

HOW WELL DO DIAGNOSIS-RELATED GROUPS EXPLAIN VARIATIONS IN COSTS OR LENGTH OF STAY AMONG PATIENTS AND ACROSS HOSPITALS? METHODS FOR ANALYSING ROUTINE PATIENT DATA[†]

ANDREW STREET^{a,*}, CONRAD KOBEL^b, THOMAS RENAUD^c, JOSSELIN THUILLIEZ^{d,e}
ON BEHALF OF THE EURODRG GROUP[†]

^aCentre for Health Economics, University of York, York, UK

^bDepartment of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Innsbruck, Austria

^cInstitut de Recherche et Documentation en Économie de la Santé, Paris, France

^dCentre d'Économie de la Sorbonne, University Paris 1, Paris, France

^eEcole des hautes études en santé publique, Rennes, France

ABSTRACT

We set out an analytical strategy to examine variations in resource use, whether cost or length of stay, of patients hospitalised with different conditions. The methods are designed to evaluate (i) how well diagnosis-related groups (DRGs) capture variation in resource use relative to other patient characteristics and (ii) what influence the hospital has on their resource use. In a first step, we examine the influence of variables that describe each individual patient, including the DRG to which the patients are assigned and a range of personal and treatment-related characteristics. In a second step, we explore the influence that hospitals have on the average cost or length of stay of their patients, purged of the influence of the variables accounted for in the first stage. We provide a rationale for the variables used in both stages of the analysis and detail how each is defined. The analytical strategy allows us (i) to identify those factors that explain variation in resource use across patients, (ii) to assess the explanatory power of DRGs relative to other patient and treatment characteristics and (iii) to assess relative hospital performance in managing resources and the characteristics of hospitals that explain this performance. Copyright © 2012 John Wiley & Sons, Ltd.

KEY WORDS: DRGs; hospital costs; length of stay; hospital performance

1. INTRODUCTION

Since the development of the first classification of diagnosis-related groups (DRGs) in the 1970s (Fetter *et al.*, 1980), the number of DRG systems has proliferated, with many countries developing their own versions, which are periodically overhauled (Kobel *et al.*, 2011). The original intention was that DRGs would classify patients into a manageable number of resource homogenous groups (Fetter *et al.*, 1980), and this remains the fundamental basis for classification. However, herein lies a puzzle: is variation in medical practice and resource use so great across countries that each requires its own patient classification system? Or are some DRG systems better than others at categorising patients into resource homogenous groups?

*Correspondence to: Centre for Health Economics, University of York, York YO10 5DD, UK. E-mail: andrew.street@york.ac.uk

[†]www.eurodrdg.eu/EuroDRG_group.pdf. Funding: The presented results were conducted within the research project 'EuroDRG - Diagnosis Related Groups in Europe: towards efficiency and quality', which was funded by the European Commission under the Seventh Framework Programme. Research area: HEALTH-2007-3.2-8 European System of Diagnosis-Related Groups, Project reference: 223300

Table I. Diagnostic and procedural codes used to identify patients to each EoC

EoC	Main diagnosis code (<i>ICD-10</i>)	Procedure code (<i>ICD-9-CM</i>)
Acute myocardial infarction	I21, I22	exclude 36.1 ^a
Appendectomy	K35-K38	47.0
Breast cancer surgery	C50, D05	85.20–85.23, 85.33–85.36, 85.4
Childbirth ^b	Z37, O80–O84	72, 73, 74
Cholecystectomy	K80	51.2
Coronary artery bypass graft	–	36.1
Inguinal hernia surgery	K40	17.1, 17.2, 53.0, 53.1
Hip replacement	–	00.7, 81.51–81.53
Knee replacement	–	00.8, 81.54–81.55
Stroke	I61, I63, I64	–

^aAll cases that receive a bypass are to be excluded.

^bIn each country, the definition that best matches national coding habits should be used. The aim is to include all cases of childbirth.

To address these questions, we consider patients admitted to hospitals in 10 European countries¹ for one of 10 conditions, which we define as *episodes of care* (EoCs). The EoCs are listed in Table I. The number and construction of DRGs used to categorise patients in each EoC varies markedly across these 10 countries.

The empirical articles in this special issue of *Health Economics* apply a common analytical strategy to examine the DRG classifications used in each of these countries in terms of how well they explain variations in resource use for patients having the same EoC. In this article, we set out the strategy adopted to examine why resource requirements might vary among patients. Our approach builds on the literature that examines hospital costs using patient-level data (McClellan, 1997; Kessler and McClellan, 2002; Dormont and Milcent, 2004; Olsen and Street, 2008; Laudicella *et al.*, 2010; Dormont and Milcent, 2005; Bradford *et al.*, 2001).

We have two types of dependent variables, either cost (y^c) or length of stay (LoS; y^s), both of which can be considered as measures of resource use (Martin and Smith, 1996; Gilman, 2000; Norton *et al.*, 2002). Although the analysis of cost variation is preferred, patient-level cost data are not available for all countries. Where cost data are absent, variation in LoS is examined instead.

LoS has the advantage that it is defined in a straightforward manner, calculated as the difference between the date of discharge and the date of admission. The disadvantage, of course, is that LoS is an imperfect indicator of resource use, particularly for surgical patients. In theory, costs should better reflect actual incurred resource use than LoS, but their accuracy depends on the process used to allocate costs to individual patients (Tan *et al.*, 2011). In those countries where patient-level costs are available, general guidance is set out detailing how hospitals should calculate costs, but this guidance differs in important respects across countries, as summarised in Table II. These differences render it illegitimate to pool cost data across countries. Instead, all analyses are performed separately for each country.

