$P(x|V, S) = \left\{ \mathbf{y}: \mathbf{y} \leq \mathbf{z} \cdot \mathbf{M}, \mathbf{z} \cdot \mathbf{K} \leq \mathbf{x}, \mathbf{z} \in \mathfrak{R}, \sum_{i=1}^n z_i = 1 \right\}, \quad (8)$$

Replacing this assumption with the one of weak disposability of the undesirable output leads to the change in the description of the production technology shown in expression (9).

$$P(x|V, S/W) = \{(\mathbf{y}^S, \mathbf{y}^W): \mathbf{y}^S \leq \mathbf{z} \cdot \mathbf{M}^S, \mathbf{y}^W = \mu \cdot \mathbf{z} \cdot \mathbf{M}^W, \mathbf{z} \cdot \mathbf{K} \leq \mathbf{x}, \mathbf{z} \in \mathfrak{R}, \\ 0 \leq \mu \leq 1, \sum_{i=1}^n z_i = 1\} \quad (9)$$

The congestion score is obtained by the ratio of the technical efficiency scores obtained under the strong disposability assumption and the weak one, as shown in equation (10).

$$C_0(x, \mathbf{y}|V) = \frac{\theta^{S^*}}{\theta^{S/W^*}} \geq 1, \quad (10)$$

where θ^{S^*} and θ^{S/W^*} are found by solving the linear and non-linear programming problems outlined in expressions (11) and (12), respectively.

$$\begin{aligned} & \max_{\mathbf{z}, \theta^{S^*}} \theta^{S^*} \\ & s. t. \\ & \theta^{S^*} \cdot \mathbf{y} \leq \mathbf{z} \cdot \mathbf{M} \\ & \mathbf{z} \cdot \mathbf{K} \leq \mathbf{x} \\ & \mathbf{z} \in \mathfrak{R} \\ & \sum_{i=1}^n z_i = 1 \end{aligned} \quad (11)$$

⁴ For a complete explanation on the different disposability assumptions see Färe et al. (1989).

$$\begin{aligned}
& \max_{\mathbf{z}, \mu, \theta^{S/W^*}} \theta^{S/W^*} \\
& s. t. \\
& \theta^{S/W^*} \cdot \mathbf{y}^S \leq \mathbf{z} \cdot \mathbf{M}^S \\
& \theta^{S/W^*} \cdot \mathbf{y}^W = \mu \cdot \mathbf{z} \cdot \mathbf{M}^W \\
& \mathbf{z} \cdot \mathbf{K} \leq \mathbf{x} \\
& 0 \leq \mu \leq 1 \\
& \mathbf{z} \in \mathfrak{R} \\
& \sum_{i=1}^n z_i = 1
\end{aligned} \tag{12}$$

Once this congestion measure is obtained, we can go to the baseline model described without any reference to quality measures and ask whether those units that are deemed efficient present higher congestion scores than the ones lying below the frontier. If this were to happen, then the higher efficiency scores could be explained by differences in the congestion scores, i.e., efficient units are only so at the cost of longer waiting lists. To see whether this is a valid assumption, a Wilcoxon rank sum and a Kruskal-Wallis test are performed.

To complement the congestion analysis, we calculate perceived quality adjusted congestion scores to answer whether the efficient units identified in our baseline scenario operate under significantly different congestion levels. This new measure aims to capture the changes produced by focusing in a quantity/quality combined accountability of output. To do so, we perform again a Wilcoxon rank sum and a Kruskal-Wallis test.

5. Results

5.1 Patient Perceived Satisfaction

Table 2 presents the output oriented efficiency scores⁵ descriptive statistics.

