

This is a pre-publication version of this paper. Some of the details may differ in the published version. Please do not quote without prior written permission from the Corresponding Author.

Identifying the Scope for Savings At Inpatient Episode Level: An Illustration

Applying DEA to Chronic Obstructive Pulmonary Disease

Emmanuel Thanassoulis, Aston University, Birmingham, UK

(e.thanassoulis@aston.ac.uk)

Maria Portela¹, CEGE, Católica Porto Business School (csilva@porto.ucp.pt)

Mike Graveney, University of Warwick, UK (mikegraveney09@gmail.com)

Abstract

Chronic obstructive pulmonary disease (COPD) is characterized by a largely irreversible obstruction of the airways, and is one of the leading causes of chronic morbidity and mortality worldwide. This paper illustrates the use of Data Envelopment Analysis (DEA) to assess the potential for cost savings at COPD inpatient episode level. The analysis uses the length of stay of each episode as a surrogate for expenditure on that episode while allowing for the medical condition of the patient and the quality of care received. We find substantial possible reductions in length of stay which would translate to cost savings. The paper also explores differences both between hospitals and between care teams within hospitals so that cost efficient protocols of treatment can be identified and disseminated.

Keywords: Health care, COPD, efficiency analysis

1. Introduction

Chronic obstructive pulmonary disease (COPD) is characterized by a largely irreversible obstruction of the airways, and encompasses both emphysema and chronic bronchitis (Afonso et al, 2011). According to the World Health Organization (WHO) COPD is one of the leading causes of chronic morbidity and mortality worldwide. This paper investigates the comparative *efficiency* of certain UK hospitals in handling this disease using data at inpatient episode level. It sets the length of stay (LOS) of an episode against the medical condition of the patient, including co-morbidities with COPD, age and gender with the view to identifying the potential for reduction of LOS at no detriment to the quality of treatment the patient has received. While performance is assessed at inpatient episode level, the results are then aggregated at higher levels, including hospital level, with a view to

¹ Corresponding author. Católica Porto Business School. Rua Diogo Botelho 1327, 4169-05 Porto, Portugal

identifying systematic best practice at hospital or constituent unit level. We assess the scope for savings using Data Envelopment Analysis (DEA). This is because of its flexibility in estimating a piece-wise-linear cost frontier which envelops the observed data rather than requiring the specification of an assumed functional form for that frontier. We thus avoid the risk of mis-specifying a frontier functional form and of making assumptions about the probability distribution of any inefficiencies. Further, DEA has the advantage of identifying benchmark episodes of treatment which are most suitable for each episode which has scope for savings. Using episode-specific benchmarks enables non-technical hospital managers to gain confidence in the findings when comparing instances of treatment within their hospital on costs and quality of service. Episode – specific benchmarks also are likely to be the most suitable examples from which best practice for adoption can be drawn.

In the UK there has been a growing interest for analyses at patient level since the introduction of patient-level costing systems in the National Health System (NHS) in the mid 2000's. However, "in 2011 less than half of NHS trusts had patient-level information and costing systems (PLICS)" (Blunt and Bardsley, 2012). NHS hospitals are paid on the basis of a fixed price per diagnosis-related admission, designated 'tariff' under the Payment by Results (PbR) framework. The tariff is computed from national reference costs for each Healthcare Resource Group (HRG). Each medical diagnosis maps to an HRG. The tariff can be complemented in cases where the patient requires specialist care or longer length of stay. As will be seen in section 3, the cost data returned by hospitals for treating each inpatient episode revealed serious problems of reliability. This made it impossible to use cost data in our analysis. However, we could use Length of Stay (LOS) data, which is far more reliable. LOS can be seen as a surrogate for the expenditure by the hospital in treating a patient, given that we have controlled for primary diagnosis and use surrogates for any co-morbidities the patient has.

To the best of our knowledge this is the first paper to analyse the potential savings in the treatment of COPD patients. The DEA models used for this purpose, though not totally new, possess some methodological novelties necessitated by the fact that the input variable used was truncated (as only episodes with LOS of 2 or more days were considered). Hospitals are compared on their ability to reduce costs through reductions in LOS.

Consultants within each hospital are also compared on their ability to reduce costs given the characteristics of the episodes each one handles. This study contributes to the scarce literature on applying non-parametric frontier models to patient level data.

We find substantial possible reductions in length of stay of COPD patients, which would translate in cost savings. This is in line with previous literature suggesting that hospital capacity can be increased through reduction of LOS without compromising quality (see Berk and Moinzadeh (1998)). We find, however, in some cases that there is a trade-off between LOS and time to re-admission, with too short lengths of stay generating sooner readmissions for patients with a high disease weight. We also find that hospitals and care teams within hospitals can have differing performance on length of stay, sometimes by patient gender, not entirely justified by the disease weight of the patients treated. We suggest such differences be investigated with a view to identifying treatment protocols, which are economical without detriment to quality of care.

2. Motivation for Analysing Health Care Costs at Patient Level

The assessment of cost efficiency or productivity of health service entities at the aggregate level has been criticized by many authors (e.g. Hollingsworth, 2008 and Bradford et al., 2001). This is primarily because at the aggregate level one cannot reflect accurately the multiplicity of differences between treated patients. The best predictor of a patient's treatment costs is the medical diagnosis for which the particular instance of treatment is intended and the accompanying general health condition, including age and gender of the patient. This has long been recognised, and most cost analyses in the health care field use one or other 'case-mix' measures as an explanatory factor for treatment costs when analyses are at aggregate level. However, such a 'case mix' is still too broad a category to reflect accurately the treatment needs of a cohort of patients.

In this paper we narrow down case-mix to cases where the 'primary' diagnosis relates to just 4 Health Resource Groups (HRGs), all related to the COPD condition. This is the first step in controlling for the disease weight each patient presents, which drives their treatment cost. However, it is well known that even within each HRG a large variability of costs exist

due to various factors. This can be because both the HRG does not fully capture the medical condition of a patient, and therefore his or her clinical needs, and also because of the efficacy of the operating practices both clinical and managerial of a hospital, which can impact costs. The aim in this paper is to isolate the latter costs by controlling for the clinical needs of the patient. One way to control for the clinical needs of the patient is to use patient and indeed episode level data. These reflect more accurately the disease weight of the patient at the point of admission. Moving from aggregate data (e.g. hospital or General Practice level) to patient episode level data, makes it possible to derive findings specific to a particular health condition (diagnosis).

Most studies in the health literature using data at patient/episode level have relied on multilevel regression models, since patients are nested into services and services are nested into hospitals (see e.g. the literature review of Diez-Roux, 2000). These multilevel studies have been mainly concerned with identifying the cost drivers of patients with a certain disease, and assess the impact of each cost driver and the variability of costs across hospitals. For example, Dormont and Milcent (2004), Olsen and Street (2008), Kristensen et al. (2010), Laudicella et al. (2010), and Daidone and Street (2011) used multilevel models to analyse the costs of treating patients with various diseases. Dormont and Milcent (2004) analysed cost variability of French hospitals based on 7314 stays for acute myocardial infarction during 1994-1997; Olsen and Street (2008) were interested in patients attending vascular departments in 8 Danish hospitals in 2004; Kristensen et al. (2010) analysed all diabetic patients in the UK in 2005/2006, in a total of 31371 patients, and Laudicella et al. (2010) analysed patients from 136 obstetrician departments in the UK in 2005/2006. Daidone and Street (2011), on the other hand, analysed UK patient data from 2008/2009 and assessed 12m patients not restricting to a single disease or specialty, but considering all patients receiving diverse types of treatment. As it would be impossible to include in the regression models dummy variables for about 1400 HRGs, Daidone and Street (2011) divided each patient cost by the mean cost of all patients within the same HRG. They used this standardised cost variable as the dependent variable in a regression model. The use of this variable implied that the authors were investigating the drivers of cost deviations from average and the effect of the hospital on these deviations.

All the above multilevel studies attempted to explain costs of treatment, where a number of control variables (relating mostly to patients) have been taken into account. Examples of such types of variable are the age, gender, race, socio-economic status, number of operations, number of diagnoses, existence of other conditions or co-morbidities in addition to the disease being analysed, emergency admissions, transfers, death, the HRG group and, diagnostic markers based on ICD-10 (the international classification of disease, 10th revision). In multilevel models, it is possible to isolate the proportion of total variation in cost that is due to patient characteristics and that due to hospital/department characteristics. The latter is usually called the hospital/department effect. This effect can also be explained by additional variables (this time measured at the hospital or department level). DEA models too can consider the multilevel nature of data, but through a different approach where all units are analysed together (with reference to a pooled frontier) and also within a certain hierarchical level (e.g. the hospital).