Our analysis involves assessing the extent to which variation in costs or LoS of patients in the same EoC is explained by the following:

- The DRGs to which the patients are allocated. As noted earlier, there is considerable variation across the 10 countries in the number and construction of DRGs used to describe patients in each EoC.
- Other variables that describe each patient including their demographic characteristics, diagnoses, treatment and quality of care. We specify a set of variables that can be constructed from routine patient-level data in all 10 countries.
- The hospital in which treatment took place. Our analytical models allow us to explore the extent to which the relative influence that each hospital has on the costs or LoS of its patients is driven by hospital-level characteristics.

¹Austria, England, Estonia, Finland, France, Germany, Ireland, Poland, Sweden and Spain (Catalonia).

Table II. Summary of approaches to costing

Country	Overhead cost allocation to medical departments	Direct cost allocation to patients	The following costs included/excluded in patient costs				
			Investments and buildings	Interest	Taxes/VAT-related costs	Research and teaching	User charges
England	Step down	Top-down microcosting to DRG level	Included	Included	Excluded	Excluded	Excluded
Estonia	Direct	Bottom-up microcosting	Excluded	Excluded	Excluded	Excluded	Excluded
Finland	Direct	Bottom-up microcosting	Included	Included	Excluded	Excluded	Excluded
France	Step down	Mixture of top-down and bottom-up microcosting	Included	Included	Included	Excluded	
Germany	Step down (preferably)	Bottom-up microcosting	Excluded	Excluded ^b	Excluded	Excluded	
Spain	Top-down with reciprocal imputation (Catalonia)	Bottom-up (activity based costing)	Included	Excluded	Included	Excluded	
Sweden	Direct	Bottom-up microcosting	Included ^a	Included ^a	Excluded	Excluded	Included

^aAll indirect costs except research, teaching, external projects, ambulances and counties' politicians and their staff are included.

^bIf not related to capital loans.

For six countries, the analytical sample for each EoC consists of all hospital patients that have one of the specific diagnostic or procedural codes as detailed in Table I recorded in their medical record. The analyses for Finland, France (for cost), Germany and Spain (Catalonia) are based on all patients admitted to a sample of hospitals. Data are for 2007/2008, with the exception of France and Poland where data relate to 2007 and 2009, respectively. All admitted patients are included whether or not they were treated on a day-case basis, as DRGs are designed to capture the full caseload of admitted patients, irrespective of their LoS (Fetter *et al.*, 1980). Indeed, in some countries, DRG payments have been designed with the expressed intention to incentivise hospitals to shift from overnight to day-case treatment (Street and Maynard, 2007). Patients attending outpatient or ambulatory clinics, whether before or after hospital treatment, are excluded as different classification systems are used to describe this activity (Berlowitz *et al.*, 1995; Goldfield *et al.*, 2008).

In what follows, we first specify in general terms the models used to estimate variation in costs or LoS for these patients. We describe the construction of the variables to be included in these models and then provide a guide to comparative interpretation of the models. We conclude with a summary of how we build on the existing literature, the analytical choices made given the constraining features of the EuroDRG project and the suggested directions for future research.

2. EMPIRICAL FRAMEWORK

2.1. Overview

An analysis of why resource use differs among patients can be undertaken by specifying a multilevel (or hierarchical) model, which recognises that patients (level 1) are clustered within hospitals (level 2). We explore variations at each level of the hierarchy by undertaking the analysis as a two-stage process (Saxonhouse, 1976; Lewis and Linzer, 2005; Jusko and Shively, 2005; Laudicella *et al.*, 2010). In the first stage, we assess the influence on costs or LoS of a set of variables defined for each individual patient. After controlling for these patient-level factors, we assess the influence that hospitals have on costs and LoS in the second stage.

2.2. Analysis of patient-level costs

When cost data are available, in the first stage, we analyse the influence of patient characteristics on cost. The model takes the following general form:

$$y_{ik}^c = \beta_0 + \boldsymbol{\beta}' \mathbf{x}_{ik} + u_k + \varepsilon_{ik}$$

where y_{ik}^c is the (logarithmic) cost of patient i in hospital k and \mathbf{x}_{ik} is a vector of characteristics of patient i in hospital k , which will be discussed in detail in Section 3.1. For characteristics that enter as dummy variables, their proportionate influence is calculated as $p = \left\{ \exp(\hat{\boldsymbol{\beta}}) - 1 \right\}$ (Halvorsen and Palmquist, 1980). u_k captures the hospital influence on costs over and above the patient characteristics, whereas ε_{ik} is the standard disturbance.

Typically, the distribution of patient costs is positively skewed. Although we exclude outliers from the analysis,² we also avoid any remaining high-cost patients exerting undue influence on the estimated relationships by converting costs into logarithmic form, which serves to normalise the overall distribution of costs.

With only two levels to the hierarchy (patients clustered in hospitals), the previously mentioned equation can be estimated using log-linear models with fixed effects, which are appropriate given our aim to make inferences about hospitals themselves (Rice and Jones, 1997).³ These fixed effects, u_k , can be interpreted as a measure of hospital performance, with higher values implying that this hospital's costs are above average after taking into account the characteristics of the patients being treated (Dormont and Milcent, 2004; Dormont and Milcent, 2005).

2.3. Analysis of patient-level length of stay

For those countries where patient-level cost data are unavailable, LoS is analysed as a proxy for resource use. As LoS is count data, the model outlined above has to be adjusted to incorporate this.