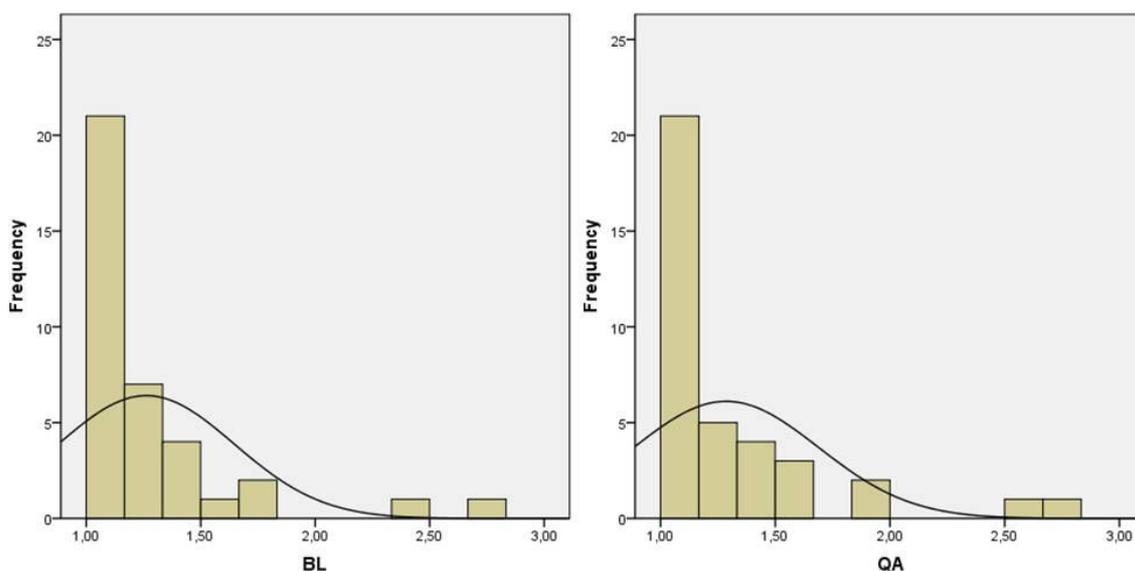
The high percentage of units in the frontier reflects the limited number of observations and the broad definition of the technology. Nevertheless, under this specification, the perceived quality adjustment is significant at a 1% level of significance.

Table 2 - Results descriptive statistics

| Model | Frontier (%) | Mean Effic. Score | Std. Dev. | Min | Max | Skewness | p-value* |
|-------|--------------|-------------------|-----------|-----|----------|----------|----------|
| BL | 48,6 | 1,084355 | 0,112904 | 1 | 1,420455 | 1,420455 | 0,001 |
| PQA | 43,2 | 1,105519 | 0,134322 | 1 | 1,447178 | 1,447178 | |

*asymptotic (1-tailed) p-value obtained for the Wilcoxon matched-pairs test.

Figure 1 - Results distribution



The results shown in the previous table demonstrate that the efficient units remain in the

⁵ Choosing an input orientation would not change significantly the results.

‘best-practice’ frontier regardless the treatment given to perceived quality indicators. This could happen for two reasons: (i) the perceived quality indicators do not provide useful information with respect differentiating the decision making units; or (ii) the most efficient units are also the ones that operate under the highest perceived quality standards. The empirical evidence, as given by the results of the one-tailed Wilcoxon matched-pairs test, seems to favor the second alternative. If alternative (i) were to stand against our estimation results, then the underlying null hypothesis of the one-tailed Wilcoxon matched-pairs test could not be rejected at the usual significance levels, which is not the case. Although we see some significant shifts in the rankings of the several units, the most pronounced effect is the increase in the skewness of the results’ distribution (see figure 1). This indicates that the perceived quality effect leads to a decrease in the efficiency scores of the units located below the frontier. Therefore, we find evidence that not taking quality into account leads, in general, to an overestimation of the efficiency scores of inefficient hospitals. Some exceptions do exist, though.

The results presented above show that the empirical evidence is not sufficient to identify a clear trade-off between efficiency and quality in the hospitals under review, implying the possibility that efficiency gains may achieved without a significant sacrifice of service quality. Nevertheless, there is enough evidence to conclude that analyzing hospital efficiency without consideration of differences in quality of service will generate biased results. When perceived quality is brought to the analysis, the gap between efficient and inefficient units tends to widen, as confirmed by the one-tailed Wilcoxon matched pairs test.

5.2 Timeliness of Care

Table 3- Congestion analysis

| Model | Mean Congestion | Kruskal-Wallis* | Wilcoxon rank sum* |
|--------------|------------------------|------------------------|---------------------------|
| Efficient | 1,06409 | | |
| Inefficient | 1,06674 | 0,60 | 0,60 |

* asymptotic (2-tailed) p-values.

Although the congestion scores are slightly lower for the baseline scenario efficient units, the results in table 3 show that at the usual levels of significance there is not enough evidence to

conclude that there is a difference among the two groups of hospitals. Therefore, the loss of efficiency due to the undesirable output does not explain the difference in efficiency scores, i.e., there does not seem also to exist here a trade-off between efficiency and quality.