DEA has been used before for patient/episode level analyses. For example, Puig-Junoy (1998) used DEA to assess 16 Intensive Care Units in Catalonia, Spain using data on 993 patients, and Dervaux et al. (2009) used a directional distance function in a Free Disposal Hull (FDH) technology using data on 15029 patients treated in 25 Parisian ICUs during 2000. Bradford et al. (2001) used Stochastic Frontier Analysis (SFA) to compare the costs of two technologies for treating coronary artery disease in 645 patients observed in the year of 1994, while Dismuke and Sena (2001) used a Malmquist-Luenberger framework to analyse quality-adjusted measures of productivity change for three diagnostic technologies commonly used in the main Portuguese hospitals over the period 1992–1994.

The Bradford et al. (2001) model was very similar to the multilevel models described above in that the costs of treating patients for cardiac revascularization were explained through patient characteristics (such as co-morbidities, age, gender, etc.), and a dummy for the doctors treating the patient.

Regarding the use of DEA models, Puig-Junoy (1998) used patient level data on 2 outputs (number of days surviving in the hospital, and the discharge status - death or surviving at

hospital discharge) and some inputs reflecting patient characteristics (e.g. survival probability at admission, mortality risk level, weighted ICU days, non-ICU hospital days), and clinical practice (e.g. available nurse days per patient, available physician days per patient, and technological availability.) Puig-Junoy (1998) also used a second stage regression analysis to explain the efficiency yielded by the DEA analysis, where explanatory variables included the type of hospital (profit vs non-profit), number of beds in the ICU, proportion of patients in the same risk group, age, teaching vs non-teaching hospital, etc). This follows the same spirit of including second level variables in multilevel models that are used to explain the hospital effect. In their analysis of Parisian ICUs, Dervaux et al. (2009) followed a similar approach, but used a single output related to “the index of medical outcome” calculated as the probability of death minus the discharge status (a binary variable, where 0 represented alive and 1 deceased). The probability of death was computed through a simplified acute physiology score (SAPSII) as shown in <http://www.sfar.org/scores2/saps2.html>. Inputs related to the resources used, and they were distinguished in (i) those used only once (like insertion of a catheter or a tracheotomy) (ii) repeated treatments including angiography echocardiograms, X-rays, and (iii) treatment that is provided daily such as monitoring, dialysis, and feeding. Length of stay (LOS) was included as an input to represent the hotel type services of the hospital. Control variables were also introduced to reflect medical conditions of patients (10 diagnostic categories, mode of entry, medical or surgical treatment, and SAPSII). These variables were treated as non-discretionary variables where patients only with the same (or worst) conditions on entry were compared with each other.

The above studies share some similarities with ours in the sense that they used DEA models with data at the patient level. They, however, do not specify hospital specific frontiers as we do, nor do they try to understand whether there are substantial differences in treatment amongst hospitals and consultants. In addition, they were concerned with analyzing a variety of diseases treated in ICUs, meaning that several control variables for the actual disease of the patient needed to be accounted for. In this study we focus instead on a single medical condition - COPD (chronic obstructive pulmonary disease). In spite of there being a single condition, patients can still differ in terms of the disease weight, which we also

account for through two variables constructed from the database available to us, (note that for COPD standard measures of risk, like APACHE II, SAPS II or SOFA do not exist). In our analysis of COPD we are interested in analysing costs and isolate those that might reflect inefficiency in the management of the patient episode as distinct from costs necessitated by the medical condition of the patient. The next section outlines our approach to reflecting costs and to control for disease weight.

3. Defining the Inpatient Cohort

We are using in this paper inpatients admitted with a primary diagnosis of “Chronic obstructive pulmonary disease” (COPD). Specifically that is where the full patient coding equated to one or other of the three HRGs, which in their definition include COPD. These diagnoses correspond to ICDs J43 and J44 (3 digit code for the ICD). All episodes considered had been included in HRGs D39, D40 and D99. The percentage of episodes in each of these HRGs and their descriptions are shown in Table 1.

Table 1: Distribution of episodes per HRG

<i>HRG</i>	<i>Description</i>	<i>N</i>	<i>%</i>
D39	COPD or Bronchitis w cc*	286	31.78
D40	COPD or Bronchitis w/o cc**	496	55.11
D99	Complex Elderly with a Respiratory System Primary Diagnosis	118	13.11
Total		900	100.0

* w cc With complications

** w/o cc Without complications

All episodes considered in our analysis showed in the first diagnostic position either the 3 digit ICD J43 or J44. In a four digit coding these correspond to the diagnoses shown in Table 2.

Table 2: Distribution of episodes per first diagnostic position ICD

<i>ICD</i>	<i>Description</i>	<i>N</i>	<i>%</i>
J439	Emphysema, unspecified	29	3.22
J440	Chronic obstructive pulmonary disease with acute lower respiratory infection	326	36.22
J441	Chronic obstructive pulmonary disease with acute exacerbation, unspecified	486	54.0
J448	Other specified chronic obstructive pulmonary disease	8	0.89
J449	Chronic obstructive pulmonary disease, unspecified	51	5.67

The sample used in this study comprises 900 inpatient episodes generated by 500 patients over 5 years (year 1 - April/2005 – March/2006 to year 5 - April/2009 – March/2010). The episodes used were extracted from a larger data set on COPD-related admissions, using the criteria of age (the patients analysed all are between 65 and 75 years of age) and length of stay (all episodes had from 2 to 28 days of stay in hospital). The age range of 65-75 was used to reflect the fact that LOS can depend on age, *ceteris paribus*, and so in comparing episodes on LOS we need to control for patient age. The range 65-75 years of age is deemed similar in terms of impact on LOS, *ceteris paribus*. The lower bound of 2 days of LOS for episodes included in the analysis was used because it was felt admissions of 1 day would differ from the typical in-patient episode in terms of severity and treatment protocol and would be closer to outpatient admission. The upper bound of 28 days affected very few patients and was used to exclude from the analysis the few cases of extreme medical dependence likely to be not comparable with the bulk of the cases analysed. We only considered emergency admissions as these represent the bulk of patients with COPD. Finally we excluded episodes where the discharge status was death, as in a sense the data would reflect right-censored LOS.

The cost that hospitals report relative to an episode is mainly a function of the tariff charged for the respective HRG. As a result, the reported cost is very much unrelated with the length of stay of a patient, contrary to what we would expect. This can be readily seen in Figure 1, where reported episode cost figures (in £) are plotted against length of stay by HRG.

It is clear that up to a certain threshold of length of stay, which differs by HRG, the reported cost is constant. For example, an episode on HRG D99 with 2 days of stay is charged the same as an episode within the same HRG with 25 days of stay! This costing clearly does not reflect the true costs the hospital incurred for that episode. Costs vary per HRG, even though the three HRGs are sometimes associated with the same patient and with similar diagnoses and at other times they can be the same despite different LOS and mix of diagnoses. To see this consider the example below of a certain patient aged 67 in year 2, who visited the hospital several times with the data shown in Table3 (note: Dia_k refers to the diagnosis $k = 1, \dots, 8$ that can be attributed to the episode).

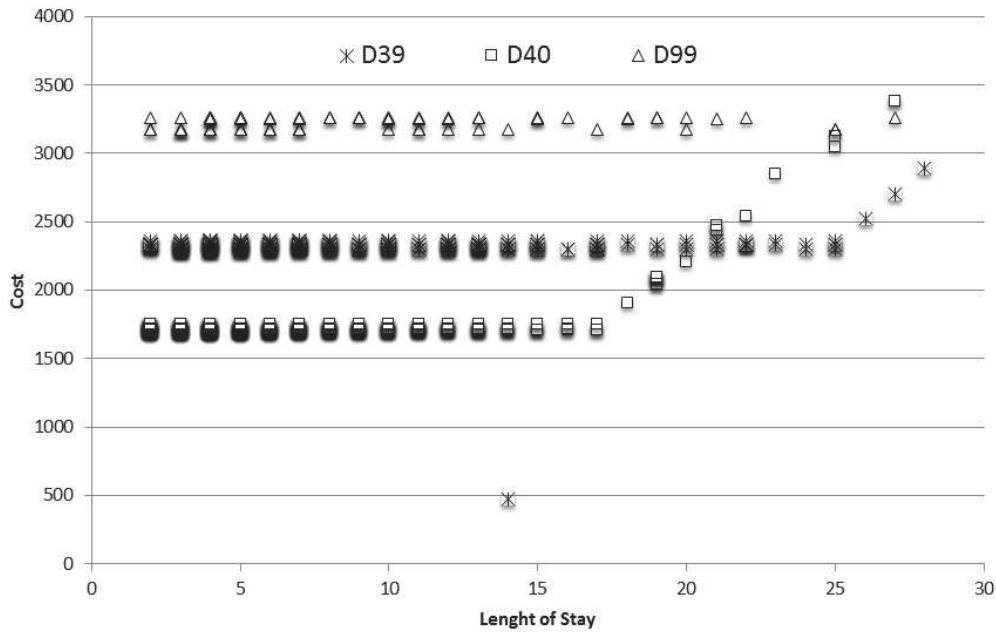


Figure 1: Cost versus length of stay

Table 3: Example of several admissions of a certain patient

Year	Los	Dia_1	Dia_2	Dia_3	Dia_4	Dia_5	Dia_6	Dia_7	HRG	cost
2	7	J440	I120	I48X	H109	I451			D40	1709
3	4	J441	I10X	I48X	E119	I500	R600		D39	2360
4	2	J441	E109	I500	I10X	E780	N189	D649	D39	2337
5	2	J440	C342	I48X	I500	E119			D99	3257