The standard model for count data analysis is Poisson regression. The underlying assumption is that equidispersion is present, that is, conditional mean equals variance, $\mu_{ik} = E[y_{ik}^s | \mathbf{x}_{ik}, h_k] = V[y_{ik}^s | \mathbf{x}_{ik}, h_k]$. The probability density function is as follows:

$$f(y_{ik}^s | \mathbf{x}_{ik}, h_k) = \frac{e^{-\mu_{ik}} \mu_{ik}^{y_{ik}^s}}{y_{ik}^s!}$$

where y_{ik}^s is the LoS (number of days) of patient i in hospital k and h_k represents the hospital effects we want to capture in the model.

For some of our EoCs, the assumption of equidispersion might be too restrictive as conditional variance exceeds the mean, a traditional way of generalization being to use negative binomial (NB) models. In the standard NB model (NB2 in Cameron and Trivedi, 1998) the variance $\omega_{ik} = V[y_{ik}^s | \mathbf{x}_{ik}, h_k]$ is assumed to be a quadratic function of the mean.

$$\omega_{ik} = \mu_{ik} + \alpha \mu_{ik}^2, \alpha > 0$$

Then, the probability density function is as follows:

$$f(y_{ik}^s | \mathbf{x}_{ik}, h_k) = \frac{\Gamma(y_{ik}^s + \alpha^{-1})}{\Gamma(y_{ik}^s + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_{ik}} \right)^{\alpha^{-1}} \left(\frac{\mu_{ik}}{\alpha^{-1} + \mu_{ik}} \right)^{y_{ik}^s}$$

²Newborns (younger than 1 year) were dropped from the sample, as were hospitals where records were available for fewer than five patients. Cost outliers are identified with a bilateral trim based on three times the standard deviation of the cost distribution. LoS distributions are highly skewed, so only the right-tailed length of stay outliers are identified on a log-transformation of LoS with an upper trim also based on three times the standard deviation threshold. Note that hospital-level variables have been calculated on the full data set (before dropping any outlier).

³Random effects can also be estimated, and for some episodes of care and for some countries, the Hausman (1978) test might indicate that these are to be preferred. If our interest was solely in costs, a random effects model would be preferable (provided \mathbf{x}_{ik} and \mathbf{u}_k are uncorrelated), as it would allow the consideration of both patient- and hospital-level characteristics in a single model. We adopted a fixed effects specification though to ensure comparability with the LoS equations and because we seek inferences about specific hospitals rather than the population from which they might be drawn.

However, NB models are required only if overdispersion is present. A formal test for overdispersion has been proposed to choose between the Poisson and the NB models (Lee, 1986; Cameron and Trivedi, 1986). Therefore, on the basis of this test, we analyse LoS with either Poisson or NB models, but consistently across countries for each EoC.⁴

The way that these count data models account for the clustering of patients within hospitals is not quite equivalent to the process applied in the cost model. This is because the count data fixed effects differ in theory (Hausman *et al.*, 1984; Cameron and Trivedi, 1998) and in implementation (see *xtpoisson* and *xtbreg* routines in StataCorp., 2009) from log-linear models with fixed effects. Instead, we apply unconditional Poisson/NB regression estimators with h_k representing the deviation of hospital k from the grand mean, each of which captures the hospital's effect on LoS. This gives us:

$$\mu_{ik} = E(y_{ik}^s | x_{ik}, h_k) = \exp(\beta_0 + \boldsymbol{\beta}' x_{ik} + u_k h_k)$$

Here, u_k is the estimated hospital effect, which is analogous to the fixed effect in the cost models. It has been demonstrated that the direct estimation of the fixed effects in count data models can be achieved by introducing a dummy variable for each hospital (Cameron and Trivedi, 1998; Allison and Waterman, 2002). u_k can then be interpreted in the same way as for the cost model.

2.4. Analysis of hospital-level characteristics

In the second stage, the estimated hospital effects, \hat{u}_k , are analysed to explore reasons why some hospitals have higher average costs or lengths of stay than others. We shall consider what these characteristics are in Section 3.2, but for the moment, we summarise them as a vector \mathbf{z}_k of variables measuring hospital-level characteristics in a regression of the form:

$$\hat{u}_k = \gamma_0 + \boldsymbol{\gamma}' \mathbf{z}_k + \varepsilon_k$$

Estimated dependent variables suffer both sampling and random error (Saxonhouse, 1976; Lewis and Linzer, 2005). Here, the sampling error is the difference between the true value of the hospital effect and its estimated value from the first stage. Random error arises even if the second-stage variable was observed directly rather than being estimated. Consequently, we need to account for both sources of error, a task further complicated when the size of measurement errors is unknown. As we want observations with smaller variance to carry larger weight in the regression, we compute a GLS regression with weights proportional to the inverse of the squared standard errors⁵ and Efron robust standard errors to correct for potential heteroscedasticity. These are more accurate than standard Huber–White corrections when dealing with small samples (Davidson and MacKinnon, 1983; Lewis and Linzer, 2005).

2.5. Graphical analyses

We interpret the hospital effects derived from the cost and LoS equations as measures of relative hospital performance (Laudicella *et al.*, 2010). These capture the average cost or LoS of the hospital's patients, purged of the influence of the explanatory variables included in the first-stage equations, which are assumed to be uncorrelated with the hospital effects (Dormont and Milcent, 2005).

A visual comparison of the effects across hospitals requires standardisation of the graphical representation. For the effects derived from the cost analyses, we plot the hospital intercept \hat{u}_k standardized by the sample average to derive the relative performance of each hospital in comparison with the mean. Consequently, having conditioned on the variables accounted for in the first stage, each ratio corresponds to the mean cost for the

⁴This test usually gave consistent results across countries about whether it is appropriate to run a Poisson or an NB model, providing assurance about the appropriateness of the chosen specification for each episode of care.