The results in table 4 are in line with the ones shown in the previous congestion analysis. Again, the empirical evidence does not support a efficiency-quality trade-off.

Table 4 - Perceived quality adjusted congestion analysis

| Model | Mean Congestion | Kruskal-Wallis* | Wilcoxon rank sum* |
|--------------|------------------------|------------------------|---------------------------|
| Efficient | 1,074557 | | |
| Inefficient | 1,068635 | 0,24 | 0,24 |

* asymptotic (2-tailed) p-values.

6. Robustness

The high number of units deemed efficient, in the specification chosen previously, recommend some caution in their interpretation. The reduced dimension of the sample, unfortunately, does not allow for a higher degree of freedom. Nevertheless, a specification with physical inputs presents a more complete description of the technology and, therefore is the main focus of our analysis.

To check the robustness of our results, we look at total costs as a measure of input. The perceived quality adjustment produces a decrease in five percentage points in the percentage of efficient units, and passes the Wilcoxon matched test at a level of 5% of significance, as we can see in table 5.

Table 5 - Results descriptive statistics

| Model | Frontier (%) | Mean Effic. Score | Std. Dev. | Min | Max | Skewness | p-value* |
|--------------|---------------------|--------------------------|------------------|------------|------------|-----------------|-----------------|
| BL | 32,4 | 1,262159 | 0,383776 | 1 | 2,71739 | 2,321821 | |
| PQA | 27,0 | 1,285839 | 0,402496 | 1 | 2,673797 | 2,127823 | 0,01 |

*asymptotic (1-tailed) p-value obtained for the Wilcoxon matched-pairs test.

Again, we find that the most of the units deemed efficient in a purely quantitative specification remain so when the perceived quality indicators are taken into account. The same happens for the generalized decrease in the inefficient units technical efficiency scores. Therefore, we conclude that our findings are robust to the input specification.

We now turn our attention towards the congestion scores. Table 6 shows, once more, that there are only slight differences in the congestion scores of efficient and inefficient units. This can be confirmed by the non-parametric test results. As such, we conclude that the absence of a trade-off between efficiency and perceived quality is robust to the technology specification.

Table 6- Congestion analysis

| Model | Mean Congestion | Kruskal-Wallis* | Wilcoxon rank sum* |
|--------------|------------------------|------------------------|---------------------------|
| Efficient | 1,054741 | | |
| Inefficient | 1,039262 | 0,4 | 0,4 |

* asymptotic (2-tailed) p-values.

The same is true for the perceived quality adjusted congestion analysis, as shown in table 7.

Table 7 - Perceived quality adjusted congestion analysis

| Model | Mean Congestion | Kruskal-Wallis* | Wilcoxon rank sum* |
|--------------|------------------------|------------------------|---------------------------|
| Efficient | 1,062614 | | |
| Inefficient | 1,045533 | 0,26 | 0,26 |

* asymptotic (2-tailed) p-values.

7. Conclusions

The purpose of this paper is to answer two questions. The first one is whether a trade-off between efficiency and quality exist in the Portuguese NHS hospitals. The second one is whether DEA efficiency scores are biased when strictly quantitative outputs are considered.

To answer the first question, we use two types of indicators. The first set of indicators is

based on data from a 2009 survey of patients, whose main goal is to provide an independent system of regular evaluation of patient satisfaction and of hospital quality, as perceived by users of Portuguese NHS hospitals. The second set of indicators is related to the timeliness of care, as measured by the waiting lists for elective surgery. Our analysis suggests that a trade-off between efficiency and quality does not seem to exist here. Therefore, strictly quantitative output specifications tend to provide a complete picture for those units deemed efficient.

To answer the second one, we focus on the perceived patient satisfaction scores. We find that for those units deemed inefficient, abstracting from quality adjustments may lead to the overestimation of their technical efficiency scores.

We conclude that the empirical evidence is not sufficient to identify a clear trade-off between efficiency and quality in the hospitals under review, implying the possibility that efficiency gains may be achieved without a significant sacrifice of service quality. Nevertheless, there is enough evidence to conclude that analyzing hospital efficiency without consideration of differences in quality of service will generate biased results.

When DEA efficiency scores are adjusted for output quality, the decision making units that lie on the technical efficiency frontier remain largely unaltered, even if a great weight is given to quality indicators over quantity indicators of output.

