HEALTH ECONOMICS
Health Econ. **24**(Suppl. 2): 140–163 (2015)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/hec.3260

COSTS AND QUALITY AT THE HOSPITAL LEVEL IN THE NORDIC COUNTRIES

SVERRE A. C. KITTELSEN[a,*], KJARTAN S. ANTHUN[b], FANNY GOUDE[c], INGRID M. S. HUITFELDT[a], UNTO HÄKKINEN[d], MARIE KRUSE[e], EMMA MEDIN[c], CLAS REHNBERG[c], HANNA RÄTTÖ[d]
ON BEHALF OF THE EUROHOPE STUDY GROUP

[a]*Frisch Centre, Oslo, Norway*
[b]*SINTEF Health Research and NTNU, Trondheim, Norway*
[c]*Medical Management Centre, Karolinska Institutet, Stockholm, Sweden*
[d]*Centre for Health and Social Economics CHESS, National Institute for Health and Welfare, Helsinki, Finland*
[e]*COHERE, University of Southern Denmark, Odense, Denmark*

ABSTRACT

This article develops and analyzes patient register-based measures of quality for the major Nordic countries. Previous studies show that Finnish hospitals have significantly higher average productivity than hospitals in Sweden, Denmark, and Norway and also a substantial variation within each country. This paper examines whether quality differences can form part of the explanation and attempts to uncover quality–cost trade-offs.

Data on costs and discharges in each diagnosis-related group for 160 acute hospitals in 2008–2009 were collected. Patient register-based measures of quality such as readmissions, mortality (in hospital or outside), and patient safety indices were developed and case-mix adjusted. Productivity is estimated using bootstrapped data envelopment analysis.

Results indicate that case-mix adjustment is important, and there are significant differences in the case-mix adjusted performance measures as well as in productivity both at the national and hospital levels. For most quality indicators, the performance measures reveal room for improvement. There is a weak but statistical significant trade-off between productivity and inpatient readmissions within 30 days but a tendency that hospitals with high 30-day mortality also have higher costs. Hence, no clear cost–quality trade-off pattern was discovered. Patient registers can be used and developed to improve future quality and cost comparisons.

Received 5 March 2014; Revised 27 April 2015; Accepted 11 May 2015

KEY WORDS: international comparisons; quality; outcomes; performance; productivity

1. INTRODUCTION

Increasing health expenditures and a growing demand for health services have put an increasing focus on cost containment and the efficiency of delivering health services. However, the pressure to contain costs through enhanced efficiency may lead to poorer quality (Gutacker *et al*., 2013), which emphasizes the need for controlling for quality. Low quality could also be linked to high wasteful costs (McKay and Deily, 2008). Previous studies investigating the relationship between costs and quality show conflicting findings and use of heterogeneous methods and measures (Hussey *et al*., 2013). Hence, more knowledge on the association between provider costs and treatment quality is needed, and the use of cross-country comparisons gives opportunities to identify similarities and differences (Häkkinen *et al*., 2015).

An important issue in exploring the association between quality and costs is the choice of quality indicators. The indicators must reflect aspects that are of value to patients or society, which imply, in the nomenclature of

*Correspondence to: Frisch Centre, Gaustadalléen 21, NO-0349 Oslo, Norway. E-mail: sverre.kittelsen@frisch.uio.no

Donabedian (1966), that they should be at least process or structural quality indicators that are related to outcomes. To be useful at the hospital level, the indicators should be relevant for a non-negligible portion of the patients and be able to statistically distinguish between hospitals. The most interesting measures would be those that reflect medical quality by showing improved health, but from an economics point of view, measures of service quality could also be relevant if they reflect aspects of value to patients.

Several studies relate hospital costs to in-hospital or post-hospitalization mortality rates (Hussey *et al.*, 2013). In-hospital mortality used as the main quality indicator however poses some challenges. On the one hand, treatment costs could be low if the patient dies quickly after the admission. On the other hand, many resources are used for patients during their last days before death. This means that costs are endogenous to health outcomes, but these problems are less severe when mortality is measured regardless of death occurring in hospital or after discharge, as in this study. Alternative indicators may be based on complications (e.g., Kruse and Christensen, 2013), which however can be quite procedure specific and hence difficult to compare across medical specialities. Readmission rates encompass aspects of both medical and service quality.

The literature suggests that there will be a U-shaped relation between costs and quality, which for higher levels of quality means that there is a trade-off between cost containment and quality improvement, while for lower levels of quality, there may be a cost-saving potential of quality improvements (Hvenegaard *et al.*, 2009; Carey and Stefos, 2011; Gryna, 1999; Hvenegaard *et al.*, 2011). The intuition of the U-shaped relation would be that at lower quality levels, investments for improving quality may lower the net cost of treatment. Meanwhile, hospitals at higher levels of quality may operate on the upward sloping part where further investments may improve quality. If hospital service production is efficient, there will be a trade-off between quality and quantity or equivalently between costs and quality. All other things being equal, one cannot then increase the quality of treatment without incurring some opportunity costs such as reducing the number of patients treated or alternatively using more resources.

In empirical cross-section studies that compare hospitals, the relation will often be negative (e.g., Kruse and Christensen, 2013). This could be because of inadequate case-mix adjustment because some patients are inherently more prone to (costly) complications and readmissions and therefore have higher expected costs. Also, if the number of cases is small, there could be a large random component in the likelihood of complications. If case-mix adjustment is adequate and the number of cases is sufficient to disregard random variations, there remains the possibility of inefficiency. If it is possible to improve quality without increasing costs or reducing quantity, then the treatment is inefficient. On a more positive note, it is possible that a hospital that provides good quality may also be good at containing costs.

In their study of cost inefficiency and mortality in Florida hospitals, Deily and McKay (2006) isolated costs due to inefficiency and found a strong association to mortality. Their study applied individual level data in a stochastic frontier analysis. In a later study including a later sample of US acute hospitals, the authors found no systematic pattern of association between cost inefficiency and hospital outcome (McKay and Deily, 2008). Carey and Burgess (1999) found a positive relationship between costs and outpatient follow-up within 30 days after inpatient discharge for a sample of Veterans Administration (VA) hospitals in the USA. Fleming (1991) analyzed the cost and mortality/readmission relationship for Medicare beneficiaries hospitalized at 659 US hospitals and found that higher cost had a cubic association with the readmission index and surgical mortality index. Total and medical mortalities were not significantly associated with cost. Morey *et al.* (1992) used a national sample of 350 US hospitals to analyze the relationship between data envelopment analysis (DEA) scores and actual to predicted in-hospital deaths. They found that a reduction of one death was associated with an increase in efficient cost of $28,926. Mukamel *et al.* (2001) found a positive relationship between costs and risk-adjusted 30-day mortality after discharge for Medicare beneficiaries.

In a recent Canadian study, Stukel *et al.* (2012) found a positive association between costs and quality in a longitudinal analysis at patient level. They analyzed the association of hospital spending intensity and mortality and readmission rates for four common conditions (acute myocardial infarction (AMI), chronic heart failure, hip fracture, and colon cancer) in 129 hospitals in Ontario. This finding was confirmed by a German study, also at patient level, where they examined health outcomes (mortality at 30, 60, 90, and 365 days after discharge) for

AMI as a function of costs and other patient-level variables in 318 German hospitals (Stargardt *et al.*, 2014). Birkmeyer *et al.* (2012) examined the relationships between hospital outcomes (complication rates at inpatient surgery) and risk-adjusted, 30-day episode payments for four acute and elective procedures in US hospitals. It appeared that the complication rate was positively associated with Medicare payments, indicating a negative association between costs and quality. There was no statistical significant association between costs and mortality, however.

The survey by Hussey *et al.* (2013) attributed the divergent conclusions on the cost–quality association partly to differences in the unit of analysis (hospital, department, or patient group), measurement of costs and quality, as well as the adapted methodology. Hospital studies were slightly more likely to report a positive association between costs and quality than studies using other levels (such as nursing homes or areas) of analysis.