Admission 1 (in year 2) is coded under HRG D40, admission 2 (happening 315 days later in year 3) is coded under HRG D39, and its cost is higher even though the LOS is just 4 days compared to 7 days of the year 2 admission. One could assume that the two admissions have different mix of diagnoses (apart from one), which may explain the difference in costs. At the opposite end of the spectrum we have similar costs for inpatient episodes where we would expect them to be different. For example, the admissions in years 3 and 4 have the same primary diagnosis and HRG, but the mix of diagnoses is different and length of stay in year 4 is half that in year 3. Yet reported costs are virtually the same for both inpatient episodes. Clearly differences in the HRGs assigned to episodes will reflect the mix of diagnoses of each episode. For example the fact that the HRG D99 was assigned to this patient on the last admission may be related with the malignant neoplasm (C342) diagnosed in this admission and not in the previous ones. The description of each diagnosis

in Table 3 is shown in Table 4. Tables 3 and 4 demonstrate the complexity some patients can present on admission and the varying nature of that complexity over time. This in turn has cost implications for inpatient episodes. Yet, those costs, as reported by hospitals during the time period our data covers, do not on the face of it seem to capture the true level of expenditure by the hospital in treating the patient concerned. In view of this we have adopted the view, often found in the literature, that the variable cost in treating an inpatient can be reflected by the length of stay of that patient. It should be noted in this respect that our aim is simply to compare episodes on costs rather than derive their treatment costs in absolute terms.

Table 4: Description of the diagnoses of the patient in Table 3

code	description
C342	Malignant neoplasm of middle lobe, bronchus or lung
D649	Anaemia, unspecified
E109	Insulin-dependent diabetes mellitus without complications
E119	Non-insulin-depend diabetes mellitus without complication
E780	Pure hypercholesterolaemia
H109	Conjunctivitis, unspecified
I10X	Essential (primary) hypertension
I120	Hypertensive renal disease with renal failure
I451	Other and unspecified right bundle-branch block
I48X	Atrial fibrillation and flutter
I500	Congestive heart failure
J440	Chronic obstruct pulmonary dis with acute lower resp infec
J441	Chron obstruct pulmonary dis wth acute exacerbation, unspec
N189	Chronic renal failure, unspecified
R600	Localized oedema

4. Setting up a DEA Model to Compare COPD Inpatient Episodes on Costs

Following reflection on how to capture the disease weight a patient presents on admission to a COPD inpatient spell, and within the confines of the data available to us, we arrived at the input-output variables in Table 5.

Table 5: Input and output set

<i>Input</i>	<i>Outputs</i>
LOS of the episode being assessed	<ul style="list-style-type: none"> - Number of different diagnoses the patient had in the 365 days prior to the episode being assessed, including the current diagnoses; - Sum of LOS of all admissions the patient had (in the previous 365 days) - Time to readmission of the patient from discharge of the episode being assessed

As noted earlier the aim of our analysis is to isolate expenditure on treating COPD inpatient cases, which could represent waste. For this purpose on the input side ideally we need the treatment cost by COPD inpatient episode and on the output side variables that would reflect the clinical need of the patient, which drives the uncontrollable part of the cost of an inpatient spell. Our input-output variables in Table 5 are the closest surrogate variables for cost on the one hand and clinical need on the other that we had data for, at patient level. Beyond the variables in Table 5 (on the output side) we also control for **age** and **gender** of the patient as the data was analysed by gender, within a given age range. Note that several authors have reported the importance of age, gender and social status as primary determinants of the health condition of patients. For example Sørensen et al. (2009) report impacts on general practices' expenditures, which are different by gender, by age band and by professional status, while age and gender were also amongst the explanatory variables of treatment costs used in Olsen and Street (2008).

The rationale behind the choice of each variable in Table 5 is as follows.

Length of stay serves as a proxy for the costs incurred by the hospital in the course of a COPD inpatient episode. As noted earlier, the costs returned by hospitals cannot be used directly because for LOS up to a certain duration there is no variability in costs returned. Moreover, even when the mix of diagnoses, and therefore the clinical need of a patient changes, so long as the HRG stays the same so do by and large the treatment costs returned by the hospital.

In this choice we adopt the same view as others who have also relied on LOS to reflect expenditure. For example, Street et al. (2012) analysed a dataset of patients with one of 10 previously defined health conditions (associated with specific ICDs) who were admitted to

hospitals in 10 European countries. In the absence of data on cost for some countries LOS was used instead as a proxy. In this respect the authors note that “there are advantages to analyzing LOS, notably because information is more readily available and less subject to discretionary measurement than costs. Analysis based on LOS rather than cost may also prove more powerful at fostering behaviour change if it prompts clinicians to ask why their patients are staying longer in hospitals than are those treated elsewhere” (Street et al., 2012, p. 14).

The histograms by gender for LOS in our sample are shown in Figure 2.

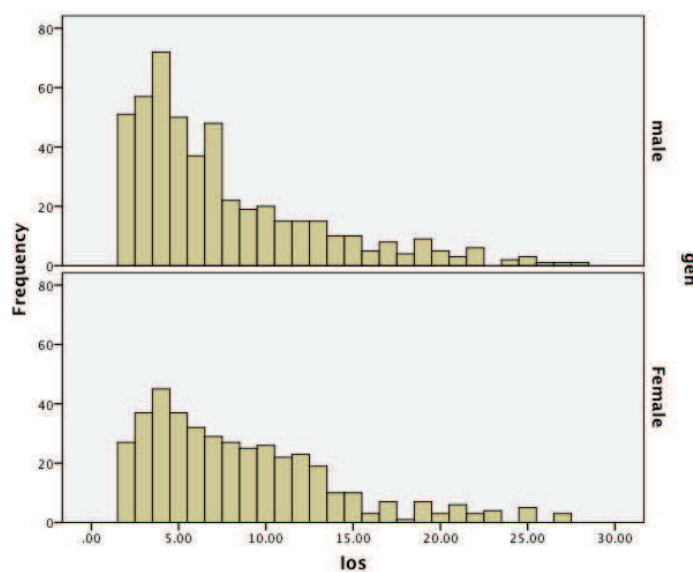


Figure 2: Histograms of length of stay

The histograms show that length of stay both for men and women have a skewed distribution where most episodes show a short LOS and just a few show LOS above 15 days. Note that the LOS distribution for males shows a higher preponderance for shorter LOS than that of females. Average LOS for females is 8.64 days while for males it is 7.57 days, which partly justifies their separate analysis.

Diagnoses - Number of different diagnoses in the 365 days prior to the episode being assessed, including the current diagnoses.

In the English NHS records, an inpatient episode would have a primary diagnosis and anything up to seven additional diagnoses, depending on the condition of the patient. The higher the number of diagnoses the higher the number of co-morbidities of the patient. We

attempted to reflect the varying complexities of different diagnoses and their impact on costs for COPD inpatient spells. We attempted this several ways. For example, we used the cost of the HRG that is most frequently linked with a certain diagnosis, as a way of ascertaining the disease weight reflected by the diagnosis. However, following various regressions we concluded that the way an additional diagnosis impacts the actual disease (COPD) being treated is almost impossible to isolate with the data at our disposal. That is, we may be able to attribute a level of severity to a certain diagnosis (e.g. through a Charlson index (see Charlson et al. 1987)), but a severe disease like cancer, may in fact have little, or less impact, on the COPD inpatient spell of a patient than say his/her hypertension. That is, we were not interested on the severity of a disease per se, but on the way that disease impacted the stay in hospital and treatment of a COPD inpatient spell. This proved impossible to establish.

Thus the sole information we could derive from the diagnoses linked to an inpatient spell is that the more the diagnoses accompanying a primary COPD diagnosis the higher the likelihood of the patient having an elevated clinical need. This interpretation is compatible with our use of the number of diagnoses as an output variable in Table 5.

We have used in our model the count of diagnoses going back 365 days from the current inpatient episode. The rationale for this is that the patient's medical history prior to an admission does convey information about the general condition of the patient, which in turn could affect his/her clinical needs at the point of admission. The cut off of one year look-back is subjective but accords with intuition that if a condition has not led to some admission in the preceding year then it is likely to be in abeyance or absent. The use of historical data one year pre-admission meant that admissions during the first year of our data (April/2004 – March/2005) could not be analysed for cost efficiency, as we would not have diagnoses going back to 2003/4.

