Measuring cost efficiency in the Nordic Hospitals— a cross-sectional comparison of public hospitals in 2002

Miika Linna · Unto Häkkinen · Mikko Peltola · Jon Magnussen · Kjartan S. Anthun ·
Sverre Kittelsen · Annette Roed · Kim Olsen · Emma Medin · Clas Rehnberg

Received: 23 February 2010 / Accepted: 6 July 2010 / Published online: 3 August 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract The aim of this study was to compare the performance of hospital care in four Nordic countries: Norway, Finland, Sweden and Denmark. Using national discharge registries and cost data from hospitals, cost efficiency in the production of somatic hospital care was calculated for public hospitals. Data were collected using harmonised definitions of inputs and outputs for 184 hospitals and data envelopment analysis was used to calculate Farrell efficiency estimates for the year 2002. Results suggest that there were marked differences in the average hospital efficiency between Nordic countries. In 2002, average efficiency was markedly higher in

Finland compared to Norway and Sweden. This study found differences in cost efficiency that cannot be explained by input prices or differences in coding practices. More analysis is needed to reveal the causes of large efficiency disparities between Nordic hospitals.

Keywords Hospitals · Organisational efficiency · DEA · DRG · International comparisons

1 Introduction

In recent years there have been growing efforts and interest in comparing the relative performance of various health systems [1]. Many OECD countries are developing national performance measurement frameworks for monitoring and comparing the overall efficiency of their health care systems. These frameworks cover important dimensions of health care provision and usually include a selection of various indicators for health outcomes, distribution, productivity, and patient satisfaction. In this setting, output measurement based on various activities may provide a useful means to assess and compare the technical aspects of hospital production (productivity and technical efficiency measurement based on intermediate types of output such as patients grouped by DRGs). However, this approach still faces many difficulties. It is not a trivial task to make output measures comparable since the output grouping definitions, coding classifications and coding practices usually vary considerably across countries. Recently, the OECD has noted the need for a generic patient grouping system (e.g. Diagnosis Related Groups, DRG) that can take into account the varying patient classifications across the OECD countries [2]. The report by the OECD recognises the importance and need for a general mapping of various procedure coding systems. However, the use of different

M. Linna (✉) · U. Häkkinen · M. Peltola
THL, Center for Health and Social Economics,
Helsinki, Finland
e-mail: mika.linna@thl.fi

J. Magnussen
Norwegian University of Technology and Science,
Oslo, Norway

K. S. Anthun
SINTEF Health Research,
Oslo, Norway

S. Kittelsen
Ragnar Frisch Centre for Economic Research,
Oslo, Norway

A. Roed
Danish National Board of Health,
Copenhagen, Denmark

K. Olsen
Danish Institute for Health Services Research,
Copenhagen, Denmark

E. Medin · C. Rehnberg
Karolinska Institutet,
Stockholm, Sweden

procedure classifications is not the only obstacle in the harmonisation of DRGs: the local applications of standardised disease classifications, such as the ICD-10 are also known to vary across European countries [3].

Due to the difficulties in the measurement of output and case-mix, international comparisons of hospital efficiency are relatively scarce in the literature. Hansen and Zwanziger [4] used cost functions to compare marginal costs in general acute care among US and Canadian hospitals. Mobley and Magnussen [5] and Magnussen and Mobley [6] examined the relative performance of Norwegian (regulated, public) and Californian (unregulated, competitive) hospitals using DEA and empirical data from 1997. Steinman et al. [7] compared a sample of Swiss and German hospitals and Linna et al. [8] compared hospital efficiency between Norway and Finland using data envelopment analysis. All these studies have found considerable differences in the average efficiency across the countries, but did not report explicitly on how the output measurement and case-mix adjustments were accomplished. It is obvious that several compromises in the accuracy of the output measurement had to be made in most of the previous comparative studies.

In this study we explored the possibility of pushing the limits of an international comparison of hospital efficiency further by using patient-level data from several countries. One of the main objectives in this study was to investigate the output grouping definitions with the primary data and how to improve validity in the total output measure when employed in international comparisons. Typical problems and differences in the output definitions which may affect DRG-based efficiency comparisons are reported in this article.

The study focused on hospital care since all Nordic countries have fairly good administrative and register data in hospitals (hospital discharge data) and even the possibility of linking registers using the personal ID number. The study was done in four Nordic countries (Denmark, Norway, Sweden and Finland) in a setting where the structure of organising hospital care and the available data (e.g. coding and used primary classifications) are sufficiently similar. In addition, each of the Nordic countries applied similar DRG grouping systems based on a common Nordic NordDRG grouping system.

2 Hospital financing and organisation in the Nordic countries

2.1 Hospital financing

In the Nordic countries, health care has been decentralised to local or regional authorities who usually purchase and also provide the health services. In all Nordic countries the funding of hospital care is a mixture of global budgeting and activity-based funding (ABF). Activity-based funding is a payment model based on the volume and type of

services provided to each patient for hospital care whereas in global budgeting the volume and total price is fixed prospectively. In Denmark, activity-based financing was not yet widely implemented during the study period. In addition, in Finland the implemented activity-based financing did not have the same incentive effects as in Norway and Sweden which applied DRG based funding to a greater extent. However, Norway, Sweden and Finland could be classified as using ABF and Denmark global budgeting.

Since health and hospital care is mainly publicly provided in the Nordic countries, taxation is the main source for financing hospital care. In 2002, the total expenditure on health as a percentage of GDP was 9.8 in Norway, 9.3 in Sweden, 8.8 in Denmark and 7.8 in Finland. The hospital care providers were regional health enterprises in Norway (after 2002), county councils in Denmark and Sweden, and hospital districts in Finland.

While municipalities and counties for the most part were responsible for arranging all health care, these responsibilities were divided differently in each country. The responsibility for hospital care in Norway, Denmark and Sweden was given to regional authorities but in Finland to local authorities. In Finland, hospital care is financed mostly by municipalities, which get their income from local taxes and non-earmarked state subsidies. In Sweden and Denmark hospital care is financed by county council taxes (with additional general grants) while in Norway there have been two public financiers (Counties and the state). After 2002 the state has been the only financier.