Studies under the EuroHOPE project have made advances in the comparison of healthcare costs between countries and relate the costs to outcomes and quality (e.g., Iversen *et al.*, 2015; Heijink *et al.*, 2015), but these studies look at a restricted set of diagnoses at a time. A recent study of the Organisation for Economic Cooperation and Development (OECD) countries analyzed the association between costs and efficiency for hospitals as a whole (Varabyova and Schreyögg, 2013). This article aims to expand such comparisons to include the quality of care as well, measured by selected case-mix adjusted quality variables. While this study relates to the EuroHOPE project, it includes only the four major Nordic countries (Norway, Sweden, Finland, and Denmark) in the comparison because only these countries have nationwide patient registers applicable for usage of the same hospital-wide case-mix (diagnosis-related group (DRG)) system. The homogenous definition of hospital outputs used in patient registers in the Nordic countries facilitates fair comparisons across countries.

Previous studies have indicated that Finnish hospitals have significantly higher average productivity than hospitals in Sweden, Denmark, and Norway and a substantial variation within each country (Kittelsen *et al.*, 2008; Linna *et al.*, 2010; Medin *et al.*, 2011; Kalseth *et al.*, 2011). Controlling for within-country variations in activity-based financing, length of stay (LOS), outpatient shares, university hospital status, or capital region only contributes to a small portion of these differences.

This paper examines whether quality differences can form part of the explanation for productivity differences and attempts to uncover any quality–cost trade-off at the hospital level. The analysis uses both individual patient-level and hospital-level data while taking cross-country differences into account. Auxiliary aims are to evaluate the usefulness of available quality indices and the importance of case-mix adjustments in these analyses. The pooling of data from four countries has at least two advantages. Firstly, we have a much larger sample size; and secondly, we are able to identify whether our findings are due to nation-specific or structural factors.

2. DATA

To perform the analysis in this study, we use data on hospital input and both quantitative and qualitative outputs. The productivity analysis utilizes a single input of hospital costs and three DRG-weighted outputs (medical inpatients, surgical inpatients, and outpatient visits) based on patient-level discharge registry data from 2008 to 2009. Individually identifiable patient data were not available in Norway before 2008. To calculate 365-day mortality, demographic data have been collected also for 2010. The Danish data are affected by the strike among hospital nurses in non-acute functions in 2008. Although one might expect a productivity penalty from the strike, both DRG production and costs would be reduced, and the impact on productivity should be minor. This section describes the hospital costs and patient-level discharge data sets, their sources and definitions, as well as the quality indicators and the case-mix adjustment variables used in the analysis (more details are available in Medin *et al.* (2013) and Anthun *et al.* (2013)). In the study, only somatic hospitals with a 24-hour emergency department or at least two medical or surgical specialities are included.

DOI: 10.1002/hec

2.1. Cost data

The hospitals costs include all production-related costs from the hospitals. Costs were harmonized across the countries by excluding costs for ambulances, value added tax (VAT), capital costs, purchased care, and costs for teaching and research.[1]

In Sweden, the cost data were assembled mainly from the Swedish Association of Local Authorities and Regions through the cost per patient database, from hospital annual reports, and from Statistics Sweden. The hospitals not recorded in these sources were sent a cost survey. For six Swedish counties, it was not possible to create data at the hospital level; so for these counties,[2] the output was also aggregated to the county level.

In Norway, the cost data were derived from the SAMDATA database of Norwegian specialized care published annually by the Directorate of Health. The National Institute for Health and Welfare in Finland collects hospital cost data annually as part of hospital productivity statistics production, while annual productivity reports published by the Ministry of Health contained the Danish cost data.

2.1.1. Cost level deflator. The collected cost data were measured in nominal prices in each country, and the costs were deflated to create real costs in each country. There were differences in currencies and input prices between the countries, and to allow for comparison between countries, the cost level had to be harmonized.

Wage indices were calculated for nine of the most important personnel groups. The wage indices were based on official wage data for the nine separate groups and included all personnel costs such as wage taxes and pension contributions (Anthun *et al*., 2013; Kittelsen *et al*., 2009; Medin *et al*., 2013). Personnel costs are the most important component with about 60% of total hospital costs. For the other costs, we use the purchasing power parity-adjusted gross domestic product price index from OECD. To form the aggregate cost level deflator, the nine personnel group indices and the index for other costs were weighed with fixed Norwegian shares for 2008, as personnel shares were not available for the other countries.

2.2. Patient-level data

Patient-level data were collected from national administrative patient registries in all four countries. The level of data was departmental (speciality) discharges. Outpatient visits registered during inpatient stays were excluded.

Death outside of hospitals was collected by linking patient-level data with other registries. In Norway, this linkage is automatically carried out in the patient registry through a link with the National Population Registry. The Danish patient data were manually linked with the Population Register. In Sweden and Finland, the time to death was collected by manually linking with the cause of death registries.

2.2.1. Diagnosis-related group grouping and weights. Norway, Sweden, and Finland each have a national version of a common grouping system for the hospital visits, NordDRG, developed at the Nordic Casemix Centre.[3] Denmark used to be part of NordDRG but changed to a national system DkDRG in 2002. The DkDRG system applies similar rules but is not completely comparable at the DRG level (Medin *et al*., 2013). Even though three of the countries have highly comparable systems, a common grouping is to be desired in order to enhance the comparability of the output measures and quality indices and to remove some of the idiosyncrasies inherent in each health system. All four countries have patient registers that use the same diagnosis and procedural classification systems, and Datawell Oy Finland has developed a common Nordic grouper for use in this and other projects based on definitions from the Nordic Casemix Centre. This grouper allows for similar grouper logic to be applied to all four Nordic countries. All patient-level data were regrouped in this grouper.

[1]Some additional costs were also excluded, details available in Anthun *et al*. (2013).
[2]Blekinge, Västmanland, Jämtland, Dalarna, Gävleborg, and Värmland. Kronoberg, Södermanland and Gotland have additionally been excluded from the productivity analysis because of problems in the cost data.
[3]http://www.nordcase.org/

DOI: 10.1002/hec

Common DRG weights are also needed to compare the countries. A set of cost weights were calculated from pooled 2008 and 2009 cost per patient data from Helsinki and Uusimaa hospital district in Finland grouped with the common Nordic grouper. As a robustness exercise, we have also calibrated weights for each of the Nordic DRGs using the average Swedish DRG weights of the Swedish patients assigned to that Nordic DRG.

Table I. Definitions of variables used

Group	Variable name	Definition
Quality indicators		
Readmissions	Readm30_Emergency	Patient admitted acutely to inpatient care in hospital within 30 days of the discharge
	Readm30_Inpatient	Patient admitted to inpatient care in hospital within 30 days of the discharge and at least two days after discharge
Mortality	Mort30_LastAdmittance Mort90_LastAdmittance Mort180_LastAdmittance Mort365_LastAdmittance	Out of hospital mortality from any cause. Dummies for 30, 90, 180 and 365 days after admission. Only set for last admission within the specified period.
Patient safety indicators		PSI indicators as defined by OECD
	PSI12_vt_pe	Pulmonary embolism/Deep vein thrombosis
	PSI13_Sepsis	Sepsis
	PSI15_AccidCutPunc	Accidental cut, puncture, or haemorrhage during medical care
	PSI18_ObstTrauma	Obstetric trauma
	BedSores	Bed-sores
Case-mix adjustment variables (used in models 0-5)		
Model 0: Nordic DRG	DRG	Diagnosis related group based upon common Nordic grouper
Model 1: + Patient characteristics		
	Male	1=Male, 0=Female
	Agegrp0 Agegrp1_9 … Agegrp80_89 Agegrp90	Age dummies for the groups: 0, 1-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90+
Model 2: + Treatment variables		
	TransInOwnHospital TransInOtherHospital TransOutOwnHospital TransOutOtherHospital	Dummies for transfer into and out of hospital department stay within one day before or after this stay. Not based on original coding but calculated from dates of patient registry directly.
	Charlson	Charlson index based upon secondary diagnosis
	NumSecDiagnoses	Number of secondary diagnoses
Model 3: + Length of stay	LOS	Length of stay defined as discharge date – admission date + 1
Model 4: + Municipal variables for patient	Population	Population of patient home municipality
	Unemployment	Unemployment rate as % of labour force
	SocialAssist	Social assistance recipients as % of population
	SingleFamilies	Single parent families, as % of all families with children
	Foreign	Citizens of foreign countries as % of population
Model 5: + Hospital-municipal variable	Traveltime	Travelling time by car in hours between hospital and centre of home municipality
Hospital level variables		
	Costs	Deflated real operating costs in common currency, corrected for differences in input price level between countries and years.
	Average Costs	Costs divided by total DRG-points
	NumberOfPatients	Number of departmental/speciality discharges
	Case-mix index (CMI)	Hospital DRG-points divided by number of patients
	UniversityHospital	Dummy if hospital is a teaching or university hospital
	CapitalCity	Dummy if hospital is located in the capital of each country