Nevertheless, we find that outside of the frontier adjusting for quality does have an impact in efficiency scores. Our analysis of quality adjusted efficiency scores reveals that, using a Wilcoxon matched pairs test, the median efficiency scores are statistically different at the usual levels of significance.

We conclude that the empirical evidence is not sufficient to identify a clear trade-off between efficiency and quality in the hospitals under review, implying the possibility that efficiency gains may be achieved without a significant sacrifice of service quality. Furthermore, there is enough evidence to conclude that analyzing hospital efficiency without consideration of differences in quality of service will generate biased results.

References

- Afonso, A. and St Aubyn M. (2005), Non-parametric approaches to education and health efficiency in OECD countries, *The Journal of Applied Economics* 8(2): 227–246.
- Barros, C.P., Menezes, A.G., Vieira, J.C., Peypoch, N., Solonandrasana, B. (2008), “An Analysis of Hospital Efficiency and Productivity Growth Using the Luenberger Indicator”, *Health Care Management Science*, 11(4), 373-381.
- Bhat VN. (2005), Institutional arrangements and efficiency of health care delivery systems, *The European Journal of Health Economics* 3: 215–222.
- Clement, J.P., Valdmanis, V.G., Bazzoli, G.J., Zhao, M., Chukmaitov, A., (2008), Is more better? An analysis of hospital outcomes and efficiency with a DEA model of output congestion”, *Health Care Management Science* 11, 67-77.
- Coelli, T. (2005). *An introduction to efficiency and productivity analysis*, Springer Verlag.
- Färe R., Grosskopf S., Lovell CAK. (1994) *Production frontiers*. Cambridge University Press.
- Färe, R., Grosskopf, S., Lovell, C.A.K., Pasurka, C., (1989), Multilateral productivity comparisons when some outputs are undesirable, *The Review of Economics and Statistics* 71, 90–98.
- Ferrier G., Rosko M., Valdmanis V., (2006), Analysis of uncompensated hospital care using a DEA model of output congestion, *Health Care Management Science* 9(2):187–188.
- Fried, H. O., C. A. K. Lovell, et al. (2008), *The measurement of productive efficiency and productivity growth*, Oxford University Press, USA.
- Gouveia M., Alvim J., Carvalho C., Correia J. and Pinto M. (2006), *Resultados da Avaliação dos Hospitais SA, Comissão para a Avaliação dos Hospitais Sociedade Anónima, Technical Report*.
- Harfouche, A. P. (2008), *Hospitais Transformados em Empresas Análise do Impacto na Eficiência: Estudo Comparativo*, Instituto Superior de Ciências Sociais e Políticas, Lisboa.
- Herr, A. (2008), Cost and technical efficiency of German hospitals: does ownership matter?, *Health Economics*, 17 (9): 1057-1071.
- Hollingsworth, B. (2008), The Measurement Of Efficiency And Productivity Of Health Care Delivery, *Health Economics* 17: 1107–1128.
- Navarro-Espigares, J. L. and Hernández, E. (2011), Efficiency and quality in health services: a crucial link, *The Service Industries Journal*, 31:3, 385-403.
- Nayar, P. and Ozcan, Y. (2008), Data Envelopment Analysis Comparison of Hospital Efficiency and Quality., *J. Med. Syst.* 32(3): 193-199.
- Rego, G., Nunes, R. and Costa J. (2010), The challenge of corporatisation: the experience of Portuguese public hospitals, *European Journal of Health Economics* (2010) 11:367–381.

Retzlaff-Roberts D, Chang CF, Rubin RM. (2004), Technical efficiency in the use of health care resources: a comparison of OECD countries, *Health Policy* 69(1): 55–72.

Simões, P. and Marques R., (2011), Performance and congestion analysis of the Portuguese hospital services, *Central European journal of operations research* 19:39–63.

Valdmanis, V., Kumanarayake, L., Lertiendumrong, J. (2004), Capacity of Thai public hospitals and the production of care for poor and non-poor patients, *Health Services Research* 39: 2117-2134.