In Denmark, the overall budget was negotiated each year in a Budget Cooperation between state government and the local governments, who were represented by the County Council Association and the Association of Municipalities. In this co-operation an overall ceiling of the growth in the local tax rate was agreed upon and the level of state block-grants was negotiated. In Sweden the central government allocated financial assistance and acted as supervisor of activities in the county councils. In Norway after 2002, the hospital care was financed directly from the state budget as the state now owned the public hospitals, reflecting the lowest possible decentralisation level.

2.2 Hospital organisation

The public hospitals in all of the Nordic countries were responsible for producing the majority of secondary and tertiary level services in health care. In each of the Nordic countries one could, during the study period, still observe a rather similar structure in hospital organization where hospitals could be classified roughly into three: University teaching hospitals, central hospitals and local hospitals (as an exception there does exist a few specialized hospitals e.g. for orthopaedics and ophthalmology).

University teaching hospitals provided the most specialized tertiary level services and were usually organized to serve regions having populations between 0.5 and 1.0 million. University teaching hospitals were also the main sources of teaching and research output by providing medical education, and conducting clinical research.

Central hospitals provided services to a region/county/district and provided somewhat less specialized tertiary services compared to the university teaching hospitals. Typical central hospitals included several (8–15) specialty clinics and emergency departments.

Local hospitals were usually smaller hospitals supplying secondary level services to a group of municipalities within a county/district. A typical local hospital included inpatient wards for 3–5 specialties (e.g. internal medicine, general surgery, obstetrics/gynaecology) and some additional specialties on an outpatient basis.

In Sweden the 21 county councils (including two regions and one municipality responsible for health care delivery) were divided into six regions having eight university and region hospitals providing tertiary level services. The total number of hospitals in this study was 49, of which seven were university hospitals. In Sweden there were 69 hospitals in total, but 17 hospitals were included in the final accounts of another hospital and did not have the decision making power (and hence could not be defined as individual decision making units). In addition three very small hospitals Mora, Skellefteå and Lycksele, were excluded. However, the excluded Swedish hospitals represented sufficiently closely the overall distribution of hospital characteristics in Sweden, and the rejection did not create any significant bias to our comparison.

In Finland, the 20 hospital districts owned and ran more than 40 acute care public general hospitals, which were divided into three categories: university teaching hospitals (five hospitals), central hospitals (15 hospitals) and other municipal hospitals. Each hospital district hosted one central hospital which in five hospital districts were university teaching hospitals and in 15 hospital districts (non-university) central hospitals.

After the 2002 reform Norwegian hospital care was provided by public enterprise and financed by the central government. In Norway, the Health Enterprise Act 2002 changed crucially the hospital care system. Hospital ownership was transferred from counties to the central government, and currently there are five Regional Health Enterprises (RHEs) that are reporting to the Ministry of Health and are responsible for delivering health services in their regions. In Norway, 8 of the 43 hospitals can be classed as university hospitals and the rest as central or local hospitals.

The Danish hospitals were owned and run by the 14 counties. In Denmark, there were 10 university hospitals, yet some of them were clearly smaller units compared to e.g. Swedish and Finnish university hospitals. The total

number of Danish hospitals in this study was 54 (there were approximately 60 public hospitals in total in the country).

In Finland, the executive management of hospitals usually consists of a chief physician, a chief nurse and a director of finance, while hospital managers are accountable to the council of the hospital district. In Sweden, the hospitals are managed by a combination of elected public officials sitting on boards and also hospital managers. In Denmark, for its part, a recent trend has been to merge management functions and/or create matrix organisations by merging departments from several hospitals into functional units with joint responsibilities for particular treatment areas. Somewhat on the same path is Norway as the boards of the local health trusts are managing the hospitals, though it has been noted that the management role of the trust boards is not clear [9, 10]. In the Nordic countries most of health care professionals work on salary basis and their wage level is based on nationwide contracts between employer and employee organisations. The payment of hospital staff was salary based in all Nordic countries.

2.3 Hospital data generating processes

The basic model for data generation is rather similar in all of the countries in this study. Medical personnel are responsible for feeding in the key medical record information (age, sex, primary and secondary diagnoses) in the hospital's system. After discharging the patient, the hospital's data administration merges data sets from various hospital systems (e.g. the procedures from the operating room information technology systems, systems for medical diagnostics, patient administration systems for the wards, outpatient records) to create a standardized discharge abstract. After quality checking, this discharge abstract is sent on to the national statistical authorities to form a national discharge registry.

However, although it is the physician's responsibility to give a diagnosis for the patient, there are differences in how this given diagnosis is interpreted as a valid ICD-10 diagnosis and coded accordingly in the system. In some cases it may be fed in the system by nurses or assistants on the wards, sometimes by the physicians. It is known that the diagnosis coding practices differ in the hospitals within countries and also across countries. For example, the propensity of using secondary diagnoses varies considerably across the countries and this variation may affect the comparisons markedly. It was already noted in the comparison between Norway and Finland in 1999 that the relative shares of patients with serious co-morbidities affecting the DRG grouping were significantly higher in the Norwegian hospitals [8]. Another potential weakness in the data generating process was in the outpatient records.

Many inpatient treatments were already shifted to outpatient care during the early 2000s, but the coding of diagnoses was not of sufficiently high quality in the IT systems for outpatient care. However, each of the Nordic countries in the present study shared the same types of problems in the data administration, and consequently our hypothesis was that there were not large systematic country-level differences in these aspects of data quality.

3 Data

The data set was based on national discharge databases collected from hospitals from Finland, Norway, Sweden and Denmark. In the current study, data from 2002 were used. The year 2002 was the last year when the Danish hospitals could be measured using the common NordDRG grouper since the Danish Ministry of Health decided to build their own grouping system (the sc. DKDRG grouper) and also a separate outpatient grouping system (DAGS) for ambulatory patients [11].