DOI: 10.1002/hec

Table II. Descriptive statistics for patient level variables

	Denmark	Finland	Norway	Sweden	All	
Number of observed discharges	15 753 686	12 395 963	11 124 765	18 884 433	58 158 847	
Variable	Mean	Mean	Mean	Mean	Mean	Std. Dev
Quality indicators						
Readm30_Emergency[a]	4.76 %	5.52 %	6.96 %		5.62 %	23.04 %
Readm30_Inpatient	4.95 %	12.67 %	13.84 %	9.99 %	9.93 %	29.91 %
Mort30_LastAdmittance	0.44 %	0.34 %	0.41 %	0.51 %	0.43 %	6.58 %
Mort90_LastAdmittance	0.54 %	0.43 %	0.53 %	0.68 %	0.56 %	7.47 %
Mort180_LastAdmittance	0.61 %	0.46 %	0.62 %	0.79 %	0.64 %	7.96 %
Mort365_LastAdmittance	0.72 %	0.49 %	0.74 %	0.96 %	0.75 %	8.66 %
PSI12_vt_pe	0.123 %	0.053 %	0.090 %	0.104 %	0.096 %	3.119 %
PSI13_Sepsis	0.076 %	0.044 %	0.078 %	0.077 %	0.070 %	2.667 %
PSI15_AccidCutPunc	0.005 %	0.005 %	0.024 %	0.014 %	0.012 %	1.083 %
PSI18_ObstTrauma	0.028 %	0.007 %	0.021 %	0.035 %	0.024 %	1.558 %
BedSores	0.015 %	0.005 %	0.031 %	0.028 %	0.020 %	1.434 %
Patient characteristics						
Male	43.01 %	44.80 %	45.58 %	45.31 %	44.63 %	49.71 %
Agegrp0	2.60 %	1.78 %	2.49 %	1.82 %	2.15 %	14.50 %
Agegrp1_9	4.13 %	6.22 %	6.11 %	6.72 %	5.80 %	23.37 %
Agegrp10_19	6.18 %	6.15 %	6.41 %	7.50 %	6.65 %	24.91 %
Agegrp20_29	9.34 %	8.72 %	9.35 %	8.79 %	9.03 %	28.66 %
Agegrp30_39	12.20 %	10.13 %	12.33 %	10.34 %	11.18 %	31.51 %
Agegrp40_49	11.23 %	11.21 %	11.27 %	10.14 %	10.88 %	31.14 %
Agegrp50_59	15.14 %	15.92 %	13.54 %	12.23 %	14.06 %	34.76 %
Agegrp60_69	18.36 %	17.32 %	16.30 %	16.73 %	17.21 %	37.75 %
Agegrp70_79	12.97 %	14.48 %	12.76 %	14.46 %	13.73 %	34.42 %
Agegrp80_89	6.84 %	7.29 %	8.25 %	9.80 %	8.17 %	27.39 %
Agegrp90	1.02 %	0.79 %	1.20 %	1.47 %	1.15 %	10.67 %
Treatment variables						
TransInOwnHospital	8.90 %	10.07 %	3.03 %	5.27 %	6.85 %	25.25 %
TransInOtherHospital	1.06 %	0.63 %	0.45 %	0.92 %	0.80 %	8.93 %
TransOutOwnHospital	6.18 %	9.39 %	3.16 %	5.00 %	5.90 %	23.57 %
TransOutOtherHospital	0.51 %	0.80 %	0.78 %	0.83 %	0.73 %	8.49 %
Charlson	0.113	0.047	0.265	0.196	0.155	0.665
NumSecDiagnoses	0.568	0.183	0.478	0.524	0.454	1.032
Length of stay						
LOS	1.568	1.457	1.669	1.678	1.599	3.519
Municipal variables						
Population	122 740	113 970	112 512	142 442	125 312	184 461
Unemployment	3.72	9.35	2.23	6.44	5.52	3.35
SocialAssist	1.38	6.73	2.53	4.52	3.76	2.46
SingleFamilies	10.89	20.33	19.87	21.02	17.91	5.48
Foreign	5.75	2.67	5.62	6.17	5.20	3.08
Hospital-municipal variable						
Traveltime	0.461	0.450	0.788	0.446	0.516	0.830

[a]Sweden lacks information on emergency status, therefore this variable only has 39 274 414 valid observations.

2.2.2. Quality indicators. We have calculated performance measures based on 11 quality indicators. Table I lists and defines the variables used in the analysis, and Table II gives descriptive statistics by country. All the indicators are binary variables at the patient level and are therefore presented as rates at the hospital or country levels.

Unlike planned readmissions, emergency readmissions within 30 days of a hospital discharge (but no sooner than the next day) are commonly viewed as a signal of poor medical quality if proper case-mix adjustment has taken place (Leng *et al.*, 1999). Only inpatients are included in this indicator as coding practice for outpatients

DOI: 10.1002/hec

varies between countries. Although some level of readmissions is unavoidable, an emergency readmission could be a sign that the initial treatment was not adequate or that the discharge was premature. We include emergency readmissions for any reason, because poor quality in the initial treatment (e.g., an operation) could well cause a readmission with another diagnosis (e.g., an infection). Country differences in the readmission rates in Table II are considerable, with Denmark at less than 5% and Norway at almost 7%. In Sweden, the coverage of the variable reflecting whether the admission is acute or planned is bad. As a substitute, we also included an indicator for all readmissions as an inpatient, regardless of emergency status. This is clearly more difficult to interpret as a sign of quality, as planned readmissions may be valid parts of a hospital treatment episode. However, in many cases, it will have a service quality dimension, because going in and out of hospitals is usually not appreciated by patients. Table II reveals that there is substantial variation in inpatient readmission patterns between countries.

Mortality rates[4] are the most widely accepted quality indicators. Even though some of the mortalities are unavoidable, lowering mortality is always an improvement. It has the additional advantage of being coded with little possibility of error. We have included four variants with different time perspectives, death within 30, 90, 180, and 365 days. There is a possibility of a person having several hospital stays within the last days of life, so the differential readmission patterns between countries would influence this indicator if the mortality was attributed fully to all hospital stays. We have therefore calculated a mortality dummy only if the stay is the last in the data before death, in order to attribute the death to this particular admission. In order to calculate 365-day mortality, we have collected patient data for the two years 2008 and 2009, and deaths also for 2010.

Patient safety indices (PSIs) are based on OECD standards using secondary diagnoses (Drösler, 2008). Most of these are applicable only to special patient groups, and Table II reveals very small raw rates, almost all less than a 10th of a percent. These also vary considerably between countries, with the Finnish numbers particularly low. The PSIs are based on secondary diagnoses, and we are aware of large differences in coding practices between countries. Secondary diagnoses are rarely reported in Finland, and the rate of PSIs is closely correlated with the reporting of secondary diagnoses (OECD, 2009). Thus, we cannot determine how much of the variation between countries is due to differences in quality and how much to coding, but within-country comparisons may still be valid.

Several other PSI definitions are available but could not be calculated from the available patient registers. Two more PSIs were initially included but turned out to be so infrequent that case-mix adjustments and hospital differences were meaningless. Numerous other quality indicators have been suggested and discarded, most because data were not available for several countries. In many cases, the data available for these indicators were not reliable. Time from referral to admission (‘waiting time’) could not be included because the definitions of referral date differed across countries and were not available at all for Sweden. Similarly, the time from admission to first procedure (‘lead time within hospital’) was not registered in Denmark and Sweden.