LosPrevious adm - Sum of LOS of all admissions (in the previous 365 days)

In a further attempt to account for the complexity the patient presents on admission to an inpatient spell, we considered the aggregate length of hospital stays of the patient in the

year preceding the current admission, but excluding the LOS of the current inpatient spell. This is compatible with the notion that the longer a patient may have been in hospital prior to the current admission the more the disease weight the patient represents embarking in the current inpatient spell. This variable is constructed by looking at all admissions of the patient in the year prior to the current admission and summing the LOS of these previous admissions. A majority of patients had no admission in the year prior to their inpatient spell under consideration and so 60% of our observations had the value of zero for this output variable. The multitude of episodes where this output is zero means there is a good size sample of observations, where like with like is being compared when it comes to assessing LOS relative to disease weight and quality of treatment. Similarly, those episodes where this variable is positive are prevented from taking advantage of this output and putting a high weight on it to justify a high LOS even if the other two outputs are poor, because again the 40% sample size where 3 outputs are positive is large enough to ensure the model does not yield spurious estimates of potential reductions in LOS.

TTR – Time to readmission

This variable has been used to reflect the quality of the clinical care the patient has received. The short LOS for an episode, which then leads to readmission in a short time after discharge is taken to signal inadequate treatment during the hospital stay. What constitutes ‘short duration’ to re-admission is subjective and normally taken to be 28 days. We have doubled this and took any re-admission within the first 56 days after discharge to signal less than adequate care. Moreover, the longer the stretch to re-admission the better the quality of care. Taking the threshold to 56 days meant we could discriminate better on quality of care than if the threshold had remained at 28 days. The cut-off of 56 days is nevertheless still subjective. However, the longer this is taken the more the observations that cannot be assessed as their re-admission data would not be available in the last year of our data. Moreover, the longer the time to re-admission the more likely it is that the episode represents a new flare up of the chronic condition of COPD than a reflection of inadequate care in the previous inpatient spell. After 56 days of time to readmission we treated all re-admissions as if they reflect the same quality of treatment and attributed them a value of 56 for TTR. The same was the case where there was no re-admission. So, for the

purposes of our analysis a TTR of 56 represents reaching a satisfactory level of medical outcome upon discharge. In line with the foregoing, of the 900 episodes in our sample some 80% were assigned a TTR of 56.

As noted earlier we have excluded from our analysis episodes where the discharge status was death. This was on account of the fact that LOS and TTR in these cases is right-censored and so the episodes are not comparable with the rest of our episodes. Further, our analysis does not reflect cases where discharge was not death but the patient may have then died before he or she could be re-admitted within the time span of our data. Death outside the hospital is in fact confidential and it is therefore not possible to take into account these data. In total 13 episodes, whose discharge was death were excluded from the initial dataset.

Table 6 shows summary statistics of our input-output data per HRG. It is clear that patients with episodes in HRG 99 are on average older, than those in the other HRGs, and have a longer length of stay. In addition, when one counts the total number of diagnoses for each episode, on average patients in HRG D99 have 6.42 diagnoses, whereas patients in HRG D40 have only 3.95 on average. The average of LOS from previous admissions is higher for HRG D39 and the time to readmission is the lowest for patients in HRG D99, indicating that in HRG99 it is more likely patients will be readmitted following discharge.

Table 6: Episode characteristics per HRG and gender

		% male	age	LOS	# Diagnoses	LosPrevious adm	TTR
D39 (286 episodes)	Total Mean	47.91%	69.48	9.27	6.12	8.92	49.44
	Total StDev	---	3.13	5.91	2.84	17.03	14.90
	Mean Male		69.55	9.45	6.07	6.6	50.66
	Mean Female		69.42	9.09	6.16	11.05	48.32
D40 (496 episodes)	Total Mean	55.65%	70.28	7.07	3.95	6.04	49.39
	Total StDev	----	3.26	4.72	2.54	16.46	14.46
	Mean Male		70.39	6.49	4.12	5.9	49.39
	Mean Female		70.13	7.8	3.73	6.21	49.90
D99 (118 episodes)	Total Mean	64.41%	72.57	9.29	6.42	6.44	48.81
	Total StDev	----	1.82	6.08	2.76	12.34	16.41
	Mean Male		72.93	8.12	6.5	6.33	48.75
	Mean Female		71.9	11.4	6.29	6.64	48.93

From Table 6 it is also evident that male and female episodes exhibit different characteristics: average LOS is generally higher for females, except for episodes coded in HRG 39; and female cases have on average a higher number of days of previous stay in hospital than male cases. In conclusion we note that apart from the factors featuring as outputs in our model, and for which we thus allow in estimating the LOS for an inpatient episode, there can be other patient characteristics that may affect LOS and indeed the likelihood of patient readmission. These include the patient's propensity to take their medicines and to stick to prescribed post admission routines, the level of home care available to the patient on discharge, etc. These factors are patient-specific and could not be taken into account in our analysis for lack of data. This means that our estimates of scope to reduce length of stay or to increase TTR may be over-estimates.

5. The DEA Model Used

Our aim with the DEA model is to estimate the lowest the LOS of each inpatient episode could have been, compared to the rest of the episodes in our set, controlling for disease weight and quality of care as reflected in our output variables. Accordingly, we construct a non-oriented DEA model so as to give top priority to the minimisation of LOS of each episode. We gave second priority to the potential delay in 'time to re-admission (TTR)' of the patient, to enable the model to 'trade – off' between LOS and TTR. Finally the lowest priority went to estimating how much larger the number of diagnoses and the number of days LOS in the previous year could have been after LOS is minimised and TTR is maximised. This is purely for estimating Pareto efficient episodes as benchmarks, given that the last two output variables are exogenously fixed at admission of the patient. Note that the non-oriented model used was inspired by the geometric distance function (GDF) of Portela and Thanassoulis (2005, 2007), but the formulation as in this paper is an original one. See also section 3.3.3 of Thanassoulis et al. (2008) on the various types of non-radial efficiency measures in the context of DEA.

Consider COPD inpatient episode j ($j = 1, \dots, n$), \mathbf{X}_j reflecting the input LOS of episode j and y_{1j} reflecting the number of diagnoses, y_{2j} reflecting the sum of previous LOS and y_{3j} the TTR of episode j . In mathematical terms the model we propose is defined in (1).

$$\text{Min}_{\theta, \beta_1, \beta_2, \lambda_j} \left\{ \begin{array}{l} \theta - 0.05\beta_1 - 0.8\alpha\beta_2 \mid \sum_{j=1}^n \lambda_j x_j \leq \theta x_o, \sum_{j=1}^n \lambda_j y_{1j} \geq \beta_1 y_{1o}, \sum_{j=1}^n \lambda_j y_{2j} \geq \beta_1 y_{2o}, \\ \sum_{j=1}^n \lambda_j y_{3j} \geq \beta_2 y_{3o}, \sum_{j=1}^n \lambda_j = 1, \lambda_j \geq 0, \beta_1, \beta_2 \geq 1 \end{array} \right\} \quad (1)$$

Model (1) is non-oriented, and reflects an assumption of variable returns to scale (VRS) as intuitively we cannot assume LOS to be proportional to disease weight and TTR in the manner an assumption of constant returns to scale would require. The constraint that the β s should be at least 1 reflects the fact that the optimal LOS must cope with at least the disease weight and the TTR pertaining to the patient episode concerned. No constraint that θ should be less than 1 is imposed as we assume that under certain circumstances it may be desirable to increase the length of stay in order to improve the quality of care received (e.g. to avoid quick readmissions).

In order to reflect different priorities, we use a common weight of 0.05 on the expansion factors associated with disease weight (number of diagnoses and sum of LOS of previous admissions), a much higher weight of 0.8 for the expansion factor associated with TTR, and the maximum weight of 1 associated with LOS. These weights are subjective but are compatible with the rationale that the expansion of LOS and TTR should be given priority over disease weight, as the former two are controllable by the service hospital while disease weight is exogenously fixed. For the aforementioned priorities to be effective we needed to adjust the weight placed on β_2 (multiplying it by a constant α in the objective function of model (1)), to control for the fact that β_2 has a longer range of feasible values than β_1 and θ .

The range of feasible values for θ is 1 to 2/28 (i.e 1/14 to 1) while for β_1 it is from 1 to 15. As model (1) seeks to minimise the value of x_j^*/x_j for θ and the value of y_{1j} / y_{1j}^* for β_1 , where the * denotes benchmark (target) value for the variable concerned, in both cases the range of this ratio is about the same, to 1/14 for θ and 1 to 1/15 for β_1 . However, the corresponding range for β_2 it is from 1 to 1/ 56. To control for this we are using the scaling factor $\alpha = 15/56$

in the objective function to model (1). This in the context of the objective function to model (1) converts to 15 the maximum target of 56 for TTR. Thus in all 3 cases the range of observed to best benchmark value is approximately the same, namely 1 to 1/14 in the case of θ and 1 to 1/15 in the case of β_1 and β_2 .