Outputs were measured as (DRG) weighted discharges in inpatient care, day surgery, day-care and the number of outpatient visits. DRG weights were averages calculated from the national cost weight sets. Inputs were measured as costs in real terms, using wage and consumer price indices to adjust operating costs. Descriptive statistics are shown in Table 1.

3.1 Output data—DRG grouping

In 1996, the Nordic countries launched a modified DRG system based on the Nordic version of the ICD-10 and a new Nordic Classification of Surgical Procedures (NCSP) introduced in 1996 (<http://www.norddrg.net/norddrgmanual/>). The current version of the NordDRG applies to Nordic diagnosis and procedure codes but imitates the classification rules in the 12th edition of the DRG classification issued by the US-HCFA in 1994. With a common Nordic version, each country has its own localised national versions. Fortunately, during the study period the national versions were not too diversified and thus in this study we could group inpatient admissions and day care episodes using each country's own NordDRG versions. However, it was necessary to do slight modifications in the primary groupings to ensure comparability. Primary grouping of the patient data was done with yearly versions of national groupers. First of all, all cases where DRGs were split into subgroups in the national versions were aggregated back to the original grouping.¹ Secondly, in Norway there were considerable volumes in some DRG groups that did not exist in the other countries. For example, the number of normal newborns were grouped

and counted as output whereas in other Nordic countries delivery related DRGs included only the hospital discharges for mother's stay at the maternity wards. This was necessary only for the normal newborns since any other types of problems with the newborn would be counted in paediatric DRG groups. In addition, the Norwegian inpatient grouping included significant volumes of rehabilitation, dialysis treatment and radiation therapy. These treatments were provided mainly in day care or outpatient visit settings in the other countries. These discrepancies were adjusted in the data used in this study.²

3.2 Output data—the definition for a DRG case (discharge)

Before grouping the output data, one critical task was to harmonize the definition of a 'discharge'. It turned out that in Norway and Denmark discharges were defined as 'hospital discharges' while in Sweden and Finland the discharges were defined as 'specialty discharges'. Specialty discharge means that if the patient is transferred to other clinics within the same hospital, a new discharge is counted. Due to the varying definitions for a discharge, Swedish and Finnish hospital data had to be fixed by merging patient discharge data in cases where clinical transfers were found. The main diagnosis for the hospital discharge was inherited from the specialty discharge having the largest DRG cost weight.

3.3 Output data—the definition for day care

Day care included cases where the patient did not stay overnight in the patient ward, but where treatment was considerably more resource intensive than in the usual outpatient visits. Short stay surgery (e.g. cataract surgery) is a typical example of day care in the operative specialties, renal dialysis in the medical specialties. Similarly, there were large differences among the Nordic countries within the organization and supply of day care services. Day care cases were grouped using the local NordDRG groupers and weighted accordingly.

3.4 Output data—the definition for outpatient visits

Outpatient visits were measured as simple counts without case-mix weighting. Outpatient visits included all emergency

¹ Stroke subtypes 14A and 14B were merged into DRG 14 using average cost weights.

² All hospital admissions with DRG 391 (Normal newborn) and DRG 433 (Alcohol/drug use with comorbid conditions) were removed. Day-care admissions within DRGs 317 (Admit for renal dialysis), 410 (Chemotherapy without acute leukaemia as secondary diagnosis) and 462 (Rehabilitation) were counted as outpatient admissions. All admissions within DRGs 373 (Vaginal delivery without complicating diagnoses) and 381 (Abortion with d&c, aspiration curettage or hysterotomy) were treated as regular inpatient treatments (and thus not as day-care).

Table 1 Descriptive statistics

Variable	Country			
	Denmark	Sweden	Finland	Norway
Number of hospitals	54	49	38	43
Operating costs, wage adjusted (MEUR)	71.3	134.8	72.6	72.2
Input price index	1.199	1.147	0.930	1.214
DRG-weighted inpatient surgical cases (IS)	6,794.9	10,496.0	8,345.4	7,106.6
DRG-weighted inpatient medical cases (IM)	9,073.5	13,448.0	9,278.9	9,057.1
DRG-weighted surgical day-care cases (DS)	0.0	140.0	1,430.8	1,546.1
DRG-weighted medical day-care cases (DM)	0.0	328.7	89.7	199.0
Other DRGs (ODRG)	1,369.9	948.4	490.9	851.9
All DRGs	17,238.3	25,361.1	19,635.5	18,760.7
Outpatient visits	123,610.7	160,260.4	124,563.2	72,006.0
Average DRG case-mix index	1.00	1.05	0.93	1.00
Average % difference in DRG adjusted LOS (observed-expected)	2.6	2.8	−9.8	−11.9

and scheduled visits, and first and follow-up visits. The treatment in a typical outpatient visit may include minor diagnostic procedures or treatments, but usually more resource intensive cases are defined as day care (see above). It is possible that Nordic countries use slightly different definitions for outpatient visits and conservative day-care. Thus there may be slight differences in the outpatient visit output. In operative day care the inclusion criteria was rather clear since an explicit list of procedure codes was used.

3.5 Output data—DRG cost weights

All patient cases (discharges and day-care) were grouped and weighted using the estimated average costs in each country separately. Because the national DRG weights varied, we used weighted average DRG-weights to aggregate inpatient care and day care to obtain a common set of weights. Because cost-accounting data at the patient level were not produced regularly in all hospitals in the Nordic countries, the cost weights were derived from the cost information of samples of hospitals in each country. In Finland, the cost weights were derived from the cost information of the largest hospital district, Helsinki-Uusimaa district (HUS). The costing sample covered approximately 30% of all acute hospital care in Finland. Cost items included diagnostic tests, procedures, medical services, support services and overhead costs. In Norway the cost weights are adjusted annually in a national price list for the DRG system and were based on a sample of the hospitals based on average price per DRG per year. In Sweden the cost weights, since 1999, are derived from the calculated cost per patient information of a sample of hospitals that covers about 50% of all acute hospital care in 2008 (with an overweight of regional hospitals).