Recent FEP Working Papers

| | |
|--------|---|
| Nº 434 | Octávio Figueiredo, Paulo Guimarães and Douglas Woodward, " <u>Firm-Worker Matching in Industrial Clusters</u> ", October 2011 |
| Nº 433 | Susana Assunção, Rosa Forte and Aurora A.C. Teixeira, " <u>Location Determinants of FDI: a Literature Review</u> ", October 2011 |
| Nº 432 | António Brandão, Luís Guimarães and Carlos Seixas, " <u>The Relationship between Trigger Price and Punishment Period in Green and Porter (1984) Game made Endogenous</u> ", October 2011 |
| Nº 431 | Argentino Pessoa, " <u>The Cluster Policy Paradox: Externalities vs. Comparative Advantages</u> ", October 2011 |
| Nº 430 | Susana Assunção, Aurora A.C. Teixeira and Rosa Forte, " <u>Do Countries' Endowments of Non-renewable Energy Resources Matter For FDI Attraction? A Cross-country Econometric Analysis</u> ", October 2011 |
| Nº 429 | Óscar Afonso and Armando Silva, " <u>Non-scale endogenous growth effects of subsidies for exporters</u> ", September 2011 |
| Nº 428 | Mariana Dias and Aurora A.C. Teixeira, " <u>Geopolítica e International Business: uma tentativa de síntese e proposta de enquadramento teórico para aplicação prática</u> ", September 2011 |
| Nº 427 | Carina Silva and Aurora A.C. Teixeira, " <u>Empreendedorismo político local em Portugal. Uma análise exploratória</u> ", September 2011 |
| Nº 426 | Marta Couto and Aurora A.C. Teixeira, " <u>Festivais de Música de Verão em Portugal: determinantes da participação e a identificação dos seus patrocinadores</u> ", September 2011 |
| Nº 425 | Luis Carvalho and Aurora A.C. Teixeira, " <u>Where are the poor in International Economics?</u> ", September 2011 |
| Nº 424 | Maria Inês Veloso Ferreira and Aurora A.C. Teixeira, " <u>Organizational Characteristics and Performance of Export Promotion Agencies: Portugal and Ireland compared</u> ", September 2011 |
| Nº 423 | Pedro Cosme Costa Vieira, " <u>Está na hora de Portugal sair da Zona Euro</u> ", September 2011 |
| Nº 422 | Márcia Daniela Barbosa Oliveira and João Gama, " <u>How we got Here? A Methodology to Study the Evolution of Economies</u> ", July 2011 |
| Nº 421 | Vitor M. Carvalho and Manuel M. F. Martins, " <u>Macroeconomic effects of fiscal consolidations in a DSGE model for the Euro Area: does composition matter?</u> ", July 2011 |
| Nº 420 | Duarte Leite, Pedro Campos and Isabel Mota, " <u>Computational Results on Membership in R&D Cooperation Networks: To Be or Not To Be in a Research Joint Venture</u> ", July 2011 |
| Nº 419 | Sandra T. Silva, Isabel Mota and Filipe Grilo, " <u>The Use of Game Theory in Regional Economics: a quantitative retrospective</u> ", June 2011 |
| Nº 418 | Marisa R. Ferreira, Teresa Proença and João F. Proença, " <u>An Empirical Analysis about Motivations among Hospital Volunteers</u> ", June 2011 |
| Nº 417 | Marlene Grande and Aurora A.C. Teixeira, " <u>Corruption and Multinational Companies' Entry Modes. Do Linguistic and Historical Ties Matter?</u> ", June 2011 |
| Nº 416 | Aurora A.C. Teixeira, " <u>Mapping the (In)visible College(s) in the Field of Entrepreneurship</u> ", June 2011 |
| Nº 415 | Liliana Fernandes, Américo Mendes and Aurora A.C. Teixeira, " <u>A weighted multidimensional index of child well-being which incorporates children's individual perceptions</u> ", June 2011 |
| Nº 414 | Gonçalo Faria and João Correia-da-Silva, " <u>A Closed-Form Solution for Options with Ambiguity about Stochastic Volatility</u> ", May 2011 |
| Nº 413 | Abel L. Costa Fernandes and Paulo R. Mota, " <u>The Roots of the Eurozone Sovereign Debt Crisis: PIGS vs Non-PIGS</u> ", May 2011 |
| Nº 412 | Goretti Nunes, Isabel Mota and Pedro Campos, " <u>Policentrismo Funcional em Portugal: Uma avaliação</u> ", May 2011 |
| Nº 411 | Ricardo Biscaia and Isabel Mota, " <u>Models of Spatial Competition: a Critical Review</u> ", May |