Given the weights used here are subjective we carried out sensitivity analysis varying the weights on θ and β_1 keeping the weight on β_2 as scaled above relative to these two weights. Table 7 shows the optimal values of θ , β_1 and β_2 as the weights of weights on θ and β_1 vary.

Table 7: Sensitivity Analysis on the objective function weights of model (1)

Trial weights	Optimal values			# efficient	
	θ and β_1	θ	β_1		β_2
1 and 0.05		0.403	2.677	1.972	9
1 and 0.5		0.683	3.98	1.971	11
0.5 and 0.05		0.431	2.991	1.985	11

Clearly the weight on β_1 (disease weight) needs to be quite close to that on θ for it to have any impact on the optimal values of the θ and β variables. Such a high weight on β_1 is unjustified due to the exogenous nature of disease weight on admission. Thus our findings will be quite robust for sensible differences in weights between θ and β_1 .

The objective function of model (1) does not yield an efficiency score. However, such a score can be derived from the optimal values of β_1 , β_2 , and θ in model (1). Inspired by the geometric distance function (GDF) of Portela and Thanassoulis (2005, 2007), we define the efficiency score of each episode as $\theta / \sqrt{\beta_1 \beta_2}$, which works as a proxy for 'cost efficiency' of the assessed episode. The higher the value of the ratio $\theta / \sqrt{\beta_1 \beta_2}$ the better the performance of the hospital on the episode o assessed via model (1). This is because the higher the value of θ the lower the potential to reduce LOS, suggesting the hospital keeps patients closer to the shortest time observed for the disease weight of the patient concerned. The larger the value of the inverse of β_1 , and β_2 the lower the values of β_1 , and β_2 which means there is little scope to extend the TTR or to cope with even heavier disease weight than found at the episode being assessed. Clearly if θ , β_1 and β_2 are all 1, $\theta / \sqrt{\beta_1 \beta_2}$ will also be 1 and the episode is efficient in the sense that it cannot improve further on any one of its observed

values on LOS, TTR and disease weight compared to other episodes in our set. However, clearly $\theta / \sqrt{\beta_1 \beta_2} = 1$ does not in itself mean the episode cannot improve on any one of θ , β_1 and β_2 . Note that slacks are non-existent on the input side, nor for the output TTR, and only one of the two slacks associated with the two disease weight outputs could be positive at the optimal solution of (1). This aspect of the performance of the hospital is not reflected in our ratio measure as these are ‘exogenously fixed’ outputs as noted earlier. However, any slacks in these variables could be taken into account in any comparison of episodes on a case by case basis if comparisons are an issue.

We also tested a model where the non-linear efficiency score $\frac{\theta}{\sqrt{0.05\beta_1 0.8\alpha\beta_2}}$ (weighted the same way as the linear function in model (1)) was used on the objective function of model (1). The results from this non linear model are very close on average with those derived using the linear objective function in model (1), though there are differences for some episodes between the two models. In the remainder of the paper we report the results from model (1) with the linear objective function.

6. The Findings from the COPD Episodes and Their Potential Use in Practice

6.1 Overview of findings

We assessed separately the male from the female episodes and, as noted earlier, in both cases only patients between 65 and 75 years of age were considered. In both sets of episodes we first attempted to identify outliers in the sense of episodes having a very low LOS for the disease weight of the patient, which could set an unrealistic benchmark for the rest of the episodes. Outliers were identified through an analysis of the efficiency scores of θ , β_1 and β_2 in model (1), through the number of times each episode appeared in the reference set of other episodes, and also through super-efficiency scores. The reasons for deeming an episode outlier were not the same in all cases. For males, in the set of 489 episodes at our disposal we identified only 1 outlier episode. This was because when this episode was included in the reference set more than 90% of the remaining episodes used it as their sole

reference (revealing a high influence of a single episode). For females, two episodes revealed influential in the foregoing sense. Computing efficiency without allowing these two episodes in the reference set revealed high super-efficiency scores (particularly for β_1 (value of 0.22 leading to a super-efficiency score of 455%, corresponding to the maximum) in one case, and for θ , β_1 and β_2 (score of 115% for θ , of 133% for $1/\beta_1$, and maximum score of 108% for $1/\beta_2$) in the other), suggesting that these 2 episodes should not be allowed in the reference set. After re-running the models without allowing these two units in the reference set, one additional female episode revealed influential (appearing 409 times out of 411 as a reference to the other episodes) and its super-efficiency scores also revealed that the episode should not be considered in the reference set of other episodes. To sum up, 3 female episodes were considered outliers and just 1 male case was considered so.

In the remaining episodes some 11 female and 7 male episodes were respectively at the frontier in the sense that no scope to reduce LOS or raise output levels was identified. In these frontier episodes the foregoing outliers are included although they were not allowed to define the location of the frontier.

In what follows we report the optimal θ and β variables as obtained through model (1). We do recognise, however, that confidence intervals for each value could lead to more robust conclusions. Such confidence intervals could in principle be estimated through bootstrapping techniques (see e.g. Simar et al., 2012). However, the development of bootstrapping methodologies for non-oriented models of the type we are using is beyond the scope of this paper.

The average female LOS efficiency (θ value in model 1) was 40.3% and average male LOS efficiency 43.1%. This suggests the average episode LOS can be more than halved! Using the θ values to compute 'efficient' LOS at episode level it is found that the total scope to reduce length of stay is an aggregate of 2508 days for females and 2637 days for males. These figures represent 71% of total length of stay both for males and females. On average length of stay could be reduced by 6.1 days from 8.64 days for females and down by 5.4 from 7.57 days for males. Thus large reductions in LOS are possible in principle at the aggregate level. This finding is in line with that in Berk and Moinzadeh (1998), applied to

intensive coronary care, who suggest that early discharge can be an effective way of increasing a hospital’s capacity without detriment of quality of care.

Our model gave the second highest priority (after LOS) to scope for improving the quality of care, captured through time to re-admission (TTR). The potential for improving re-admission at episode level is reflected in the optimal value of β_2 . Finally our model, through the optimal value of β_1 , also reflects for an episode how much more heavy the disease weight of the person could have been for the optimal LOS and TTR determined. In average terms the optimal values obtained for β_1 and β_2 are shown in Table 8.

Table 8: Average of observed outputs in relation to potential outputs

	β_1	β_2
Female episodes	2.68	1.97
Male episodes	2.66	1.88

The average values of β_1 are very similar for female and male episodes, meaning that existing secondary diagnoses or prior LOS days could have been more than twice those recorded and still the reduced LOS would have been sufficient. These interpretations, however, need to be treated with caution as our model does not discriminate on secondary diagnoses or prior LOS inpatient days as to the degree of relevance they have for affecting LOS for an inpatient COPD case at admission. A key finding, however, is that for many episodes, a reduced length of stay would have been sufficient, even if the episode had had higher level of comorbidities with COPD.

As noted earlier some 60% of our observations had no prior admission as inpatient. This could in principle offer the remaining 40% of episodes the opportunity to raise their efficiency artificially by placing a high weight on this output variable to outperform ‘artificially’ those episodes where the variable has a zero value. To test the susceptibility of our model to this possibility we solved it without the output variable SumLos. We found, as we expected, little difference in the results. The scope to increase outputs increases slightly (β_1 rises from 2.68 and 2.66 to 2.76 and 2.89 and β_2 from 1.97 and 1.88 to 1.99 and 1.88 for female and male episodes respectively). The slightly higher β values mean that efficiency values deteriorate slightly by the exclusion of this variable (the θ values also decreased

slightly). The change in efficiency values is small because the number of episodes with zero and positive SumLos value are still significant in absolute terms so as to provide a representative comparative set in each category of episode. In view of this finding, and in order to allow for patients with several previous stays in hospital, SumLos was kept in the model and these are the results being reported in this paper.

Regarding time to readmission (TTR) the average value of β_2 is 1.97 for female episodes and 1.88 for male episodes. This suggests the 'efficient' LOS still offers scope for time to re-admission to rise substantially. Caution is needed in using the numerical values of β_2 . This is because we have right-censored data on TTR when re-admission might have been beyond the last year for which we had data. Moreover, we have not discriminated within the model for re-admissions beyond 56 days from discharge. So in fact the β_2 values should be taken as indicative (in a monotonic sense) of scope for rise in time to re-admission rather than as exact estimates of potential.

In order to better understand why there is scope to reduce LOS and raise TTR in some episodes we compared episodes at the top quartile of efficiency (efficiency being defined as $\theta / \sqrt{\beta_1 \beta_2}$) with those at the bottom (see Table 9). The bottom 25% episodes show the largest average length of stay for treating patients who have the least complex conditions, as measured by the number of previous diagnoses and the cumulative LOS of previous admissions. In addition, these episodes also show a low value for the time to readmission, suggesting that some patients were not treated with the adequate level of quality.