Finally, for the analysis in the present study, the output was grouped to five categories: inpatient medical case types (IM); inpatient surgical case types (IS); day care/outpatient surgical products (OS), day care medical cases (OM), other DRG cases (ODRG) and outpatient visits (OV).

3.6 Input data

The cost data were based on the year-end accounts (according to the hospital book-keeping) and internal reports from the hospital accounting systems. In this study, hospital costs include all production related costs in a hospital, excluding capital costs and costs of teaching and research. In order to compile the cost data into a common currency, hospital costs were adjusted using the input price index. Input price adjusted operating costs (ADJ_COST) were used in all analyses.

3.7 Input prices

International comparisons of health expenditure and health prices must be based on a common currency. Purchasing power parities (PPP) are rates of currency conversion constructed to account for such price differences. Generally, the reported PPPs adjust for price differences at the level of the total GDP, not sub-aggregates of the GDP, such as health expenditures. However, cross-country differences in health care prices are not necessarily consistent with differences in prices in general.

To approximate an input price index, we weighted 70% (the average share of wage expenditures) of hospital operating costs using the wage index and the rest 30% (e.g. materials, equipment and rents) using the PPP conversion adjustment. The share of nurses wages accounted for 50% of the total expenditure and the physician wages 20%. Table 1 presents

the price indices employed for each of the Nordic countries. The input prices were substantially lower in Finland than in other Nordic countries.

4 Methods

The measurement of cost efficiency is usually accomplished by using parametric stochastic frontier (SF) methods or nonparametric Data Envelopment Analysis (DEA). Parametric estimation of the stochastic frontier needs a behavioural hypothesis for cost minimisation. In addition, the econometric approach is parametric and confounds the effects of misspecification of functional form with inefficiency. DEA approach is nonparametric and, due to milder conditions set for the form of the technology, is less prone to this type of specification error [12]. DEA is based on relative efficiency measures proposed by Farrell [13], and in this framework a hospital is judged to be efficient if it is operating on the best practice production frontier. By employing the input price adjusted operating costs as the input variable our definition of cost efficiency only approximates Farrell's measure of total efficiency [13].

In assessing the cost efficiency of hospitals in this study we used DEA which utilizes linear programming techniques in the calculation of unit-specific efficiency scores [14]. DEA constructs a piecewise linear efficient frontier which serves as the reference in the evaluation of efficiency. If a hospital is efficient, it lies on the frontier and will receive an efficiency score of 1.0 (100% efficiency). Inefficient hospitals will receive a score lower than 1.0. For example, if the score for a hospital is 0.80 as measured in the input direction, its inefficiency is 20% and it could produce its output with 20% less input. Alternatively, with an output-efficiency score it produces 80% of its potential and it could increase its output by 25% using the same resources. With the CRS assumption the scores would be the same whether measured in the input or output direction. Bias-corrected efficiency estimates can be obtained by using bootstrapping methods from simulated distributions in pseudo samples [15, 16]. If the hospital size distribution is not similar, scale assumptions may influence not only the individual efficiency measures, but also the group averages. Therefore we presented results using models with both constant and variable returns to scale.

Cost efficiency was calculated by solving the following linear program:

$$\begin{aligned} & \text{Min}_{\lambda, z_{CE}} z_{CE} \\ & \text{s.t.} \quad \lambda \cdot Y \geq y_0, \\ & \quad \lambda \cdot C \leq z_{CE} \cdot c_0, \\ & \quad \lambda_i \geq 0 \\ & \quad \lambda \cdot i = 1 \end{aligned} \quad (1)$$

where Y is an $n \times m$ matrix of observed outputs and λ is a $1 \times n$ vector of intensity variables. c_0 is a scalar representing a hospital's cost level, i is a column vector of 1s and C is the $n \times 1$ matrix of observed costs. Eliminating the last equality constraint changes the model to constant returns to scale.

Three models were constructed in order to study the differences in estimated DEA efficiency scores across different model specifications. The number of output variables were varied using different data aggregations. In Model 1, DRG grouped inpatient and day-care cases were aggregated and outpatient visits were used as the second output. In Model 2 the specification was based on 3 outputs where inpatient and day care cases were separated. Model 3 used 6 outputs where inpatient discharges and day care were further split into subgroups: medical, surgical and other cases in the inpatient care and medical and surgical for day care. Each of the three models was calculated using assumptions for constant returns to scale (CRS) and variable returns to scale (VRS). All used models were input-oriented and solved using a robust solver XA from Sunset Software Technology. The characteristics of the models are summarised in Table 2.

However, DEA models are known to be sensitive to variable selection. For example, in DEA every inclusion of an input or output variable increases the resulting average efficiency scores, as noted by Farrell [13]. Moreover, the production unit is evaluated along its best dimension(s) and will continue to be efficient no matter how many variables are added to the model. It is important to note that the bootstrapping methods do not fix the curse of dimensionality problem and thus it is informative to report the results using rank-order methods (e.g. the number of units on the frontier) to supplement mean and standard deviations. The curse of dimensionality occurs in DEA when there is an excessive number of inputs and outputs in relation to the number of decision-making units. Following the recommendations for model specification discussed in [12], three models with varying number of output variables were used to test the sensitivity of the results.