⁵See the Stata Manual (StataCorp, 2009) and Lewis and Linzer (2005) for a more detailed discussion about the different options.

hospital in question relative to the overall mean for all hospitals in the country. For LoS, we adapt this method to our unconditional Poisson/NB regression with predicted fixed effects. Each hospital's ratio corresponds to its relative LoS compared with the national average.

The 95% confidence intervals of the hospital effects from the cost and LoS analyses assume a normal distribution of these fixed effects. Some hospitals with very wide confidence intervals are omitted from the graphs. These were omitted if, for the EoC in question, they either had very low volumes of activity or a handful of patients with markedly different costs or LoS to most other patients in the same hospital. In both cases, the hospital's effect will be estimated imprecisely.

The interpretation of the ratios and graphs is straightforward. A value of 1.3 means that patients in this hospital have 30% higher costs or LoS compared with the average for all hospitals in the country that deliver this type of care. Hospitals are ordered (from left to right) from those with the lowest mean effects to the highest. Patients in hospitals to the left have lower cost or LoS, on average, than those in other hospitals (this not being due to the explanatory variables accounted for in the first stage of the analysis). We interpret the position of each hospital relative to its national peers as a measure of its relative efficiency in managing resources.

3. EXPLANATORY FACTORS

3.1. Patient-level variables

The data used to define the variables included in the models are drawn from routine sources in each country, as summarised in Table III. Our vector, \mathbf{x}_{ik} , comprises two groups of explanatory factors, namely, those that capture the DRG to which each patient is allocated \mathbf{x}_{ik}^d and a set of patient-level characteristics \mathbf{x}_{ik}^p , including demographic, diagnostic, treatment and quality information, such that

$$\mathbf{x}_{ik} = (\mathbf{x}_{ik}^d, \mathbf{x}_{ik}^p)$$

Our initial specification (M_D) explores the extent to which DRGs explain variation in costs or LoS of patients having the same EoC:

$$M_D : \begin{cases} y_{ik}^c = \beta_0 + \boldsymbol{\beta}^{d'} \mathbf{x}_{ik}^d + u_k + \varepsilon_{ik} \\ \mu_{ik} = \exp(\beta_0 + \boldsymbol{\beta}^{d'} \mathbf{x}_{ik}^d + u_k h_k) \end{cases}$$

Table III. Databases by country

Country	Data sources	Year
Austria	Performance-oriented Hospital Financing Framework Database	2008
	Private Hospitals Financing Fund Database	2008
England	Hospital Episode Statistics	2007/2008
	National Health Service Reference Costs	2007/2008
Estonia	Estonian Health Insurance Fund database	2008
Finland	Hospital Discharge Register, hospitals of Helsinki and Uusimaa	2008
France	National hospital cost study (ENCC; representative sample of voluntary hospitals)	2007
Germany	Hospital inpatient activity database (PMSI MCO; exhaustive sample of all hospitals)	2008
	Research database based on patient-level data according to §21 Hospital Remuneration Act (KHEntG)	2008
	National G-DRG cost accounting standards by the Institute for the Hospital Remuneration System (InEK)	2008
Ireland	Hospital In-Patient Enquiry	2008
Poland	Central register of healthcare services and reimbursements	2009
Spain (Catalonia)	Public Hospital Network of Catalonia	2008
	Spanish Network of Hospital Costs	2008/2009
Sweden	National Case Costing Database	2008

Each patient is allocated to a single DRG, and a dummy variable is used to identify which DRG. The number of DRGs to which patients are allocated varies according to the EoC and country in question. For the 10 countries in the study, on average, 6 DRGs are used to describe the vast majority of patients having a knee replacement whereas up to 16 DRGs are used to describe patients admitted for acute myocardial infarction (AMI). The number of DRGs used to describe patients in the same EoC also varies markedly across countries because DRG classifications differ (Kobel *et al.*, 2011). Consequently, the set of DRG dummy variables varies across both EoCs and countries.

For comparative purposes, we adopted the following procedure in constructing and ordering each country's DRG dummy variables. First, the reference DRG (which is omitted from the model) is that which captures most of the country's patients in that EoC. Second, we include a dummy variable for any other DRG if it accounts for at least 1% of patients in that EoC. In most cases, these are reported in an ascending order of their associated reimbursement rate. Third, a dummy variable captures all other patients in the EoC that are not assigned to the identified DRGs.

Our second specification (M_P) considers a set of explanatory characteristics other than DRGs to explain variation in resource use among patients. Thus, we consider the vector \mathbf{x}_{ik}^p in the following specification:

$$M_P : \begin{cases} y_{ik}^c = \beta_0 + \boldsymbol{\beta}^p \mathbf{x}_{ik}^p + u_k + \varepsilon_{ik} \\ \mu_{ik} = \exp(\beta_0 + \boldsymbol{\beta}^p \mathbf{x}_{ik}^p + u_k h_k) \end{cases}$$

These variables describe patients according to their demographic characteristics, diagnoses, treatment and quality of care.

We construct age categories based on quintiles chosen according to the observed distribution of age for the EoC in question, with the second age category generally being the reference group. A dummy variable identifies whether the patient was male.

We consider the impact on resource use of the number of diagnoses and procedures performed, whether the patient was transferred from or to another institution as part of their care pathway, whether they were admitted through the emergency department and whether the patient died in a hospital. We also specify various variables that capture common diagnostic characteristics and procedural techniques that are specific to the EoC in question. Details of these variables are provided in the empirical articles.