2.2.3. Case-mix adjusting variables. For the case-mix adjustment procedure, we have used most of the variables available in the patient registers. Ideally, the adjusting variables should capture characteristics of the patients and their illnesses that possibly influence the outcome, whatever the treatment given by the hospital. The primary risk adjustor used is the DRG formed with the common Nordic grouper. Because the division into the more than 700 DRGs is designed to capture most measurable patient differences that may influence costs, they will also capture many of the aspects that influence the expected values of the quality indicators.

The group of patient characteristics shown in Tables I and II comprises gender and age in 10-year groups, with a special infant group of less than 1 year. For data privacy reasons, the precise age was not available in the pooled cross-country dataset. Although partly endogenous, treatment variables are also allowed to adjust for risk, because these may reflect severity. The variables we coded for describing patient transfers in and out of hospital or department (where patient came from and where they went) do not distinguish between transfer

[4]We use the term ‘mortality rates’ rather than ‘case fatality rates’ because the latter are usually defined for a specific medical condition rather than for all hospital admissions.

DOI: 10.1002/hec

to/from home, a (non-hospital) health clinic, or a nursing home as we had to use information available in all four countries. Comorbidity is included both as the number of secondary diagnoses and as the Charlson comorbidity index, which in turn is based on information from secondary diagnoses (Charlson *et al*., 1987). LOS may reflect inefficiency in addition to severity (or even quality). LOS is also an endogenous variable to the hospital.

We have also included some characteristics for the patients' residence municipality in order to capture some socioeconomic differences. These variables are, however, not without challenges. Firstly, they are likely to be dependent between patients in each hospital, because most patients come from a limited number of municipalities in the hospital catchment area. In addition, they may to a large extent capture country effects, because there are marked differences between, for example, unemployment levels following the financial crisis. Finally, we have included travel time between the center of the residence municipality of each patient and the hospital,[5] a variable that previously has shown some explanatory power on hospital costs and that may have some also on quality outcomes (Kalseth *et al*., 2011).

3. METHODS

3.1. Case-mix adjustments

For the case-mix adjusted hospital performance measures, we follow Ash *et al*. (2003) and calculate the observed-to-expected ratio of each quality indicator for each hospital. The expected value, and thus the performance measure, is estimated in each of the six different models $m \in (0, \ldots, 5)$.

Each patient i has an (binary) observable quality indicator, ω_{ihk}, and an expected quality indicator, $\hat{\omega}^m_{ihk}$, subscripted by hospital $h \in (1, \ldots, H)$ and DRG $k \in (1, \ldots, K)$. We suppress an index for which indicator we are studying (see Table I for a list of all quality indicators).

The case-mix adjusted hospital performance measures, P^m_h, are calculated by summing all observed patient outcomes and dividing by the sum of all expected patient outcomes

$$P^m_h = \frac{\sum_{k=1}^{K} \sum_{i=1}^{N_{hk}} \omega_{ihk}}{\sum_{k=1}^{K} \sum_{i=1}^{N_{hk}} \hat{\omega}^m_{ihk}}, \tag{1}$$

where P^m_h is the performance indicator for hospital h in model $m \in (0, \ldots, 5)$ and N_{hk} is the number of patients in DRG k at hospital h. Because all our quality indicators are such that a lower number implies better quality, so will a lower value for the performance measure, P^m_h.

The performance measures P^m_h for $m \in (0, \ldots, 5)$ differ in the way we predict $\hat{\omega}^m_{ihk}$. In our simplest model, $m = 0$, we exploit that each hospital has a different composition of DRGs. The predicted quality indicator for patient i, $\hat{\omega}^0_{ihk}$, is thus just the average value of the quality indicator *within* each DRG for all patients across all hospitals. The predicted outcomes of this model can be written as

$$\hat{\omega}^0_{ihk} = \frac{\sum_{g=1}^{H} \sum_{j=1}^{N_{gk}} \omega_{jgk}}{\sum_{g=1}^{H} N_{gk}}, \tag{2}$$

which is independent of i and h and thus equal for all patients in DRG k.

The predicted quality measure, $\hat{\omega}^m_{ihk}$, can be further improved by conditioning on patient characteristics and municipality-specific variables. Because all our quality indicators in this study are binomial variables, the appropriate method is to estimate the conditional probability by the logit model (Greene, 2000; Hosmer *et al*.,

[5]Traveling times are calculated by Google maps using a STATA procedure from Ozimek and Miles (2011).

DOI: 10.1002/hec

2013). However, given the large number of observations, we need not assume that the explanatory variables have the same impact in all DRGs. Rather, *for each* DRG k, we calculate the expected value as the predicted value based on the maximum likelihood estimation of

$$\omega_{ihk}^{m} = \frac{e^{\beta_{0k}^{m} + \boldsymbol{\beta}_{k}^{m}\mathbf{z}_{ihk}^{m} + \varepsilon_{ihk}^{m}}}{1 - e^{\beta_{0k}^{m} + \boldsymbol{\beta}_{k}^{m}\mathbf{z}_{ihk}^{m} + \varepsilon_{ihk}^{m}}}, \tag{3}$$

where ω_{ihk}^{m} is the quality measure for patient i in DRG k at hospital h; the coefficient vectors $\beta_{0k}^{m}, \boldsymbol{\beta}_{k}^{m}$ are specific to each DRG k and model $m \in (1 \dots 5)$; $\mathbf{z}_{ihk}^{m}$ is a vector of individual case-mix adjusting variables; and ε_{ihk}^{m} is the error term, which is assumed to be normally distributed.

For each of the K DRGs and 11 indicators, we estimate five different models m, where higher-order models include more explanatory variables z (confer Table I for all case-mix adjusting variables). In model 1, the explanatory variables captured by z are the patient characteristics; in model 2, the vector includes both patient characteristics and the treatment variables; in model 3, the LOS is also added; model 4 includes also municipal characteristics for the patients' resident municipality; while model 5 adds the traveling time between the resident municipality and the treating hospital. The patients' predicted quality measures, $\hat{\omega}_{ihk}^{m}$, are calculated by setting the error term in Equation (3) to zero.

Following Moger and Peltola (2014), there are no hospital dummies in the estimation of Equation (3). Rather, the individual predicted values, $\hat{\omega}_{ihk}^{m}$, are inserted into (1) to calculate a hospital-specific performance index P_{h}^{m} for each quality indicator.[6]

Because our model is an aggregation of the estimates of a large number of non-linear equations, there are no obvious measures of model performance or goodness of fit. The extremely large sample size precludes the use of the Hosmer–Lemeshow (Hosmer and Lemeshow, 1980; Hosmer *et al*., 2013) test of goodness of fit. Instead, we use the Osius and Rojek (1992) normalization of the Pearson chi-squared statistic as outlined in Hosmer *et al*. (2013, p. 164). Following Greene (2000), we also calculate the R-squared based on the sum of the squared errors $\sum_{h=1}^{H}\sum_{k=1}^{K}\sum_{i=1}^{N_{hk}}\left(\omega_{ihk} - \hat{\omega}_{ihk}^{m}\right)^{2}$ to indicate the share of explained variance. Finally, we calculate and report the area under the curve (AUC) from receiver operating characteristic analysis. The AUC is commonly used for evaluating the ability of predictions from a logistic regression model in discriminating between outcomes and can be interpreted as the probability that, for example, the fatality prediction for a randomly selected patient who died is greater than the fatality prediction for a surviving patient.

3.2. Productivity estimates

The productivity estimates for the hospitals are based on Farrell (1957) who defined (the input oriented) *technical efficiency* as

$$E = \text{Min}\{\theta | (\theta\mathbf{x}_{i}, \mathbf{y}_{i}) \in T\}, \tag{4}$$

where $(\mathbf{x}_{i}, \mathbf{y}_{i})$ is the input/output vector for an observation i and T is the technology or production possibility, usually assumed to be specific to year and country. For an input/output vector $(\mathbf{x}, \mathbf{y})$ to be part of the production possibility set, we need to be able to produce $\mathbf{y}$ using $\mathbf{x}$. Efficiency is then the minimal proportionality factor θ on inputs that is consistent with feasibility, that is, what proportion of inputs are necessary to produce the given output vector.