5 Results

This study revealed considerable differences in cost efficiency between Nordic hospitals. In 2002, the average efficiency was highest in Finland, where the mean efficiency was between 0.73 and 0.80 in models using CRS and between 0.86 and 0.88 in models using VRS (Table 3). In Denmark the average efficiency was closest to the Finnish average, with a difference of only 0.00–0.09 efficiency units. The difference in the average efficiency between Danish and Finnish hospitals was clearly larger using the VRS models. Sweden appeared to have the least

Table 2 Models used in the analysis

No	Model name	Inputs	Outputs	# of output variables	Scale	Comments
1	Model1_CRS	ADJ_COST	(IM+IS+DM+DS+ODRG), OV	2	CRS	All DRGs grouped as one output
2	Model2_CRS	ADJ_COST	(IM+IS+ODRG), (DM+DS),OV	3	CRS	Separate inpatient, daycare and other DRGs
3	Model3_CRS	ADJ_COST	IM, IS, DM, DS, ODRG, OV	6	CRS	Separate medical and surgical cases in the DRGs
4	Model4_VRS	ADJ_COST	(IM+IS+DM+DS+ODRG), OV	2	VRS	All DRGs grouped as one output
5	Model5_VRS	ADJ_COST	(IM+IS+ODRG), (DM+DS),OV	3	VRS	Separate inpatient, daycare and other DRGs
6	Model6_VRS	ADJ_COST	IM, IS, DM, DS, ODRG, OV	6	VRS	Separate medical and surgical cases in the DRGs

Costs are adjusted using the input price index

efficient hospitals with a difference of 0.13–0.20 efficiency units compared to Finnish hospitals in the average efficiency.

While the individual hospital scores and even the (country) average efficiency scores varied markedly in different model specifications, the rank in the (country) group averages remained the same in all of the used models. The differences in group averages seemed generally to be higher in models with less output variables and models using constant returns to scale.

In Fig. 1 the bias-corrected efficiency scores and the bootstrapped 95% likelihood intervals are displayed for each of the hospitals for Model5. According to preliminary results there was more variation in cost efficiency among Swedish and Finnish hospitals, whereas the variance of efficiency scores was smallest in Norwegian hospitals (Fig. 1). Likelihood intervals seemed to be somewhat wider in the Finnish hospitals (Appendix 1).

In all countries there were fully efficient hospitals, depending on the model specification. The least number of efficient units were in Model1 where only one Danish and one Finnish hospital were estimated to be fully efficient. Among the most efficient hospitals there were many types and sizes of hospitals ranging from small local

hospitals to largest university hospitals. However, in the CRS models the small local hospitals were clearly overrepresented whereas using the VRS models tended to significantly increase the relative performance of larger university hospitals.

6 Discussion

Our findings showing substantial differences in average cost efficiency were quite robust using different models. Moreover, after adjusting for input price differences, there still remained a difference of 0.13–0.20 in the efficiency measures, suggesting that efficiency may have been clearly lower in Norwegian and Swedish hospitals. However, although the average efficiency in Danish hospitals was slightly lower than in Finnish hospitals (0.00–0.09 units), the ranking for Denmark may be sensitive to changing some of the assumptions used in this comparison.

The country-level average differences in efficiency have been found to be surprisingly stable in time. Based on updated data sets from 2005 to 2007, hospital efficiency in the Nordic countries was assessed in a recent working paper [17]. The average efficiency in Finnish, Swedish and

Table 3 DEA efficiency scores and the number of efficient units

		Mean DEA scores and the number of efficient units							
		Denmark		Sweden		Finland		Norway	
		Mean	# efficient	Mean	# efficient	Mean	# efficient	Mean	# efficient
CRS	Model1	0.68	1	0.53	0	0.73	1	0.64	0
	Model2	0.77	1	0.60	0	0.78	2	0.67	1
	Model3	0.80	8	0.65	0	0.80	5	0.74	3
VRS	Model4	0.77	5	0.64	2	0.86	6	0.75	1
	Model5	0.84	7	0.69	2	0.86	11	0.77	3
	Model6	0.85	18	0.75	8	0.88	15	0.80	9

Fig. 1 Bias-corrected efficiency scores for Model5

Norwegian hospitals converged only slightly and calculations using preliminary and partly incomplete data suggested that significant differences in average efficiency had prevailed also in 2005–2007.

The present study also demonstrated some important issues which have to be taken into account in country-level international comparisons: 1) cost efficiency differences may turn out to be substantial, explaining a large of part of the differences in health expenditure, 2) The estimates of average cost efficiency are however somewhat sensitive to the model specification, and if data comparability is ensured only the country ranking can be reliably demonstrated 3) PPP adjustments can be misleading in cases where input prices are clearly different in the health sector, such as in our case where, compared to Finnish hospitals, the hospital wage index in 2002 was 30% higher in Norway, and 28% higher in Denmark. However, the differences in the average wage levels of hospital personnel used in the present study were consistent with the reported official statistics for health workforce remuneration by OECD [18]. Further investigations of a proper input price index are still needed.

The search for explanations for cost efficiency differences using the present data proved to be difficult. Our cross-sectional analysis does not allow us to make far-reaching conclusions on any causal relationships. In the previous comparative study between Finland and Norway, the institutional, structural and technical explanations were discussed [8]. First, there had been actions to improve the organisation of health care delivery between primary and specialised health care in Finland, whereas in Norway patients crossed an “administrative border” when leaving the hospital. There was also discussion that municipalities in Norway were able to shift part of the costs of health care to the hospitals, thereby creating longer lengths of stay and lower levels of efficiency.

Third, according to the previous study there were substantial differences in the organization of care for some patient groups between Norway and Finland [8]. This pointed to structural differences and the most distinct difference was the higher number of outlier days. The reason why the number of outlier days/discharge were significantly higher in Norway may be due to a) different DRG case-mix, b) generally longer LOS in Norway, c) larger variation in inpatient care (bed-days).

The fourth factor explaining the differences is more technical and relates to coding issues and the accuracy of the case-mix adjustment. According to the results, regional hospitals, which usually provide more specialized services, seemed to be less efficient compared to local hospitals using the constant returns to scale assumption. The regional hospitals were mostly large hospitals, which made it difficult to distinguish between technology and other correlates of efficiency. One possible explanation could be that the large regional hospitals are often teaching hospitals. As shown in previous studies, teaching hospitals probably have special characteristics which may affect efficiency comparisons [19]. It may be that teaching and research activities absorb more inputs than are compensated for teaching and research activities. It is also possible that the DRG case-mix adjustment does not capture all case complexity adequately. However, since the hospitals’ size distribution in the countries was not similar, this may have affected our findings relating to differences in country-level efficiency. For example, the size of the Swedish hospitals was clearly larger judged by the statistics in Table 1.