We take account of the comorbidities used in the construction of the Charlson index (Charlson *et al.*, 1987; Quan *et al.*, 2005). Rather than using the index itself, we define three distinct patient groups based on their Charlson comorbidities. The first involves specifying six of the 17 Charlson comorbidities as 'severe', these being hemiplegia/paraplegia, renal disease, cancer, moderate or severe liver disease, metastatic solid tumour and AIDS/HIV (Charlson *et al.*, 1987). The other comorbidities are designated 'non-severe'. We then define a dummy variable indicating whether the patients suffered a single non-severe comorbidity and another dummy variable indicating at least one severe or two non-severe comorbidities; all other patients suffered no comorbidity (and form the reference group).

We modify this process if the diagnoses included in the Charlson categories are directly related to the EoC analysed (in which case they are not comorbidities). The disregarded comorbidities are myocardial infarction for the analysis of AMI and coronary-artery bypass graft (CABG), cancer and metastatic solid tumour for breast cancer and cerebrovascular disease and hemiplegia/paraplegia for stroke.

We also include a set of variables that capture adverse events (Loke *et al.*, 2008). We use some of the patient safety indicators (PSIs) that have been developed for use with routine hospital administrative data (Quan *et al.*, 2008). A dummy variable indicates whether any one of the following occurred: PSI5 (foreign body left in during procedure), PSI7 (infection and inflammatory reaction due to other vascular device, implant, etc.), PSI12 (pulmonary embolism/deep vein thrombosis), PSI13 (sepsis) and PSI15 (accidental cut, puncture, perforation or haemorrhage during medical care). For childbirth, we take account of PSI18, PSI19 and PSI20 (obstetric trauma) rather than the other indicators.

Finally, we define two dummy variables measuring whether urinary tract infection (International Classification of Diseases, 10th Revision (*ICD-10*) codes N30.x, N39.0, O23.x and O86.2) or postoperative surgical infection (T81.4) are suffered during hospitalisation.

The third model is our ‘full’ specification (M_F) that includes both the vectors of DRG variables and patient characteristics:

$$M_F : \begin{cases} y_{ik}^c = \beta_0 + \boldsymbol{\beta}^d \mathbf{x}_{ik}^d + \boldsymbol{\beta}^p \mathbf{x}_{ik}^p + u_k + \varepsilon_{ik} \\ \mu_{ik} = \exp(\beta_0 + \boldsymbol{\beta}^d \mathbf{x}_{ik}^d + \boldsymbol{\beta}^p \mathbf{x}_{ik}^p + u_k h_k) \end{cases}$$

The hospital effects from the full model are those used for the graphical presentation and for the second-stage analysis, designed to identify hospital characteristics that may drive average costs or LoS. We now turn to the discussion of these characteristics.

3.2. Hospital-level variables

Our second-stage analysis is designed to explore the explanatory power of various hospital characteristics, incorporated in the vector \mathbf{z}_k , on costs or LoS. This analysis is conducted only for countries that have a sufficient number of hospitals for the EoC to be included in the regression; as a rule of thumb, Montenegro (2001) suggests a minimum of 10 observations per parameter.

The characteristics considered include the hospital’s teaching or ownership status, the amount and range of activity undertaken and the quality of care. We discuss the reasons why these characteristics may be related to resource use in detail elsewhere (Street *et al.*, 2010) and offer a brief summary in Table IV.

The observed relationships between these characteristics and our second-stage dependent variables are summarised in each of the empirical articles that follow. Tables of results are not reported but available from www.eurodrgeu.

4. MODEL INTERPRETATIONS AND COMPARISONS

In the empirical analyses for each EoC, we estimate and compare our models to gain insight into three broad sets of issues.

First, we identify what factors explain variation in cost and LoS and assess whether these explanatory factors are generally consistent across cost and LoS equations. As is usual practice, this involves the consideration

Table IV. Hospital-level variables

Variable	Description/rationale for consideration
Teaching status	The funding of teaching and research may not accurately reflect the full cost of these activities, opening up the possibility of cross-subsidisation with patient care. Teaching hospitals might systematically attract more severe patients within each DRG.
Ownership status	Public and private or for-profit and not-for-profit status may confer different incentives on staff and may also mean that hospitals are subject to different regulatory constraints (Duckett, 2001; Mason <i>et al.</i> , 2009).
Volume of activity	Hospitals might benefit from economies of scale, experiencing decreasing average costs as volume increases. Beyond a particular size, diseconomies of scale may arise. Measured by (i) the size of the hospital (number of patients treated annually, in thousands) and (ii) the share of the hospital’s patients that fall under the EoC in question (in percent).
Specialization	If there are economies of scope, increasing the range of activity leads to lower costs (Panzar and Willig, 1981). However, hospitals that specialise may have lower costs than general hospitals (Dranove, 1987). We enhance the measure of concentration of activity across Major Diagnostic Categories (MDCs) proposed by Daidone and D’Amico (2009), measuring specialisation as the deviation from the national average hospital morbidity across MDCs.
Adverse events	We consider the rate per 1,000 of infections (rate 1) and of PSIs for all patients (rate 2) of the hospital. For childbirth, we consider obstetric trauma instead.

of each of the variables included in the vector \mathbf{x}_{ik} and of overall explanatory power as given by the adjusted R^2 statistics for the cost equations. For the LoS equations, we use the deviance R^2 proposed by Cameron and Windmeijer (1996) with adjustment for the number of parameters. The deviance R^2 represents the sum of squared deviance residuals, which is a generalization of one of the many ways to express the normal R^2 statistics (Kvålseth, 1985). In addition, it can be expressed as the relative log-likelihood increase compared with the maximal possible log-likelihood increase of a saturated model (Cameron and Windmeijer, 1996).⁶

Second, we assess the explanatory power of the set of DRGs to which each patient is allocated, \mathbf{x}_{ik}^d , relative to the set of characteristics included in the vector \mathbf{x}_{ik}^p and in relation to the fully specified model. We evaluate the explanatory power of model M_D relative to model M_F by measuring $a = (R_{M_F}^2 - R_{M_D}^2) / R_{M_F}^2$. Similarly, the explanatory power of model M_P relative to model M_F is defined as $b = (R_{M_F}^2 - R_{M_P}^2) / R_{M_F}^2$.