Table 9: Characteristics of top and bottom episodes on efficiency (Average values)

Female	LOS	#Diagnoses	LOSPrevious adm	TTR	N	%D39	%D40	%D99
Bottom 25%	14.4	3.98	2.71	43.50	103	7.54%	13.14%	4.38%
TOP 25%	4.25	5.92	16.09	52.92	103	8.27%	14.11%	2.68%
100%	6.15	10.92	50.23	50.23	13	1.05%	0.5%	0.85%
Male								
Bottom 25%	13.47	4.24	2.83	44.10	122	7.98%	12.68%	4.29%
TOP 25%	3.50	6.45	12.17	50.81	127	6.54%	14.52%	4.91%
100%	4.44	10.22	54.11	48.78	9	0.17%	0.6%	0.85%

Note that both for male and female episodes those with 100% efficiency show the highest average complexities (as measured through the number of diagnoses and the cumulative sum of days of previous inpatient stays), the highest time to readmission and yet a lowest length of stay. A key difference between the most and the least efficient episodes is the LOS and the LOS of previous admissions. Episodes at the bottom 25% on efficiency display clearly signs of concern, as they relate to patients that do not show, on average, high disease weight and yet have stayed for a long time in the hospital and were on average readmitted soon after discharge. In Table 9 we also show the percentage of episodes from each of the HRGs considered that fall into the three categories of episodes. We see that the percentage of episodes from each HRG falling in the top and bottom efficiency band are very similar suggesting there is no association between cost efficiency of episode and its HRG designation.

The episodes showing less than 100% efficiency have scope either to reduce LOS and/or to improve TTR and cope with heavier disease weight. Figure 3 categorises episodes on the basis of the areas where they have scope for improvement. We present the results in aggregate for both genders rather than by gender as the results were very similar across genders. Some 90% of COPD inpatient episodes have scope to change LOS and of these the vast majority (99.3% of cases) have scope to reduce LOS. The rest (only 6 cases) have scope to increase LOS. The latter 6 cases all have scope to increase TTR (and no scope to increase the other outputs). This suggests that for these 6 episodes the short length of stay might have been at the expense of the quality of their treatment.

For the vast majority of episodes, those where there is scope to decrease LOS, there are various additional improvements possible. The most prevalent one is that the episode could have also presented a heavier disease weight (as measured through the number of diagnoses and the sum of previous LOS). Some 69% of episodes where LOS can be reduced could have presented heavier disease weight. Other than this, some 5% of episodes had potential to reduced LOS and simultaneously increase TTR (but not the other outputs) and the rest in approximately equal numbers could either improve on all outputs or on none (see Figure 3).

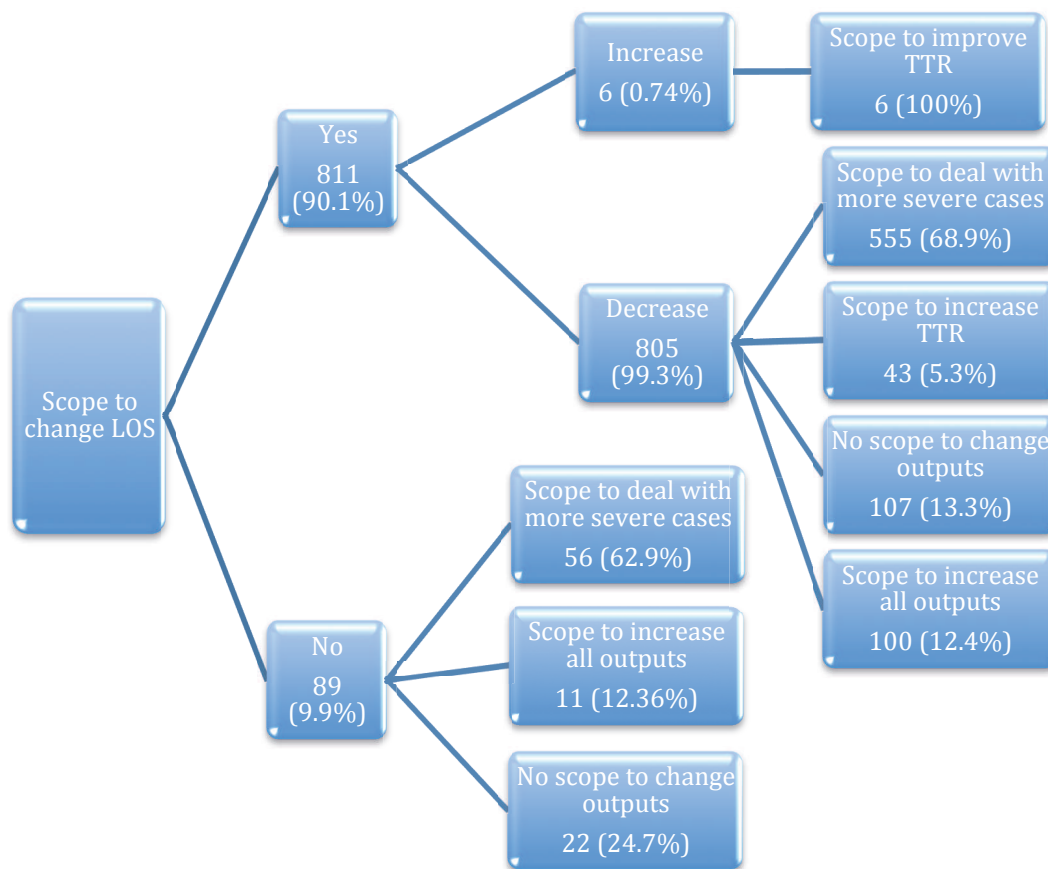


Figure 3: Categorisation of episodes by nature of potential improvements

For the 89 cases for which no scope to reduce LOS was found, only 22 lie at the best practice frontier, in the sense that no scope to change outputs was identified either. The remaining cases cannot be considered 'efficient' because, although LOS cannot be reduced, it was found that episodes with identical LOS had dealt with more severe health conditions (in 63% of cases), or with more severe health condition and yet had longer time to readmission (12% of cases). In a practical setting, for episodes where the model identifies no scope to reduce LOS yet there is scope to treat cases with heavier disease weight and/or lengthen the time to re-admission, one can investigate on a case by case basis whether LOS could indeed have been shorter – something our model could not identify directly.

6.2 Using the Findings to Identify and Disseminate Best Practice in Attaining Cost Efficiency in the Treatment of COPD

In this section we look in more detail at the findings from the perspective of a hospital that wishes to improve cost efficiency in treating COPD admissions. We will do this by group of episodes as shown in Figure 3. The groups reflect different combinations of efficiency gains feasible in principle.

We start with the group of episodes for which there is scope to reduce LOS while dealing with even heavier disease weight. This group of episodes is the largest for both males and females and it is one where quality of care is good in the sense that we do not identify scope for extending time to re-admission (TTR). In Table 10 we show the average of observed and target values as identified through DEA for this group.

Table 10: Average observed and target values when $\beta_1 > 1$ $\beta_2 = 1$ and $\theta < 1$

	Female		Male	
	Target by DEA	observed	Target by DEA	observed
Los	2.16	9.04	2.00	8.28
#Diagnoses	8.44	3.69	9.46	4.33
LosPrevious adm	8.63	1.22	5.12	2.34
TTR	56	56	56	56

There is a large scope for decreasing length of stay in this subset of episodes. Average LOS can drop from about 9 days for females and 8.3 days for males, to about 2 days in both cases. Notwithstanding the potential for reduced LOS it is also found that previous diagnoses and cumulative previous LOS could have been double from that actually observed for males. For females cumulative prior LOS could have been over 7 times that observed! These large factors could be in part a consequence of the fact that many of the prior stays in hospital and diagnoses often have little bearing on the required LOS for a given admission for COPD. Thus both Male and Female inpatient COPD episodes of this category show substantial dissimilarity from benchmark episodes both in LOS and in disease weight as captured through cumulative prior LOS and diagnoses.

While at an individual case there may be good reasons for the 'cost inefficiency' through

prolonged LOS we identify here, at the aggregate level of several hundred episodes many individual effects should cancel out leaving the suggestion that a considerable part of the scope for savings identified could be true. It is noteworthy that quality of care as captured by TTR for these episodes is satisfactory in the sense that we do not identify on average early re-admissions. LOS is clearly affected by operating practices beyond the patient's medical condition. For example Singh KC and Terwiesch (2011), find that LOS and mortality rate in the case of cardiac episodes are affected by hospital focus in terms of type of patient admitted.

In contrast to the foregoing the vast majority of episodes where LOS could be reduced, we have also identified a small number (3 female and 3 male episodes), for which LOS could have been longer. Table 11 shows the details.

Table 11: Average observed and target values for episodes where LOS could rise.