In this study we observed that there were some differences in the average case-mix between the Nordic countries (Table 1). This may be due to coding differences. Using patient-level data, it was possible to provide insight into the differences of patient care by looking more closely at the produced DRGs in each country. For each of the DRG

groups the relative share of cases within each DRG group was compared. In addition, we compared the average case-mix, LOS and case-mix weighted cases/population in every DRG group and wider aggregates of DRGs, the major diagnostic categories (MDC) (Fig. 2).

According to our analysis of individual DRGs, the largest differences were found among the typical day surgery/day care cases, such as tonsillectomies, hernias, cataract surgeries, abortions, vein ligations and carpal tunnel releases. These cases were underreported in the Swedish and Danish hospitals compared to Finnish and Norwegian hospitals, possibly a reflection that in Sweden and Denmark these cases are mainly reported as outpatient visits. Thus, the total output in Swedish and Danish hospitals may be slightly underestimated.

Another large discrepancy was found in the conservative treatment of some high-volume patient groups in internal medicine. There were substantially fewer cases of atherosclerosis with co-morbidities in the Finnish hospitals, perhaps due to differences in the use of secondary diagnoses. A similar difference could be observed in the cardiac arrhythmia cases with and without complications. Moreover, the number of chest pain and angina pectoris cases was clearly lower in Finnish hospitals, which may be due to either coding or treatment differences in cardiovascular diseases.

The percentage of acute myocardial infarctions (AMI) with cardiovascular complications (DRG 121) in the total number

of AMIs leaving hospital alive was 48.5% in Norway but only 33.1% in Finland. This must be due to differences in coding practices which favoured Norwegian hospitals in the productivity comparison. However, further observation of coding practices revealed that 48% of deliveries were accompanied with complicating secondary diagnoses in Norway (38% in 1999) and 49% in Finland (9% in 1999). The percentages for ‘complicated’ Caesarean sections were also higher in Finland, 40% and 52% respectively.

However, exploring the MDCs revealed that although there were differences in the prevalence of individual DRGs, aggregation of output to MDC level cancelled out the effects of coding individual DRGs. The variation across the Major Diagnostic Categories (MDCs) was not large, indicating that the MDC distribution of cases was rather similar in the Nordic countries. In Figure 2 the difference between the observed and expected case-mix adjusted output is presented, measured as the percentage of total hospital output in each country. In most of the MDCs the difference was less than 1%, except for a few cases. In Norway, the higher case-mix in MDC 0 (Pre-MDC groups), MDC 5 (Diseases and disorders of the circulatory system) and MDC 18 (Infectious and parasitic diseases) gave some 1–2% (measured in efficiency units) advantage in hospital-level efficiency comparisons. In Finland, cases in MDC 18 were also seemingly more severe. In Sweden, the case-mix in MDC 9 (Diseases and disorders of the skin and subcutaneous tissue) was substantially higher

Fig. 2 The difference between the observed and expected case-mix adjusted output in each of the MDCs (MDC 1–26). Results are presented as the percentage of total output

while the lower case-mix in the same MDC gave some disadvantage for Norwegian and Finnish hospitals. It can be concluded that systematic bias in the casemix measurement can explain only a small part of the observed differences in efficiency.

Although there were several potential biases in the used data and methods as discussed above, it seems that there remains unexplained differences in the average cost efficiency between the Nordic countries. Identifying the causes of these differences is, however, quite challenging since it involves both substantive and methodological problems. Firstly, while the main differences between countries identified here would be captured by country dummies in a second stage regression, there are only four countries. This makes it impossible to statistically separate effects that only vary between and not within countries, although one might hope to generate fruitful hypotheses. Secondly, the variables that vary within each country (between hospitals or over time) are often difficult to operationalize and data may not be available. Thirdly, there are methodological challenges in modelling simultaneously the causes and extent of efficiency, particularly in the bootstrapped non-parametric DEA methodology [19].

Despite the methodological challenges, our findings could be important in assessing the performance of health care and further analyses may reveal interesting policy implications.

Using a second stage analysis and panel data from 1999–2004 and 2005–2007, it was possible to investigate the effect of hospital reform (centralizing ownership), size (economies of scale), activity based financing and treatment practices (variation in the average DRG adjusted length of stay and the share of outpatient treatment) on the productivity [17, 20]. The ownership reform in Norway was found to increase productivity. Positive deviation from expected LOS was associated with reduced productivity, as expected. Interestingly, these studies failed to detect any clear effects of the changes in activity based financing. Some evidence on the diseconomies of scale was found in [17].

Kittelsen et al. 2008 concluded that structural changes and better management may be the most likely explanations for increased efficiency [20]. In a comparison of Finland

and Sweden, Finnish hospital care system would be expected to differ from the Swedish system in productivity in two dimensions. Firstly, the purchaser–producer split favours higher productivity in Finland, while the second, activity-based funding (ABF), favours higher productivity in Sweden. According to previous analysis, the impact of ABF on productivity differences between the countries was modest [20]. Higher productivity in Finland may thus be partly related to the role of the purchaser. The cost control by municipalities in Finland may to be more effective than that of the counties/regions in Sweden and Denmark, or the central government in Norway. The Finnish municipalities are responsible for other public services in addition to health care. Thus their resource allocation to hospital care must be balanced annually with the allocation of resources to other sectors, while increases in hospital costs must be financed either by increasing the local tax rate or by diminishing the resources allocated to social services including the day care for children.