Using these measures a and b , we categorise countries according to the explanatory power of their DRG system \mathbf{x}_{ik}^d relative to the set of patient characteristics included in the vector \mathbf{x}_{ik}^p . The assessment and the interpretation depend on which category a country appears in. Generally, we distinguish between several categories.

- Both a and b have relatively small values. In this case, we could conclude that \mathbf{x}_{ik}^d and \mathbf{x}_{ik}^p are equally successful at explaining variation, appear to play a similar function and can be considered interchangeable.
- Both a and b have relatively large values. Hence, \mathbf{x}_{ik}^d and \mathbf{x}_{ik}^p are only able to capture a small proportion of what the full model M_F is able to explain. Possibly, the DRG system could be refined to take into account \mathbf{x}_{ik}^p .
- If a is larger than b , \mathbf{x}_{ik}^p is better than \mathbf{x}_{ik}^d at explaining variation. This suggests that the country's DRG system lacks specific criteria that could be used for grouping purposes. The patient characteristics included in \mathbf{x}_{ik}^p include these criteria.
- If a is smaller than b , we could conclude that \mathbf{x}_{ik}^d performs better than \mathbf{x}_{ik}^p . This might be because the local DRG system, using country-specific information (such as operation codes), has valuable explanatory power.

Finally, we assess the hospital effects derived from model M_F . These effects capture the average cost or LoS of patients in each hospital, purged of the influence on resource use of the explanatory variables in \mathbf{x}_{ik} . As such, the effects can be considered measures of relative hospital performance, the relative position of each hospital not being due to differences in the characteristics of their patients or to the DRGs to which patients are allocated, which have been accounted for in the first-stage analysis. There may be other identifiable characteristics of hospitals that influence their relative performance. We are able to evaluate these characteristics statistically in those countries with a sufficient number of hospitals.

5. DISCUSSION AND CONCLUSIONS

This article builds on the econometric literature that analyses reasons why costs vary among hospital patients. We contribute to this literature in four important respects.

First, we set out methods for analysing variations in LoS that are equivalent to those used in analysing costs. There are advantages to analysing 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. LoS is, of course, only a partial measure of

⁶In addition to the deviance R^2 , we have evaluated the model performance of LoS models using other measures of fit (i.e. AIC, BIC, pseudo- R^2 and correlation R^2 ; Cameron and Windmeijer, 1996) to ensure that the conclusions are not conditional upon the measure used.

resource use, and our preference is for the analysis of cost data where these are available. When both measures are available, it is straightforward to run both analyses and compare the results. We have reserved these comparisons for country-level reports (Gaughan *et al.*, 2012).

Second, we have set out an econometric approach to evaluating the performance of DRG systems in explaining variations in resource use among patients undergoing the same EoC. Our approach is intended to complement rather than substitute for traditional means of constructing and evaluating DRGs. Typically, these involve either (i) constructing DRGs from scratch by assessing potential grouping variables often using the Classification and Regression Tree (CART) analysis and measuring reduction in variance (Lemon *et al.*, 2003) or (ii) comparing existing grouping systems by running the same set of patient data through them (Reid *et al.*, 2000; Reid and Sutch, 2008). Before the Classification and Regression Tree, it may be useful to perform econometric analysis, which can indicate the size of influence on costs or LoS of each explanatory variable and therefore provide a good starting point for identifying candidate grouping variables.

Third, unlike those studies of hospital efficiency that rely on aggregate or average hospital costs (Hollingsworth, 2008), analysis based on patient-level data is both more robust and insightful, in which hospital effects are purged of the influence of patient characteristics (Laudicella *et al.*, 2010). Although early studies based on patient-level data explored the average impact of hospital characteristics on costs (McClellan, 1997; Dormont and Milcent, 2004), here we also focus on the purged effects in their own right, interpreting them as measures of relative hospital efficiency in managing resource use. Although this has been performed previously in analysing costs (Laudicella *et al.*, 2010; Dormont and Milcent, 2005), we extend the approach to the analysis of LoS.

Fourth, we have outlined an analytical strategy for making cross-country comparisons that does not require the pooling of data or the use of purchasing power parities to rebase costs. A key feature of the empirical analyses is that they are based on routinely available data. We have constructed a set of explanatory variables that are defined consistently across all 10 countries. However, there may be systematic differences across countries in how costs are calculated and in coding practice. For example, in Finland, it is uncommon to code beyond the primary diagnosis, in contrast to greater coding depth elsewhere, for example, Germany. Differential practice prohibits the pooling of data and also means that caution needs to be exercised in making comparisons across countries. Such cautions appear in the accompanying empirical articles.

Various methodological choices reflect compromises made to ensure consistency across EoCs and countries, given the objectives of the EuroDRG project. Consistency was sought particularly in three areas: (i) to ensure comparability between those countries with cost data and those with LoS data, (ii) to ensure a common cost or LoS functional form across all countries for each EoC and (iii) to ensure a common set of explanatory variables that could be constructed from routine patient-level data in all countries and across all EoCs. In research contexts that are not subject to such overarching constraints, alternative choices might be made.