	Female		Male	
	Target by DEA	observed	Target by DEA	observed
Los	7.56	6.00	6.26	4.67
#Diagnoses	13	13	12.67	12.67
LosPrevious adm	53	53	27	27
TTR	44.34	9	49.81	11.33

Both for female and for male episodes the average observed LOS is very low in the face of the disease weight of the patient. Moreover, readmission has occurred on average very soon after discharge. The model has identified potential to increase the TTR at the expense of increasing LOS. It suggests that increasing LOS by about 34% for males and 26% for females, time to readmission rises to values that on average are close to our threshold of 56 days.

Finally there is a group of episodes which have the lowest LOS available in our data (2 days) and yet they have the potential to have dealt with heavier disease weight than was the case. See Table 12. There were 20 female and 36 male cases, in this category. In all these cases TTR was long enough to suggest quality of care was appropriate. It could be lower LOS was feasible in these cases but our model could not identify it because we did not consider episodes with less than 2 days LOS.

Table 12: Average observed and target values for subgroups where $\beta_1 > 1$ $\beta_2=1$ and $\theta=1$.

	Female		Male	
	Target by DEA	observed	target by DEA	observed
Los	2.05	2.05	2.00	2.00
#Diagnoses	7.82	3.05	9.63	4.19
LosPrevious adm	5.62	0.60	4.39	1.94
TTR	56	56	56	56

6.3 The Hospital Effect

One of the aims in our analysis was to identify whether certain hospitals showed systematically higher cost efficiency than other hospitals and to draw lessons from this regarding good quality but cost-efficient treatment protocols. There are several hospitals in our data set, but just three have a sufficiently large volume of episodes to make possible a comparison between them. Therefore, in what follows we will consider just these three hospitals coded as hospital 1, 2, and 3 respectively. These 3 hospitals are responsible for 826 cases. Table 13 shows the mean adjustment factors (θ , β_1 and β_2) obtained from the results of the pooled assessment across all hospitals and not just the 3 hospitals under analysis.

Table 13: Results per hospital (pooled assessment)

		<i>Scope to change LOS Mean (θ)</i>	<i>Scope to deal with more severe cases Mean (β_1)</i>	<i>Scope to increase TTR Mean (β_2)</i>	<i>N</i>	<i># efficient</i>
Male	hospital 1	0.44	2.82	1.83	159	4
	hospital 2	0.45	2.72	1.78	114	2
	hospital 3	0.40	2.55	2.04	166	2
Anova and K_W (p values)		0.11 & 0.35	0.486 & 0.99	0.93 & 0.56	439	
Female	hospital 1	0.38	3.27	1.28	97	0
	hospital 2	0.37	3.03	1.38	71	0
	hospital 3	0.42	2.26	2.58	219	13
Anova and K_W (p values)		0.13 & 0.26	0.001 & 0.000	0.035 & 0.006	387	

Looking at male episodes first it is seen that the mean β values are similar across the 3 hospitals and so disease weight and TTR are similar in terms of potential to improve. However, the model does identify some differences in mean potential to reduce LOS between hospitals, even though these differences are not statistically significant (Kruskall-Wallis or Anova p statistic is above 5%). Nevertheless, there is a full 5 percentage points

difference in the mean θ value between hospitals 2 and 3, with hospital 3 having scope to reduce LOS on average by a further 5 percentage points compared to hospital 2. Hospital 1 also has more scope to reduce LOS than hospital 2.

Looking at female episodes it is again found that hospital 3 is different from hospitals 1 and 2. Hospital 3 shows this time less scope than the other 2 hospitals to reduce LOS but has significantly higher scope to raise TTR (in other words female re-admissions are on average much sooner in hospital 3 than is the case of the other two hospitals). This is to an extent perhaps explained by the fact that hospital 3 deals with heavier disease weight than the other two hospitals as indicated by the lower mean β_1 value for hospital 3.

What is perhaps more interesting is not the differences between hospitals but rather the differences within-hospital, between the genders, as depicted in Table 13. Hospitals 1 and 2 show considerably more scope for reducing LOS in the treatment of female than that of male COPD cases (6 and 8 percentage points difference in the pooled assessment of hospitals). Hospital 3 differs from hospitals 1 and 2 in having only 2 percentage points difference in the scope to reduce LOS between male and female COPD cases, and in the fact that the scope to reduce LOS is lower for female cases.

Clearly each hospital would do well to examine its treatment protocols on COPD cases and identify any differences between them as applied to different genders in order to draw appropriate conclusions. For example hospital 3 does appear to have significantly earlier both discharges and re-admissions for female patients both compared to its own treatment of male COPD cases and compared to the other 2 hospitals where females are concerned. This could suggest either a treatment policy or implementation of it by different care teams varying between female and male COPD cases at hospital 3.

An additional way in which we can look for hospital differences is to assess episodes within-hospital, by gender. The identified scope for LOS or potential improvements on TTR will no longer be comparable across hospitals. However, it will still provide evidence of the comparative degrees of dispersion on LOS and TTR both within and across hospitals. Table 14 shows details of the within-hospital within gender assessments. We show in Table 14 just

the scope to change LOS for the within-hospital and for the pooled assessments, as reduction in LOS is the main driver of cost cuts. Note that we also show the hospital effect on LOS (the ratio of the mean efficiency (i.e. scope to change LOS) in the pooled to the mean efficiency derived from the within-hospital assessment). This ratio is in the spirit of Portela and Thanassoulis (2001), where the hospital effect is an average of the ratio of pooled efficiency to within-hospital efficiency, obtained for each episode. Note that the outliers identified for the pooled assessment were also treated as outliers for the within-hospital assessments for ease of comparison.

Table 14: Scope to reduce LOS and to increase outputs identified within-hospital and pooled assessments

Male episodes			
	Scope to change LOS Within- hospital assessment Mean (θ)	Scope to change LOS Pooled Assessment Mean (θ)	hospital effect on LOS Mean (Pooled/within) assessment
hospital 1	0.47	0.44	95.3%
hospital 2	0.51	0.45	91.5%
hospital 3	0.43	0.40	95.1%
Female episodes			
hospital 1	0.42	0.38	98.3%
hospital 2	0.47	0.37	81.0%
hospital 3	0.42	0.42	100.0%

Contrasting the values for the scope to reduce LOS in the pooled assessment and the within-hospital assessment (right most column in Table 14) reveals the distance of the hospital lowest LOS frontier from that of the pooled frontier. In the case of hospitals 1 and 3 scope for LOS reduction by gender is similar whether seen relative to the pooled or the within-hospital frontier. E.g. for hospital 1 the scope is down to 47% for males within-hospital and 44% within pooled frontier. (The figures for females are 42% (within) and 38% (pooled)). This means hospitals 1 and 3 have within-hospital frontiers very close to the pooled frontier, and therefore the hospital effect on LOS is higher. In contrast hospital 2 has a within-hospital frontier at considerable distance from the pooled frontier, especially so for female episodes where the scope to reduce LOS is down to 37% compared to 47% when the referent level is pooled and within-hospital, respectively. In practical terms this suggests that best practice in terms of lowering LOS in hospital 2 (controlling for disease weight and

TTR of patient) is considerably worse than is the case for all other hospitals. So there is something systematic (or institutional) in the COPD treatment practices of hospital 2 that set it apart from the rest of the hospitals involved in the study.

Note that the above conclusions need to be seen in light of unbalanced sample sizes for female episodes: Hospital 3 accounts for about 57%, while the other two hospitals for about 43% of female episodes. This results in very unbalanced samples with hospital 3 having more than double the episodes of hospital 1 and more than three times the episodes of hospital 2. Such unbalanced sample means that there may be some bias in the computation of the hospital's effects for female episodes due to unequal sample sizes (see e.g. Simpson (2005)).

6.4. The Consultant Effect

Where there are multiple care teams (usually led by a consultant) within a hospital, the results aggregated at hospital level could mask significant differences in operating practices and judgements made at care team level within the hospital. We have attempted to look for what we would call 'consultant' effect by aggregating the results from the pooled assessment to consultant (care team) level. We present here findings only in relation to consultants who cared for at least 20 COPD inpatient episodes each. The findings are summarised in Table 15. Note that we have added for each consultant the findings on both female and male episodes from the within gender pooled assessments conducted earlier.

We observe significant differences between consultants both within and across hospitals. Consultant 12 operating in hospital 3 has scope to reduce LOS to nearly a third ($\theta= 0.34$) at one extreme while consultant 3 operating at hospital 2 can only reduce LOS about to two thirds ($\theta= 0.56$). It is recalled that our analysis has controlled for disease weight and quality of treatment by way of TTR. The largest difference between consultants within the same hospital in the scope to reduce length of stay is observed for hospital 2 with a difference of 19 percentage points (C3-C5). The ranges for the other hospitals are also very high, with hospital 1 showing a range of 9 percentage points and hospital 3 a range of 10 percentage points between individual consultants. Taking the case of hospital 2, where consultant 3

shows the highest score on LOS (and lowest scope to reduce it) and consultant 5 shows the highest scope to reduce LOS, it can be seen that these differences may be the result of different styles of practice. Indeed, in terms of scope to have dealt with heavier disease weight patients the two consultants show similar values, but regarding the scope to improve TTR consultant 3 shows a higher scope, suggesting that the team's lower lengths of stay may be at the cost of shorter times to re-admission than is the case for Consultant team 5. These results may initiate some further analysis on the care practices of different consultants. For example C5 is distinct within hospital 2, because on average his/her team has the highest average LOS (8.22 days) compared to average LOS of about 6 days for other consultants' teams.