Estimates of efficiency rely on estimates for the specific technology T; but for comparing productivities, only a common reference surface for observations is needed. The literature on Malmquist productivity indices

[6]Had we estimated a single logit, we could have included hospital fixed effects and estimated the performance measure in the first stage. Our setup with DRG-specific coefficients on the z-vector is equivalent to a single estimation with a full set of interaction terms. With 783 DRGs and up to 25 covariates, this would mean simultaneous estimation of up to 20,000 coefficients. Unfortunately, we have not had the necessary programs or machine power available.

DOI: 10.1002/hec

uses an (homogenous in inputs and outputs) envelopment of the technology to estimate changes in technical productivity over time (Førsund and Hjalmarsson, 1987; Grifell-Tatjé and Lovell, 1995). To compare the productivity of two or more observations, we do not need to estimate the separate country-specific and time-specific technologies but may instead rely on an estimate of the meta-frontier or the envelopment of the underlying technologies (Asmild and Tam, 2007). Productivity estimates of individual observations are then compared with this global measure of the highest attainable productivity. Here, we will estimate productivity of a hospital by calculating the reference set T from the pooled set of all hospitals across the Nordic countries and the two years 2008 and 2009 and then comparing individual hospitals to the reference set.

The estimates of the reference set, and therefore of the productivity of each hospital observation, are made using the homogenous version of the non-parametric DEA, one of the two main methods in the productivity literature (Coelli *et al*., 2005; Fried *et al*., 2008). This does not imply an assumption of constant returns to scale technology, because the reference frontier is only a homogenous envelopment of the underlying technology.[7] Because DEA estimates are known to be biased and the statistical properties are not available in closed form, bias-corrected estimates and confidence intervals have been calculated using the bootstrapping algorithm from Simar and Wilson (1998). The average cost per DRG point, which does not use a frontier method, is also calculated.

3.3. Productivity–quality trade-off

This article offers no full behavioral model of the relationship between productivity and quality. We start by calculating the hospital level pairwise Pearson correlation coefficient for each performance indicator, average costs (operating costs/DRG points), and productivity estimates. Additionally, we estimate a simple regression model with random hospital effects, assuming that the unobserved hospital heterogeneity is uncorrelated with the included variables (Greene, 2000)[8]

$$T\hat{P}_{hct} = \gamma_0 + \hat{\mathbf{P}}_{ht}\gamma_1 + \mathbf{x}_{hct}\gamma_2 + \lambda_c + \phi_{hct}, \tag{5}$$

where $T\hat{P}_{hct}$ is the DEA bootstrapped estimate of productivity for hospital h in country c in year t and $\hat{\mathbf{P}}_{ht}$ is a vector of the performance indicators estimated in model 2 (Table V). Note that for estimating the productivity–quality trade-off, we have calculated two performance measures for each hospital, one for each year, instead of pooling both years in a single hospital performance measure. $\mathbf{x}_{hct}$ is a vector of hospital-specific variables, including municipal variables averaged at the hospital level (Table VI); and λ_c contains country fixed effects.

Equation (5) is also estimated for each country separately thus leaving out the country-specific fixed effects.

4. RESULTS

4.1. Case-mix adjustments

Summarizing the more than 700 case-mix adjustment logit regressions for each of the 11 quality indicators is not straightforward. The R-squared statistics (shown in Table A.I of the Appendix) is rather low for all models. This is common in logistic regressions as the outcomes are 0 or 1, while the predictions are almost always between. Because of the large number of observations, R-squared for all models are significantly larger than zero; and for every quality indicator, adding a block of variables significantly increases the share of explained variance.[9] This test therefore gives no direct guidance on the model specification. We note, however, that the

[7]In Kittelsen *et al*. (2015), the productivity estimates are decomposed into scale efficiency, technical efficiency, and country-specific factors using variable returns to scale assumption on the underlying technology in each country. The results there are not sensitive to the use of DEA or the competing stochastic frontier analysis method.

[8]Simultaneous multi-level modeling of the case-mix adjustment is precluded by the computational intractability of the large number of coefficients, confer footnote 6. The large number (58 million) of patient observations would not in itself be a barrier, even though all calculations take much time.

[9]This is equivalent to the adjusted R-squared increasing as we move to larger models. In fact, the numbers for R-squared and adjusted R-squared are not distinguishable with the number of decimals reported in the table.

DOI: 10.1002/hec

DRGs in model 0 explain at most 9% of the variance of the quality indicators. Adding patient characteristics as is performed in model 1 does not change this pattern and hardly adds any explanatory power. Adding the treatment variables in model 2 and LOS in model 3 increases R-squared somewhat for readmission rates, mortality rates, and for PSI13 (sepsis). The municipal and travel time variables of models 4 and 5, respectively, only slightly increase R-squared.

The normalized Pearson goodness-of-fit test (shown in Table A.II of the Appendix) fails to reject the large majority of models. Model 1 is rejected for some of the quality indices and for PSI18 (Obstetric trauma) also, model 2 is rejected; but here, it seems that the problem is that PSI18 only applies to women in specific DRGs and age groups.

Table III. Area under the curve (AUC) based on a 0.1% sample of discharges, using predictions from the full-sample DRG-specific case-mix adjustment regression models with quality indicators as dependent variables.

Model	0	1	2	3	4	5
Cummulative included independent variables	DRGs	+Patient characteristics	+Treatment variables	+Length of stay	+Municipal variables	+Travel time
Dependent variable						
Readm30_Emergency	0.72***	0.73*	0.75***	0.75	0.77***	0.77
Readm30_Inpatient	0.71***	0.73***	0.74***	0.75	0.78***	0.78
Mort30_LastAdmittance	0.92***	0.95***	0.96	0.96	0.96	0.96
Mort90_LastAdmittance	0.91***	0.94***	0.95*	0.95	0.96	0.96
Mort180_LastAdmittance	0.89***	0.94***	0.95*	0.95	0.95	0.95
Mort365_LastAdmittance	0.86***	0.92***	0.93*	0.93	0.94	0.94
PSI12_vt_pe	0.86***	0.88	0.95***	0.95	0.95	0.95
PSI13_Sepsis	0.95***	0.96	0.98***	0.98	0.99	0.99
PSI15_AccidCutPunc	0.95***	0.97	0.99***	0.99	0.99	0.99
PSI18_ObstTrauma	0.98***	0.99	1.00***	1.00	1.00	1.00
BedSores	0.97***	0.98	0.99***	0.99	0.99	0.99

Model *m* includes all variables from model *m-1*. AUC estimates for model *m* that are significantly higher than that of model *m-1* are marked at *0.10; ** 0.05; ***0.01 level. The AUC and the corresponding confidence intervals are estimated using the roctab procedure in Stata 13. In the full sample there are 58 158 847 observations except for Readm30_Emergency which is not registered in Sweden and therefore has only 39 274 414 observations.

More interesting are probably the AUC results shown in Table III.[10] The ability to discriminate between outcomes is very high for all mortality and PSI indicators. For the mortality indicators, the inclusion of patient characteristics significantly increases the AUC estimates and weakly so does the inclusion of treatment variables. For the PSIs, patient characteristics do not seem to matter but treatment variables do. LOS, municipal variables, and travel time do not contribute for these quality indicators. The readmission variables have a clearly different pattern, with lower but considerable AUCs in all models. Here, the inclusion of patient characteristics and treatment variables is significant, as well as the municipal variables. It must be noted, however, that there are large country differences in some of the municipal variables, for example, the number of foreign citizens and the unemployment rates in the wake of the financial crisis.

The statistical evidence seems to favor model 2, with some exceptions. The purpose of these models is to level the field in country and hospital comparisons. The choice of case-mix adjustment model specification must therefore also take account of the problems of country effects in the municipal variables. In addition, the LOS is to a large extent an endogenous variable for the hospital in question and may be more of a mediating than confounding variable. In the further analysis, we therefore use model 2, that is, the model without LOS, the municipal variables, and travel time, returning to these only in the hospital trade-off regressions.

[10]STATA 13 was not able to calculate AUC based on the extremely large samples, so we report AUC results for a 0.1% random subsample stratified on hospitals with 58,159 patients (39,274 patients for Readm30_Emergency).

Table IV. Country means of case-mix adjusted performance measures (model 2) with 99 % confidence intervals.