The substantial differences in productivity between Nordic hospitals warrant further investigation and there are still unknown factors which make the transformation process from inputs to health care services more efficient in Finnish hospitals. A better insight to these can only be achieved by collecting more detailed data on the various resources used (labor, materials, capital) and investigating the technical efficiency in the hospital production. One obvious expansion already started by our research group (the Nordic Hospital Comparison Group) is to include various measurements of hospital quality in the future comparisons. In addition, the costs and outcomes in several large patient groups in the Nordic secondary and tertiary care will be explored in the 7th Framework Programme's EuroDRG and EuroHOPE projects.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix 1. Bias-corrected efficiency scores and 95% confidence limits for Model5

Denmark				Sweden				Finland				Norway			
OBS	CorrEst	Low	High	OBS	CorrEst	Low	High	OBS	CorrEst	Low	High	OBS	CorrEst	Low	High
1	0.556558	0.535494	0.574196	55	0.341884	0.331373	0.349822	104	0.613064	0.591137	0.631318	142	0.514819	0.493057	0.536902
2	0.634451	0.618137	0.645262	56	0.362707	0.350231	0.371217	105	0.650127	0.552827	0.7301	143	0.522969	0.475966	0.552092
3	0.644062	0.627228	0.654814	57	0.391038	0.377697	0.401153	106	0.674864	0.625325	0.705925	144	0.591418	0.563262	0.61623
4	0.65034	0.625279	0.671888	58	0.396077	0.381252	0.409074	107	0.697308	0.666248	0.719076	145	0.592603	0.570271	0.610397

Appendix 1 (continued)

Denmark				Sweden				Finland				Norway			
OBS	CorrEst	Low	High	OBS	CorrEst	Low	High	OBS	CorrEst	Low	High	OBS	CorrEst	Low	High
5	0.651911	0.629854	0.671248	59	0.422722	0.407453	0.436016	108	0.706348	0.622935	0.763825	146	0.601751	0.576274	0.627806
6	0.668483	0.651068	0.679649	60	0.43239	0.417627	0.442831	109	0.706684	0.647355	0.740355	147	0.612856	0.590731	0.62895
7	0.675248	0.65733	0.686107	61	0.435046	0.420356	0.445828	110	0.708499	0.681284	0.736294	148	0.613085	0.589209	0.62977
8	0.676205	0.650658	0.700365	62	0.442883	0.427923	0.454645	111	0.708921	0.678244	0.739927	149	0.615219	0.591852	0.637769
9	0.687326	0.65506	0.705893	63	0.45343	0.43775	0.464227	112	0.710187	0.681403	0.738734	150	0.625916	0.601925	0.642888
10	0.688138	0.669988	0.701684	64	0.458873	0.443496	0.470525	113	0.722092	0.696558	0.743478	151	0.626625	0.589246	0.649109
11	0.695133	0.658941	0.715313	65	0.473398	0.461244	0.481161	114	0.722727	0.672225	0.753213	152	0.627566	0.60348	0.64573
12	0.706426	0.678426	0.731617	66	0.474568	0.458358	0.486136	115	0.725499	0.678518	0.759804	153	0.627596	0.603398	0.644692
13	0.709718	0.681516	0.735968	67	0.485688	0.467156	0.50184	116	0.726677	0.692599	0.749421	154	0.628423	0.602417	0.646718
14	0.71401	0.685654	0.740991	68	0.488858	0.466661	0.501778	117	0.731617	0.709429	0.749235	155	0.642054	0.619145	0.662013
15	0.714206	0.686891	0.737423	69	0.524807	0.503885	0.537154	118	0.742875	0.692767	0.775306	156	0.645972	0.62255	0.668556
16	0.714266	0.693816	0.729831	70	0.525103	0.509245	0.534363	119	0.748433	0.705475	0.775198	157	0.649202	0.625402	0.666412
17	0.717918	0.695077	0.736301	71	0.538112	0.519556	0.551758	120	0.752801	0.724932	0.774194	158	0.656229	0.632375	0.678407
18	0.725369	0.651902	0.770368	72	0.569803	0.550249	0.583158	121	0.756708	0.720678	0.782003	159	0.656624	0.632777	0.674172
19	0.726619	0.707712	0.738726	73	0.570735	0.551041	0.58475	122	0.774207	0.741106	0.796928	160	0.660499	0.631087	0.679669
20	0.728282	0.709117	0.742495	74	0.574262	0.557854	0.586835	123	0.779352	0.742091	0.803517	161	0.663102	0.638993	0.680351
21	0.737702	0.690492	0.763057	75	0.591293	0.570983	0.605493	124	0.779541	0.740771	0.805877	162	0.663384	0.6332	0.683765
22	0.746496	0.726263	0.758761	76	0.592119	0.571768	0.606241	125	0.781677	0.707264	0.838761	163	0.677643	0.555721	0.743621
23	0.74794	0.728204	0.762434	77	0.593623	0.574739	0.608572	126	0.785581	0.74435	0.81273	164	0.678996	0.647029	0.699599
24	0.755534	0.735047	0.767945	78	0.609034	0.593094	0.619967	127	0.790489	0.7616	0.818945	165	0.689521	0.661626	0.711541
25	0.75628	0.727367	0.7809	79	0.612709	0.596275	0.624948	128	0.790648	0.477907	0.992567	166	0.695329	0.665173	0.715982
26	0.757355	0.728304	0.774605	80	0.61656	0.536936	0.68217	129	0.800473	0.720145	0.861857	167	0.698812	0.660033	0.723618
27	0.759981	0.737109	0.777483	81	0.633236	0.616157	0.644013	130	0.804522	0.772034	0.832738	168	0.69942	0.673708	0.718286
28	0.77738	0.744213	0.796578	82	0.637935	0.620877	0.650669	131	0.809636	0.754801	0.850234	169	0.700372	0.670984	0.721989
29	0.779279	0.756433	0.796544	83	0.639791	0.599994	0.66084	132	0.817772	0.750438	0.856711	170	0.700941	0.652251	0.728885
30	0.793121	0.766358	0.816455	84	0.647333	0.621855	0.671923	133	0.825548	0.79293	0.848171	171	0.705544	0.677332	0.725185
31	0.795002	0.772307	0.812387	85	0.648533	0.629366	0.660084	134	0.860427	0.807986	0.893382	172	0.707102	0.681071	0.727068
32	0.803239	0.778812	0.817601	86	0.663068	0.645678	0.674655	135	0.871896	0.833187	0.901606	173	0.710762	0.681293	0.731564
33	0.810476	0.774472	0.831076	87	0.670843	0.644913	0.695979	136	0.888521	0.854792	0.916357	174	0.713708	0.687828	0.73328
34	0.815618	0.790852	0.830327	88	0.671787	0.648597	0.690891	137	0.894745	0.83967	0.935213	175	0.723757	0.672967	0.752857
35	0.81806	0.787764	0.842494	89	0.67859	0.660678	0.691706	138	0.900612	0.785013	0.970206	176	0.735467	0.701615	0.758467
36	0.822224	0.794103	0.843921	90	0.684194	0.663305	0.696315	139	0.901794	0.813947	0.951135	177	0.739203	0.712423	0.758567
37	0.824986	0.799771	0.839912	91	0.691257	0.665772	0.7127	140	0.906544	0.866042	0.937004	178	0.747511	0.72268	0.768726
38	0.826478	0.798487	0.851051	92	0.691676	0.670634	0.704165	141	0.948272	0.906722	0.993927	179	0.750723	0.722276	0.770728
39	0.826963	0.79416	0.856879	93	0.694655	0.671277	0.711725					180	0.76281	0.735132	0.789268
40	0.835946	0.807793	0.852848	94	0.707073	0.68829	0.718961					181	0.764346	0.736787	0.784658
41	0.836251	0.804127	0.863799	95	0.707308	0.688625	0.720523					182	0.783641	0.757313	0.806407
42	0.837808	0.80986	0.858316	96	0.721669	0.696684	0.742116					183	0.79115	0.418597	0.991204
43	0.848783	0.819653	0.874717	97	0.736447	0.704924	0.755175					184	0.791275	0.764496	0.811235
44	0.851999	0.822659	0.874942	98	0.74822	0.720274	0.774488								
45	0.855303	0.817963	0.876773	99	0.762517	0.736225	0.784599								
46	0.86808	0.83819	0.89146	100	0.804926	0.777336	0.82395								
47	0.871442	0.841727	0.89119	101	0.820532	0.793035	0.840573								
48	0.874064	0.742719	0.989837	102	0.887849	0.863322	0.906433								
49	0.888902	0.777279	0.966514	103	0.891121	0.832935	0.922364								
50	0.901692	0.875015	0.921761												
51	0.903189	0.872565	0.926215												
52	0.916592	0.884139	0.942317												
53	0.917916	0.88523	0.947422												
54	0.921328	0.885621	0.955546												