Table 15: Results per consultant

	Con- sul- tant	N	Scope to change LOS Mean (θ)	Scope to deal with more severe cases Mean (β_1)	Scope to increase TTR Mean (β_2)
Hospital 1	C1	28	0.51	3.54	1.87
	C2	27	0.42	2.38	1.29
Hospital 2	C3	34	0.56	3.25	2.02
	C4	32	0.47	2.95	1.47
	C5	23	0.37	3.31	1.48
	C6	22	0.46	2.34	2.54
Hospital 3	C7	50	0.42	2.21	2.49
	C8	41	0.41	1.95	2.92
	C9	41	0.39	2.52	3.69
	C10	38	0.41	2.51	2.85
	C11	37	0.39	2.49	2.32
	C12	37	0.34	1.76	2.78
	C13	37	0.44	2.61	1.57
	C14	33	0.44	2.84	1.63
	C15	31	0.43	2.04	2.14

In spite of the above attempt to compare consultants, we do recognise that more cases would be needed for each consultant in order to draw more robust conclusions regarding different styles of practice amongst doctors.

7. Conclusion

This paper presents a DEA assessment of COPD inpatient episodes focusing on cost savings that can be obtained from the reduction of length of stay. In assessing scope for cost savings we control for disease weight of the episodes being assessed through age, gender, the number of diagnoses in the 365 days prior to the current admission and the sum of the days of stay in hospital in the preceding year. In addition we also take account of quality of treatment as measured through time to readmission (on the assumption that a early readmission after discharge may signal bad quality of treatment).

Some of the variables used are censored and this implied some adjustments in the DEA model used. In particular a non-oriented model was required such that the model would be able to distinguish on performance between the episodes that showed the lowest length of stay (2 days) on the sole input used. Contributions of the paper at the empirical level relate to the use of patient level data in order to capture co-morbidities affecting costs of treatment as accurately as possible. The paper provides a template for the use of DEA to assess potential savings in treating COPD inpatient episodes. We illustrate how episodes once analysed for cost efficiency at in-hospital spell can be aggregated suitably at different levels - consultant, hospital, all hospitals - in order to identify and disseminate best practice for cost-efficient good quality treatment of COPD cases.

Our findings provide a useful input for hospital managers interested in improving productivity, a key consideration in publicly funded health services. We find scope for significant possible savings in length of stay in the treatment of COPD episodes, of the order of 70%. Moreover, we find on the face of it significant unexplained differences in length of stay both across hospitals and within hospitals by gender. Further, at consultant level we find unexplained differences in length of stay both across consultants within and across hospitals and even within consultant across genders. We have controlled for age and gender and most importantly primary diagnosis but beyond that we were forced to use surrogate variables to control for patient condition. This combination may not reflect accurately patient condition, and so our findings cannot be taken at face value to reflect potential savings. Nevertheless, the identified scope is so significant in scale and the differences between hospitals and certain consultants so significant in length of stay that

further investigation could prove valuable to identify cost-efficient operating practices for treating COPD episodes without detriment to the quality of care provided.

Our approach lays a path for investigation at the micro level, by DEA more broadly in the provision of health services. For example, we do see significant parallels in pharmaceuticals where the direction of travel is very clearly towards targeting of modified drug agents to the patient's own genomics. Our patient episode level analysis could be suitably extrapolated to patient level treatment by drugs where for example the unit of assessment might be the patient-drug combination. DEA might provide insights into the micro efficiency (however defined) of drug interventions at patient level. These findings in turn can then be aggregated to drug category, patient category, clinic level etc.

Key requirement for studies at patient level of the type we advocate is the availability of properly anonymised, very large, truly representative and full person-linked datasets. Such data is needed to tap the vast knowledge that amalgamated medical records hold, but remains hidden from view.

References

- Afonso, A.S.M., Verhamme K. M.C., Sturkenboom M.C.J.M., Brusselle, G.G.O. (2011). COPD in the general population: Prevalence, incidence and survival. *Respiratory Medicine* 105, 1872-1884.
- Berk, E. and Moynihan, K.(1998) The Impact of Discharge Decisions on Health Care Quality. *Management Science* 44(3): 400-415.
- Blunt, I. and Bardsley, M. (2012) Patient-level costing: Can it yield efficiency savings? Research Summary, Nuffieldtrust
- Bradford, W.D., Andrew, N. Kleit, M., Krousel-Wood, A., and Richard N. Re. (2001). Stochastic frontier estimation of cost models within the hospital. *The Review of Economics and Statistics*, 83(2): 302-309
- Charlson, M., Pompei, P., Ales K.L., Mackenzie, C.R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases* 40. 373-383.
- Cooper, W.W., Thompson, R.G., and Thrall, R.M. (1996). Extensions and New Developments in DEA. *Annals of Operations Research* 66, pp.3-45.
- Daidone, S. and Street, A. (2011). Estimating the costs of specialized care, CHE – Centre for Health Economics, Research paper 61. The York University
- Dervaux, B., Leleu, H., Minvielle, E., Valdmanis, V., Aegerter, P., Guidet, B. (2009). Performance of French intensive care units: A directional distance function approach at the patient level. *Int. J. Production Economics* 120, 585-594.
- Diez-Roux, A. V.(2000). Multilevel analysis in public health research. *Annu. Rev. Public Health*, 21, 171-92.

Dismuke, C. E. and Sena, V. (2001). Is there a Trade-Off between Quality and Productivity? The Case of Diagnostic Technologies in Portugal, *Annals of Operations Research* 107, 101–116.

Dormont, B. and Milcent, C. (2004). The sources of hospital cost variability, *Health Economics* 13, 927-939.

Hollingsworth, B. (2008). The measurement of efficiency and productivity of health care delivery. *Health Economics* 17(10): 1107-1128.

Kristensen, T., Laudicella, M., Ejersted, C. and Street, A. (2010). Health Economics Cost variation in diabetes care delivered in English hospitals. *Diabetic Medicine*, 27, 949–957.

Laudicella, M. Olsen, K. R. and Street, A. (2010). Examining cost variation across hospital departments: a two-stage multi-level approach using patient-level data. *Social Science & Medicine* 71, 1872-1881.

Olsen, K.R. and Street, A. (2008) The analysis of efficiency among a small number of organizations: how inferences can be improved by exploiting patient-level data. *Health Economics*, 17. 671-681.

Puig-Junoy, J., (1998). Technical efficiency in the clinical management of critically ill patients. *Health Economics* 7, 263–277.

Portela, M.C.A.S., and Thanassoulis, E. (2001). Decomposing School and School Type Efficiency. *European Journal of Operational Research*, 132/2, pp. 114-130.

Portela, M.C.A.S., and Thanassoulis, E. (2007). Developing a Decomposable Measure of Profit Efficiency Using DEA. *Journal of the Operational Research Society* 58, 481-490.

Portela, M.C.A.S., and Thanassoulis, E. (2005), Profitability of a sample of Portuguese bank branches and its decomposition into technical and allocative components, *European Journal of Operational Research* 162/3, 850-866.

Simar, L., Vanhems, A. and Wilson, P.W. (2012). Statistical inference for DEA estimators of directional distances. *European Journal of Operational Research* 220/3, 853-864.

Simpson, G. (2005). Programmatic efficiency comparisons between unequally sized groups of DMUs in DEA. *Journal of the Operational Research Society* 56(12): 1431-1438.

Singh KC,D. and Terwiesch, C. (2011) The Effects of Focus on Performance: Evidence from California Hospitals. *Management Science* 57(11): 1897-1912.

Sørensen, T.H., Olsen, K.R., and Gyrd-Hansen, D. (2009) Differences in general practice initiated expenditures across Danish local health authorities – a multilevel analysis, *Health Policy* 92, 35-42.

Street, A., Kobel, C., Renaud, T. Thuilliez, J. (2012). How well do diagnosis-related groups explain variations in costs or length of stay among patients and across hospitals? Methods for analyzing routine patient data. *Health Economics*, 21 (suppl.2). 6-18.

Thanassoulis, E, M.C.A.S. Portela, and O. Despic (2008), “Data Envelopment Analysis – The Mathematical Programming Approach to Efficiency Analysis”, In Fried, H.O., Lovell, C.A.K., and Schmidt, S.S., *The Measurement of Productive Efficiency and Productivity Growth*, Oxford University Press, 251-420.