For instance, we adopted a fixed effects specification to ensure the comparability of the cost and LoS analyses. This decision was somewhat costly, requiring us to undertake a second-stage analysis to assess whether the fixed effects are driven by provider characteristics. If cost is the only variable of interest, a random effects model might be preferable, allowing the consideration of provider characteristics in a single specification (Dormont and Milcent, 2004). That said, when considerable information is available about the first-level effects, the gains from estimating both levels in a single step are modest (Lewis and Linzer, 2005).

Similarly, we choose to model log-transformed costs because this allows an easy comparison of the proportionate influence on the costs of explanatory variables across all countries, irrespective of their currencies. If not making cross-country comparisons and subject to their distribution, the analysis of costs in linear or gamma forms may in some situations be preferable (Manning and Mullahy, 2001). Note though that the empirical gains may not be substantial for either cost or LoS analyses; for instance, Austin *et al.* (2002, 2003) compared several regression models and found they have similar ability to predict costs or LoS after CABG surgery.

Finally, the full specification, M_F used to estimate hospital effects comprises both the vectors of DRGs, x_{ik}^d and patient-level characteristics, x_{ik}^p . There may be multicollinearity between variables in these vectors, particularly as DRGs can be a function of patient characteristics. In the empirical analyses, variables that were highly collinear ($r > 0.8$) with DRGs variables were excluded. However, in any case, including irrelevant variables in the model does not bias the estimates of the hospital effects (Wooldridge, 2002).

In summary, we have set out an analytical strategy to examine variations in resource use, whether cost or LoS, of patients hospitalised with different conditions. This allows us (i) to identify those factors that explain variation in resource use across patients, (ii) to assess the explanatory power of DRGs relative to other patient and treatment characteristics and (iii) to assess relative hospital performance in managing resources and the characteristics of hospitals that explain this performance. Although the approach has been tailored to the requirements of the EuroDRG project to support the empirical articles that appear in this special issue, the methods should be of general interest to researchers evaluating variation in resource use among patients and across healthcare providers and countries.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The authors thank all members of the EuroDRG project team and, particularly, Reinhard Busse, Nils Gutacker, Unto Häkkinen, Jacqueline O'Reilly, Gunnar Rosenqvist and Susanne Strohmaier for their advice and guidance. They are also grateful to the journal's referees for their constructive comments. The usual caveat applies.

REFERENCES

- Allison PD, Waterman RP. 2002. Fixed-effects negative binomial regression models. *Sociological Methodology* **32**: 247–265.
- Austin PC, Rothwell DM, Tu JV. 2002. A Comparison of Statistical Modeling Strategies for Analyzing Length of Stay after CABG Surgery. *Health Services and Outcomes Research Methodology* **3**: 107–133.
- Austin PC, Ghali WA, Tu JV. 2003. A comparison of several regression models for analysing cost of CABG surgery. *Statistics in Medicine* **22**: 2799–2815.
- Berlowitz D, Rosen A, Moskowitz M. 1995. Ambulatory care casemix measures. *Journal of General Internal Medicine* **10**: 162–170.
- Bradford WD, Kleit AN, Krousel-Wood MA, Re RN. 2001. Stochastic Frontier Estimation of Cost Models Within the Hospital. *The Review of Economics and Statistics* **83**: 302–309.
- Cameron AC, Trivedi PK. 1986. Econometric models based on count data. Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics* **1**: 29–53.
- Cameron AC, Trivedi PK. 1998. *Regression Analysis of Count Data*. Cambridge University Press: Cambridge.
- Cameron AC, Windmeijer F. 1996. R-Squared Measures for Count Data Regression Models with Applications to Health-Care Utilization. *Journal of Business & Economic Statistics* **14**: 209–220.
- Charlson M, Pompei P, Ales KL, Mackenzie CR. 1987. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases* **40**: 373–383.
- Daidone S, D'Amico F. 2009. Technical efficiency, specialization and ownership form: evidences from a pooling of Italian hospitals. *Journal of Productivity Analysis* **32**: 203–216.
- Davidson R, Mackinnon JG. 1983. *Estimation and Inference in Econometrics*. Oxford University Press: New York.
- Dormont B, Milcent C. 2004. The sources of hospital cost variability. *Health Economics* **13**: 927–939.
- Dormont B, Milcent C. 2005. How to regulate heterogeneous hospitals? *Journal of Economics and Management Strategy* **14**: 591–621.