	Denmark	Finland	Norway	Sweden
Readm30_Emergency	0.891	1.031	1.103	-
	(0.888 - 0.893)	(1.028 - 1.034)	(1.099 - 1.106)	-
Readm30_Inpatient	0.573	1.235	1.256	0.986
	(0.572 - 0.575)	(1.232 - 1.237)	(1.253 - 1.258)	(0.984 - 0.988)
Mort30_LastAdmittance	0.927	1.037	0.751	1.011
	(0.918 - 0.936)	(1.024 - 1.050)	(0.741 - 0.760)	(1.002 - 1.019)
Mort90_LastAdmittance	0.909	1.043	0.785	1.052
	(0.901 - 0.917)	(1.031 - 1.055)	(0.776 - 0.794)	(1.044 - 1.060)
Mort180_LastAdmittance	0.907	0.989	0.808	1.071
	(0.900 - 0.915)	(0.978 - 0.999)	(0.800 - 0.817)	(1.064 - 1.079)
Mort365_LastAdmittance	0.918	0.877	0.840	1.101
	(0.911 - 0.925)	(0.868 - 0.886)	(0.832 - 0.848)	(1.095 - 1.108)
PSI12_vt_pe	1.153	0.870	0.763	0.992
	(1.131 - 1.174)	(0.842 - 0.898)	(0.742 - 0.783)	(0.974 - 1.011)
PSI13_Sepsis	1.319	1.081	0.718	0.967
	(1.288 - 1.350)	(1.043 - 1.119)	(0.698 - 0.739)	(0.946 - 0.988)
PSI15_AccidCutPunc	0.459	0.681	1.145	0.934
	(0.418 - 0.500)	(0.615 - 0.749)	(1.085 - 1.205)	(0.886 - 0.983)
PSI18_ObstTrauma	0.917	0.393	0.727	1.529
	(0.881 - 0.953)	(0.358 - 0.429)	(0.687 - 0.768)	(1.480 - 1.579)
BedSores	0.752	0.433	1.015	0.992
	(0.713 - 0.791)	(0.389 - 0.478)	(0.969 - 1.062)	(0.956 - 1.027)

4.2. Country and hospital differences

The case-mix adjustments (in model 2) change the relative performance of the countries to some extent. Table IV gives the mean performance measure at the country level, with a 99% confidence interval calculated from the individual patients' predicted values. By construction, each performance measure has a mean of 1.0 when averaging over all four Nordic countries, rendering almost all country-specific performance measures significantly different from the Nordic mean. As the quality measures used are by definition 'measures of low quality', lower performance measures denote higher quality.

The quality measures in Table IV do not give a uniform picture of the quality of care in any of the Nordic countries. Neither do they indicate any clear ranking of the countries. While Denmark has clearly fewer emergency readmissions, Norway has the lowest mortality rates. The inpatient readmission rates, on the other hand, are higher in Norway and Finland than in Denmark and Sweden. PSI12 (pulmonary/deep vein thrombosis) and PSI13 (sepsis) are the lowest for Norway; PSI15 (accidental cut, puncture, or haemorrhage during medical care) is the lowest in Denmark; while Finland has the lowest score for PSI18 (obstetric trauma) and bed sores.

Table V. Hospital differences in case-mix adjusted performance measures (model 2)

	Share of hospitals with performance measure significantly different from 1 at 95% level					ANOVA
	Denmark	Finland	Norway	Sweden	Total	F
Readm30_Emergency	89 %	91 %	81 %		86 %	774.4***
Readm30_Inpatient	100 %	97 %	87 %	87 %	91 %	3635.3***
Mort30_LastAdmittance	68 %	59 %	85 %	64 %	70 %	82.6***
Mort90_LastAdmittance	79 %	66 %	81 %	85 %	79 %	96.6***
Mort180_LastAdmittance	75 %	53 %	72 %	85 %	73 %	98.9***
Mort365_LastAdmittance	86 %	66 %	74 %	81 %	77 %	112.2***
PSI12_vt_pe	75 %	53 %	74 %	51 %	63 %	63.9***
PSI13_Sepsis	57 %	41 %	74 %	62 %	61 %	55.4***
PSI15_AccidCutPunc	85 %	34 %	30 %	32 %	41 %	13.3***
PSI18_ObstTrauma	57 %	81 %	48 %	62 %	60 %	36.2***
BedSores	54 %	80 %	39 %	43 %	51 %	19.7***
Number of hospitals	28	32	47	53	160	160

ANOVA tests for differences in hospital performance and the significance of the F-values are marked at *0.10; ** 0.05; ***0.01 level.

Hospital differences are difficult to summarize, but Table V shows the percentage of hospitals across both years with performance measures significantly different from the Nordic mean of 1. This holds for almost all of the readmission variables, and for a large majority of the mortality rates, but to a lesser and mixed extent for the PSIs. For readmission variables, Denmark and Finland have the largest shares of hospitals with performance measures different from the Nordic mean. Sweden has the largest share of hospitals with significantly different means in two mortality measures. For mortality within 30 days, Norway has the largest share of hospitals with significantly different means. The last column of Table V shows the significance of the hospitals in explaining the variation remaining after the case-mix adjustment of model 2, based on a linear ANOVA test of the difference between observed and predicted values (Greene, 2000). The results show that hospitals are significantly different from each other in their performance measures for all quality indicators. Given the very large number of patient observations, the *F*-values are not particularly high for the mortality indicators and definitely weak for the PSIs.

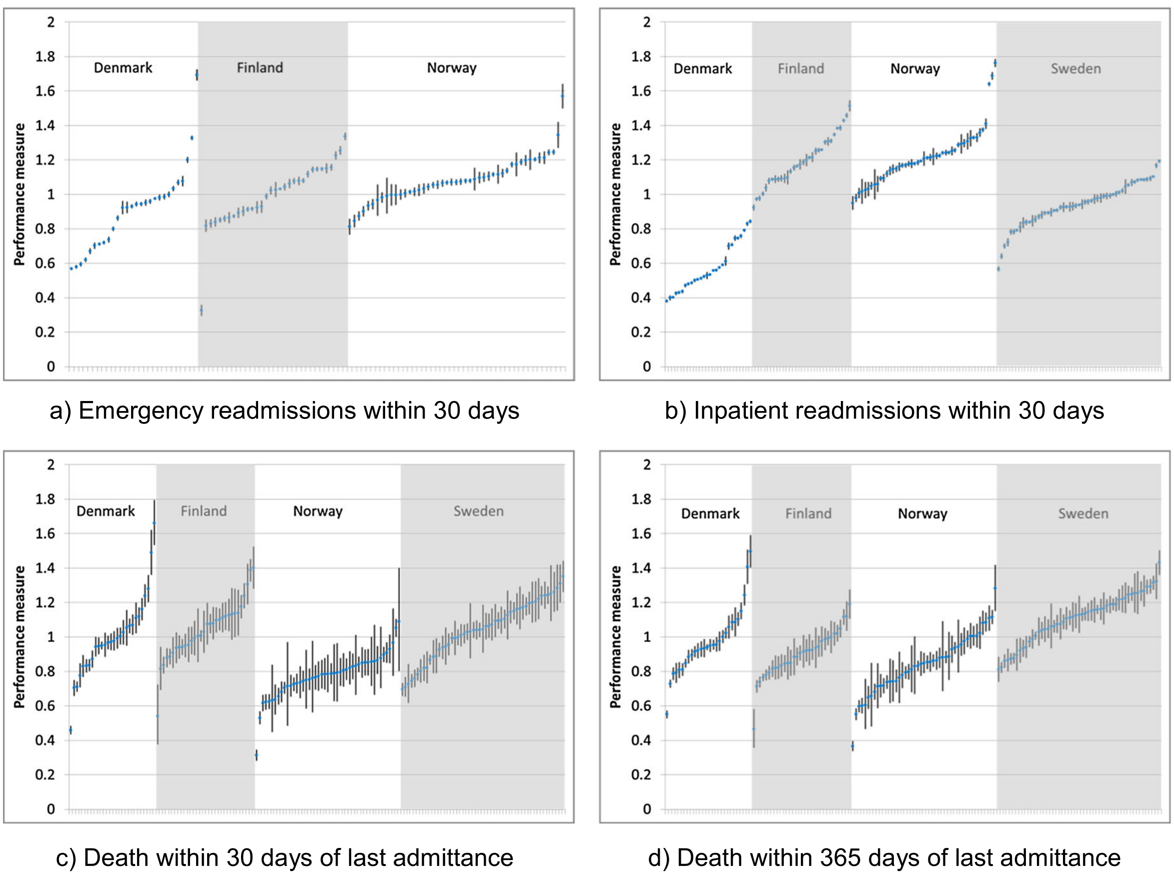

Figure 1. Selected case-mix adjusted performance measures for hospitals sorted by country, with 99% confidence intervals. Lower numbers indicate better quality.