References

- Hurst J, Jee-Hughes M (2001) Performance measurement and performance management in OECD Health systems. Labor market and social policy—Occasional papers no 47, OECD
- Häkkinen U, Joumard I (2007) Cross-country analysis of efficiency in OECD health care sectors: options for research. Economics Department Working Papers No. 554, OECD
- European Commission (2008) Hospital data project phase 2. Final report part II. Annex Diagnosis (<http://ec.europa.eu/health>)
- Hansen KK, Zwanziger J (1996) Marginal Costs in general acute care hospitals: a comparison among California, New York and Canada. *Health Econ* 5:195–216
- Mobley L, Magnussen J (1998) An international comparison of hospital efficiency. Does institutional environment matter? *Appl Econ* 30:1089–1100
- Magnussen J, Mobley L (1999) The impact of market environment on excess capacity and the cost of an empty hospital bed. *Int J Econ Bus* 6:383–398
- Steinman L, Dittrich G, Karmann A, Zweifel P (2004) Measuring and comparing the (in)efficiency of German and Swiss hospitals. *Eur J Health Econom Issue* 5:216–226
- Linna M, Häkkinen U, Magnussen J (2006) Comparing hospital cost efficiency between Norway and Finland. *Health Policy* 77:268–278
- Magnussen J, Hagen T, Kaarboe O (2007) Centralized or decentralized? A case study of Norwegian hospital reform. *Soc Sci Med* 64(10):2129–37
- Hagen T, Vrangbæk K (2009) The changing political governance structures of Nordic health care systems. In: Magnussen J, Vrangbæk K, Saltman R (eds) *Nordic health care systems. Recent reforms and current policy challenges*. Open University Press, London
- Ankjaer-Jensen A, Rosling P, Bilde L (2006) Variable prospective financing in the Danish hospital sector and the development of a Danish case-mix system. *Health Care Manag Sci* 9:259–268
- Smith P (1997) Model misspecification in data envelopment analysis. *Ann Oper Res* 73:233–252
- Farrell M (1957) The measurement of productive efficiency. *J R Stat Soc* 120:253–281
- Charnes A, Cooper W, Rhodes E (1978) Measuring the efficiency of decision-making units. *Eur J Oper Res* 2:429–444
- Simar L, Wilson P (1998) Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Manage Sci* 44:49–61
- Simar L, Wilson P (2001) Testing restrictions in nonparametric efficiency models. *Comm Stat Sim Comput* 30(1):159–184
- Kittelsen S, Kjartan S, Kalseth B, Kalseth J, Halsteinli V, Magnussen J (2009) En komparativ analyse av spesialisthelsetjenesten i Finland, Sverige, Danmark og Norge: Aktivitet, ressursbruk og produktivitet 2005–2007, Rapport A12200. SINTEF, Trondheim
- Health at a Glance. OECD indicators, OECD 2009
- Grosskopf S, Margaritis D, Valdmanis V (2001) The effects of teaching on hospital productivity. *Socio-Econ Plann Sci* 35:189–204
- Kittelsen S, Magnussen J, Anthun K, Häkkinen U, Linna M, Medin E, Olsen K, Rehnberg C (2008) Hospital productivity and the Norwegian ownership reform. A Nordic comparative study. HERO Working Paper 10/2008, Oslo University