| | |
|--------|--|
| | 2011 |
| Nº 410 | Paula Sarmento, " The Effects of Vertical Separation and Access Price Regulation on Investment Incentives ", April 2011 |
| Nº 409 | Ester Gomes da Silva, " Portugal and Spain: catching up and falling behind. A comparative analysis of productivity trends and their causes, 1980-2007 ", April 2011 |
| Nº 408 | José Pedro Figue, " Endogenous Response to the 'Network Tax' ", March 2011 |
| Nº 407 | Susana Silva, Isabel Soares and Carlos Pinho, " The impact of renewable energy sources on economic growth and CO2 emissions - a SVAR approach ", March 2011 |
| Nº 406 | Elena Sochirca and Sandra Tavares Silva, " Efficient redistribution policy: an analysis focused on the quality of institutions and public education ", March 2011 |
| Nº 405 | Pedro Campos, Pavel Brazdil and Isabel Mota, " Comparing Strategies of Collaborative Networks for R&D: an agent-based study ", March 2011 |
| Nº 404 | Adelaide Figueiredo, Fernanda Figueiredo, Natália P. Monteiro and Odd Rune Straume, " Restructuring in privatised firms: a Stata approach ", February 2011 |
| Nº 403 | Cláudia M. F. Pereira Lopes, António Cerqueira and Elísio Brandão, " The financial reporting quality effect on European firm performance ", February 2011 |
| Nº 402 | Armando Silva, " Financial constraints and exports: evidence from Portuguese manufacturing firms ", February 2011 |
| Nº 401 | Elena Sochirca, Óscar Afonso and Pedro Mazedo Gil, " Directed technological change with costly investment and complementarities, and the skill premium ", January 2011 |
| Nº 400 | Joana Afonso, Isabel Mota and Sandra Tavares Silva, " Micro credit and Territory - Portugal as a case study ", January 2011 |
| Nº 399 | Gonçalo Faria and João Correia-da-Silva, " The Price of Risk and Ambiguity in an Intertemporal General Equilibrium Model of Asset Prices ", January 2011 |
| Nº 398 | Mário Alexandre Patrício Martins da Silva, " A Model of Innovation and Learning with Involuntary Spillovers and absorptive capacity ", January 2011 |
| Nº 397 | Fernando Governo and Aurora A.C. Teixeira, " Marketing and technology sophistication as hidden weapons for fostering the demand for 'art house' cinema films: a cross country analysis ", January 2011 |
| Nº 396 | Liliana Fernandes, Américo Mendes and Aurora A.C. Teixeira, " A review essay on child well-being measurement: uncovering the paths for future research ", December 2010 |
| Nº 395 | David Nascimento and Aurora A.C. Teixeira, " Recent trends in the economics of innovation literature through the lens of Industrial and Corporate Change ", December 2010 |
| Nº 394 | António Brandão, João Correia-da-Silva and Joana Pinho, " Spatial competition between shopping centers ", December 2010 |
| Nº 393 | Susana Silva, Isabel Soares and Óscar Afonso, " E3 Models Revisited ", December 2010 |
| Nº 392 | Catarina Roseira, Carlos Brito and Stephan C. Henneberg, " Innovation-based Nets as Collective Actors: A Heterarchization Case Study from the Automotive Industry ", November 2010 |
| Nº 391 | Li Shu and Aurora A.C. Teixeira, " The level of human capital in innovative firms located in China. Is foreign capital relevant? ", November 2010 |
| Nº 390 | Rui Moura and Rosa Forte, " The Effects of Foreign Direct Investment on the Host Country Economic Growth - Theory and Empirical Evidence ", November 2010 |
| Nº 389 | Pedro Mazedo Gil and Fernanda Figueiredo, " Firm Size Distribution under Horizontal and Vertical R&D ", October 2010 |
| Nº 388 | Wei Heyuan and Aurora A.C. Teixeira, " Is human capital relevant in attracting innovative FDI to China? ", October 2010 |
| Nº 387 | Carlos F. Alves and Cristina Barbot, " Does market concentration of downstream buyers squeeze upstream suppliers' market power? ", September 2010 |

Editor: Sandra Silva (sandras@fep.up.pt)

Download available at:

<http://www.fep.up.pt/investigacao/workingpapers/>

also in <http://ideas.repec.org/PaperSeries.html>

www.fep.up.pt

FACULDADE DE ECONOMIA DA UNIVERSIDADE DO PORTO

Rua Dr. Roberto Frias, 4200-464 Porto | Tel. 225 571 100

Tel. 225571100 | www.fep.up.pt