Figure 1 plots four of the performance measures and their 99% confidence intervals for the individual hospitals sorted by countries. For emergency readmissions (panel a), the confidence intervals are very narrow, which means that there are significant differences between most hospitals. There is mostly a clear ranking of hospitals within countries, because each hospital performance measure is mainly outside the range of other hospitals' confidence intervals. As noted, Denmark has the lowest emergency readmission rates, but there is

DOI: 10.1002/hec

some overlap with the Finnish and Norwegian hospitals. It was not possible to compile this indicator for Sweden. Inpatient readmissions (panel b) show even greater differences, with all Danish hospitals having significantly lower rates than all Finnish and Norwegian hospitals. The rates of Swedish hospitals fall mostly between Danish and both Finnish and Norwegian hospitals.

For 30-day or 365-day mortality, the confidence intervals are wider, but most hospitals are still significantly different from the mean and from each other. Most Norwegian hospitals have significantly lower 30-day mortality than hospitals in the other countries, but these differences are less marked when comparing 365-day mortality (panels c and d, respectively).

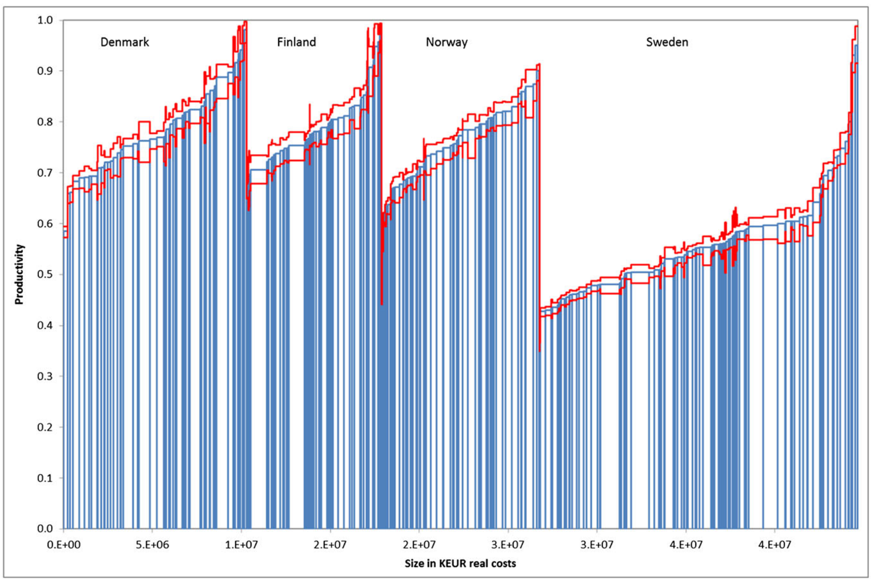

Figure 2. Salter diagram of bootstrapped DEA hospital productivity estimates sorted by country with 95% confidence intervals. The width of each column is proportional to hospital size measured by real costs.

4.3. Productivity

To look into the possible trade-off between hospital productivity and quality, we first had to estimate hospital productivity. As noted in 2.2., we here use a common Nordic version of the NordDRG grouper, which makes it possible to compile hospital output measures that are comparable between countries. Figure 2 shows the bias-corrected DEA productivity estimates of the hospitals sorted by country and productivity levels, with the width of the bars proportionate to hospital costs. Bootstrapped 95% confidence intervals are also shown.

The figure confirms the previous results that Finnish hospitals are on average more productive than in the other Nordic countries, even though Denmark is almost as productive (Medin *et al*., 2011; Kittelsen *et al*., 2008; Linna *et al*., 2010; Kittelsen *et al*., 2015). Even Norway has not much of a cost disadvantage in this analysis, a clear catching up from previous studies. Sweden, however, still lags behind. As a first robustness test, average costs (real costs per DRG point) have also been calculated and show essentially the same picture with a correlation of −88.6%. We have also recalculated the DRG points using calibrated Swedish DRG weights and results are again very similar, with a correlation between productivity estimates of 90.2%. Table A.III in the Appendix shows the mean hospital inputs, outputs, and productivity estimates for each country.

DOI: 10.1002/hec

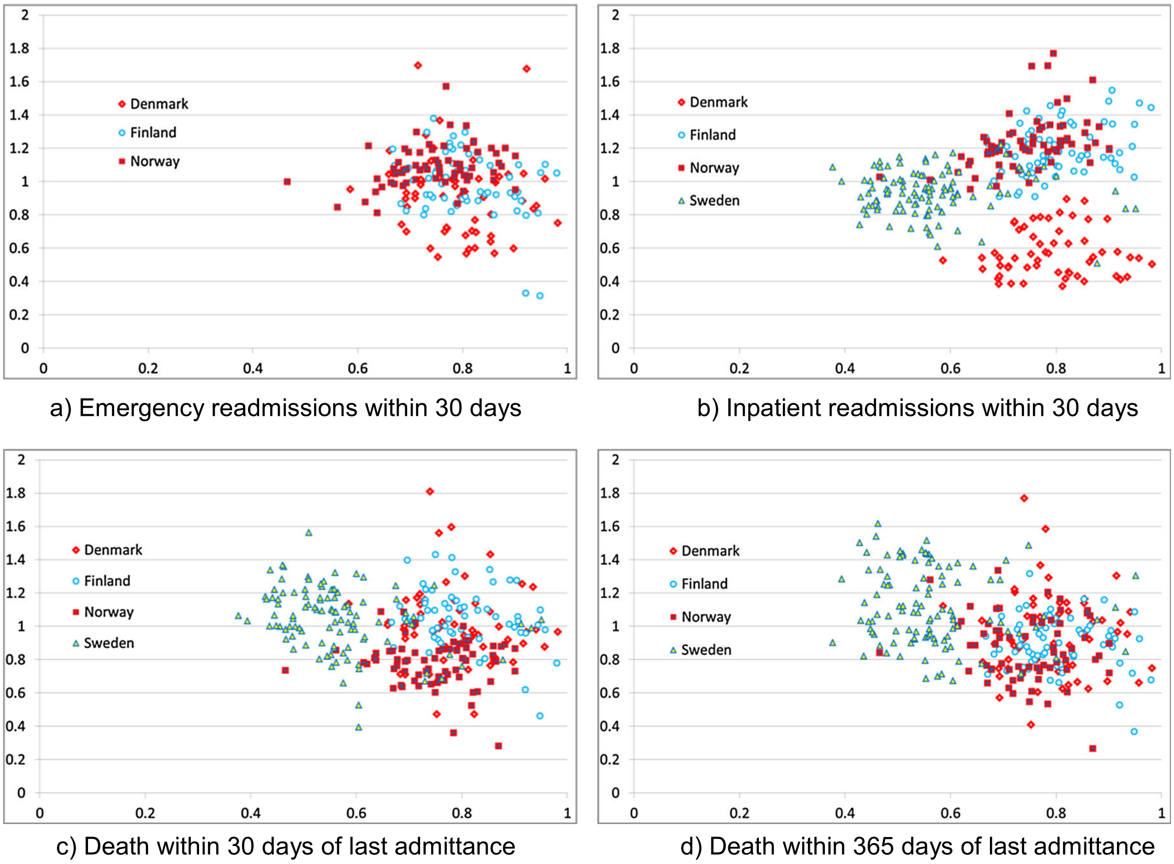

Figure 3. Selected case-mix adjusted performance measures for hospitals (vertical) plotted against estimated productivity (horizontal). Better joint performance is low performance measure and high productivity (lower right).