- Dranove D. 1987. Rate-setting by diagnosis related groups and hospital specialization. *RAND Journal of Economics* **18**: 417–427.
- Duckett S. 2001. Does it matter who owns health facilities? *Journal of Health Services Research & Policy* **6**: 59–62.
- Fetter RB, Shin YB, Freeman JL, Averill RF, Thompson JD. 1980. Casemix definition by diagnosis related groups. *Medical Care* **18**(suppl): 1–53.
- Gaughan J, Mason A, Street A, Ward P. 2012. *English hospitals can improve their use of resources: an analysis of costs and length of stay for ten treatments*. Centre for Health Economics research paper 78: York.
- Gilman BH. 2000. Hospital response to DRG refinements: the impact of multiple reimbursement incentives on inpatient length of stay. *Health Economics* **9**: 277–294.
- Goldfield N, Averill R, Eisenhandler J, Grant T. 2008. Ambulatory Patient Groups, Version 3.0-A Classification System for Payment of Ambulatory Visits. *The Journal of Ambulatory Care Management* **31**: 2–16.
- Halvorsen R, Palmquist R. 1980. The interpretation of dummy variables in semilogarithmic equations. *American Economic Review* **70**: 474–75.
- Hausman JA, Hall BH, Griliches Z. 1984. Econometric models for count data with an application to the patents-R&D relationship. *Econometrica* **52**: 909–938.
- Hollingsworth B. 2008. The measurement of efficiency and productivity of health care delivery. *Health Economics* **17**: 1107–1128.
- Jusko KL, Shively WP. 2005. Applying a two-step strategy to the analysis of cross-national public opinion data. *Political Analysis* **13**: 327–344.
- Kessler D, McClellan M. 2002. The effects of hospital ownership on medical productivity. *The Rand Journal of Economics* **33**: 488–506.
- Kobel C, Thuilliez J, Bellanger M, Pfeiffer K-P. 2011. DRG systems and similar patient classification systems in Europe. In *Diagnosis Related Groups in Europe: Moving towards transparency, efficiency and quality in hospitals*, Busse R, Geissler A, Quentin W, Wiley M (eds). Open University Press: Maidenhead.
- Kvålseth TO. 1985. Cautionary Note about R2. *The American Statistician* **39**: 279–285.
- Laudicella M, Olsen KR, Street A. 2010. Examining cost variation across hospital departments—a two-stage multilevel approach using patient level data. *Social Science & Medicine* **71**: 1872–1881.
- Lee L-F. 1986. Specification test for poisson regression models. *International Economic Review* **27**: 689–706.
- Lemon S, Roy J, Clark M, Friedmann P, Rakowski W. 2003. Classification and regression tree analysis in public health: Methodological review and comparison with logistic regression. *Annals of Behavioral Medicine* **26**: 172–181.
- Lewis JB, Linzer DA. 2005. Estimating regression models in which the dependent variable is based on estimates. *Political Analysis* **13**: 345–364.
- Loke YK, Price D, Herxheimer A. on behalf of the Cochrane Adverse Effects Methods Group. 2008. Chapter 14: Adverse effects. In *Cochrane Handbook for Systematic Reviews of Interventions*, Higgins JPT, Green S (eds). under “The Cochrane Book Series” Imprint. John Wiley & Sons, Ltd: Chichester, West Sussex.
- Manning WG, Mullahy J. 2001. Estimating log models: to transform or not to transform? *Journal of Health Economics* **20**: 461–494.
- Martin S, Smith PC. 1996. Explaining variations in inpatient length of stay in the National Health Service. *Journal of Health Economics* **15**: 279–304.
- Mason A, Street A, Miraldo M, Siciliani L. 2009. Should prospective payments be differentiated for public and private healthcare providers? *Health Economics, Policy, and Law* **4**: 383–403.
- McClellan M. 1997. Hospital reimbursement incentives: an empirical analysis. *Journal of Economics and Management Strategy* **6**: 91–128.
- Montenegro A. 2001. On sample size and precision in ordinary least squares. *Journal of Applied Statistics* **28**: 603–605.
- Norton EC, Van Houtven CH, Lindrooth RC, Normand ST, Dickey B. 2002. Does prospective payment reduce inpatient length of stay? *Health Economics* **11**: 377–387.
- Olsen KR, Street A. 2008. The analysis of efficiency among a small number of organisations: how inferences can be improved by exploiting patient-level data. *Health Economics* **17**: 671–281.
- Panzar J, Willig R. 1981. Economies of scope. *American Economic Review* **71**: 268–272.
- Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J-C, Saunders LD, Beck CA, Feasby TE, Ghali WA. 2005. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical Care* **43**: 1130–1139.
- Quan H, Drösler S, Sundararajan V, Wen E, Burnand B, Couris CM, Halfon P, Januel JM, Kelley E, Klazinga N, Luthi JC, Moskal L, Pradat E, Romano PS, Shephard J, So L, Sundaresan L, Tournay-Lewis L, Trombert-Pavioit B, Webster G, Ghali WA. 2008. Adaptation of AHRQ Patient Safety Indicators for Use in ICD-10. Administrative Data by an International Consortium. In *Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 1: Assessment)*, Henriksen K, Battles JB, Keyes MA, Grady ML (eds). Agency for Healthcare Research and Quality: Rockville (MD).
- Reid B, Sutch S. 2008. Comparing diagnosis-related group systems to identify design improvements. *Health Policy* **87**: 82–91.

- Reid B, Palmer G, Aisbett C. 2000. The performance of Australian DRGs. *Australian Health Review* **23**: 20–31.
- Rice N, Jones A. 1997. Multilevel models and health economics. *Health Economics* **6**: 561–75.
- Saxonhouse GR. 1976. Estimated parameters as dependent variables. *American Economic Review* **66**: 178–183.
- Street A, Maynard A. 2007. Activity based financing in England: the need for continual refinement of payment by results. *Health Economics, Policy, and Law* **2**: 419–427.
- Street A, Scheller-Kreinsen D, Geissler A, Busse R. 2010. *Determinants of hospital costs and performance variation: Methods, models and variables for the EuroDRG project*. Universitätsverlag der TU Berlin: Berlin.
- Tan SS, Serdén L, Geissler A, Van Ineveld M, Redekop K, Heurgren M, Hakkaart-Van Roijen L. 2011. DRGs and cost accounting: which is driving which? In *Diagnosis-related Groups in Europe: Moving towards transparency, efficiency and quality in hospitals*, Busse R, Geissler A, Quentin W, Wiley M (eds). Open University Press: Maidenhead.
- Wooldridge JM. 2002. *Econometric analysis of cross section and panel data*. The MIT Press: Cambridge.