4.4. Productivity–quality trade-off

When productivity estimates are plotted against four of the performance measures in Figure 3, one finds no strong correlations. In all panels, the optimal frontier would be at the lower right with the highest productivity and the lowest performance measure, keeping in mind that a low performance measure indicates higher quality. In panel a, there is a slightly negative correlation ($r = -0.155$) between productivity and emergency readmissions, implying no trade-off between high quality and high productivity. There is a slight tendency for low emergency readmission rates to go together with high productivity in Finland, but the main impression is a large dispersion. Panel b shows some positive correlation, implying that having high productivity goes together with high number of inpatient readmissions. There seems to be a trade-off between quality and productivity but only in so far as the inpatient readmission rate is a valid quality indicator.

For the two mortality rates shown in Figure 3, mortality within 30 days and mortality within 365 days of hospital episode, there is a clear negative correlation between productivity and performance measures, which is strongest for 365-day mortality. This would imply that there is no trade-off between productivity and quality, and it is possible to improve both productivity and quality at the same time.

The pairwise correlations between measures of productivity and quality in Figure 3 are reported in the first column of Table VI, which draws on the full correlation matrix in Table A.IV in the Appendix.

DOI: 10.1002/hec

Table VI. Productivity-performance correlations and trade-off regressions (GLS random hospital effects models)

	Pairwise correlations with estimated productivity	Linear regression on dependent variable estimated productivity					
		I. All countries	II. Without Sweden	III. Denmark	IV. Finland	V. Norway	VI. Sweden
Constant		0.938 ***	0.951 ***	1.311 ***	0.852 ***	0.688 ***	0.748 ***
		(0.092)	(0.102)	(0.328)	(0.159)	(0.157)	(0.192)
Performance measures							
Readm30_Emergency	−0.155 ***		−0.002	0.110	−0.143 **	0.042	
			(0.038)	(0.082)	(0.066)	(0.065)	
Readm30_Inpatient	0.121**	0.067 *	0.130 ***	0.023	0.196 ***	0.148 **	−0.088
		(0.040)	(0.043)	(0.120)	(0.076)	(0.072)	(0.097)
Mort30_LastDischarge	−0.233***	−0.110***	−0.149 ***	−0.307***	−0.066	−0.092	−0.040
		(0.028)	(0.036)	(0.076)	(0.058)	(0.086)	(0.046)
Hospital variables							
Number of patients	0.095	−1.73E-08	−3.48E-08	−1.66E-07	−1.36E-07	2.06E-07*	−1.17E-07
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
UniversityHospital	0.130 **	0.013	0.020	0.012	−0.006	−0.034	0.095
		(0.023)	(0.024)	(0.102)	(0.058)	(0.040)	(0.061)
CapitalCity	0.102 *	−0.009	−0.008	0.108	−0.085	−0.027	0.151
		(0.040)	(0.044)	(0.152)	(0.207)	(0.081)	(0.087)
Hospital average of municipal variables							
Population	0.058	2.01E-08	3.38E-09	8.64E-08	9.64E-07	3.58E-08	1.92E-09
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Unemployment	−0.122 **	0.002	−0.001	0.003	−0.003	−0.013	0.004
		(0.002)	(0.003)	(0.008)	(0.004)	(0.011)	(0.004)
SocialAssist	−0.112 *	−0.003	−0.002	0.009	0.012	−0.030	−0.005
		(0.009)	(0.012)	(0.078)	(0.014)	(0.042)	(0.014)
SingleFamilies	−0.360***	−0.004	−0.004	−0.019	−0.002	0.003	−0.004
		(0.004)	(0.004)	(0.028)	(0.007)	(0.007)	(0.008)
Foreign	−0.167***	0.000	−0.010	−0.031	−0.039 **	−0.005	0.006
		(0.006)	(0.007)	(0.029)	(0.019)	(0.010)	(0.010)
Traveltime	−0.079	−0.064***	−0.073 ***	0.094	−0.057	−0.055 **	−0.057
		(0.025)	(0.023)	(0.207)	(0.092)	(0.025)	(0.087)
Country dummies							
Denmark		−0.014	0.056				
		(0.062)	(0.067)				
Norway		−0.069	−0.057				
		(0.044)	(0.047)				
Sweden		−0.214***					
		(0.035)					
R-squared							
Within		0.054	0.010	0.023	0.058	0.060	0.036
Between		0.003	0.046	0.033	0.017	0.042	0.097
Overall		0.012	0.081	0.063	0.074	0.096	0.079
Number of observations		292	186	56	64	66	106

Standard errors in (). Significant coefficients are marked at * 0.10; ** 0.05; *** 0.01 level. Hausman tests reject random effects for model I (X^2=25.7), but accepts random effects for model II (X^2=11.9).

DOI: 10.1002/hec

The next columns of Table VI show the results of the random effects regressions on the bias-corrected productivity estimate with the main performance measures at the hospital level as explanatory variables (Equation (5)). For reasons of collinearity, only one mortality measure, mortality within 30 days, is included. The results for the PSIs are also deemed too weak to be valid as measures of quality for hospitals as a whole and therefore not included as explanatory variables. Column I excludes the emergency readmissions, because Sweden has no observations on this measure, while column II instead excludes the Swedish observations. The last four columns show regressions for each country separately. The country-specific models have generally less explanatory power due to the low number of observations.

The negative pairwise correlation between productivity and emergency readmissions disappears when modeled simultaneously with the other performance measures and the control variables. Looking at individual countries, only in Finland is there a significant negative coefficient, indicating no trade-off between productivity and emergency readmission rates. In Denmark and Norway, the coefficients are positive but insignificant. The lack of significance in some of the country-specific associations may partly reflect the low number of observations. For inpatient readmission, the positive correlation, indicating a trade-off, carries through to the regression coefficients, with high inpatient readmission rates being associated with high productivity, particularly in Finland and Norway.

As in Figure 3, there seems to be some association of high quality and high productivity when using 30-day mortality as a quality measure on a pooled dataset with all countries. However, in the country-specific regressions, the association is only significant for Denmark.

Of the controls, Sweden has significantly lower productivity, while the Norway and Denmark dummies are not significant. In common with previous studies (Kalseth *et al*., 2011), one finds a negative association between productivity and the traveling time, which seems to be influenced mainly by the Norwegian observations. No other controls are significant.

5. DISCUSSION

Our study confirms the productivity differences from previous studies with a similar ranking of the countries where Finnish hospitals have the highest productivity estimates followed by Denmark and Norway and last Sweden. Overall, there is a weak pattern that Norway and Denmark show higher performance in quality and Sweden lower performance; thus, there is no trade-off between productivity and quality at the country level. The distance between hospitals in Sweden and the other countries is even larger when taking quality aspects into account, although the confidence intervals for several indicators are overlapping, which makes ranking of hospitals within each country difficult.

Case-mix adjustment of quality indicators is important; and in some cases, the ranking of countries changed as a result of the adjustment. Our result shows that there are major and significant differences in the included case-mix adjusted quality indicators, especially across countries. One example is inpatient readmission within 30 days where Denmark has less than half compared with the other countries. Also, mortality within 30 days after last admittance varies considerably with Norway having the lowest rates and Finland the highest. The ranking of the countries changes for the longer periods up to 365 days after admission, where Finland has lower mortality rates than both Sweden and Denmark. The case-mix adjusted performance measures for hospitals show larger differences for readmissions than for mortality after last admission. The smaller variation for mortality is in line with other studies about mortality differences. In Sweden and Norway, 30-day mortality of AMI was lower than that in Finland; but for stroke and hip fracture, there was no difference between the three countries (Heijink *et al*., 2015; Häkkinen *et al*., 2015; Hagen *et al*., 2015; Medin *et al*., 2015; Peltola *et al*., 2015). Other quality indicators largely confirm these country level differences and are based on a low frequency of events, and some PSIs are not usable for quality comparisons at the country level because of poor or divergent coding. PSIs may still reveal important differences between and within countries with lower general standard of health care, such as in some developing countries. For most quality indicators, the performance measures reveal room for improvement in Nordic hospitals.

The hospital variables do not contribute to the explanation of the differences in productivity. The number of patients, as a proxy for size, is not significant nor is university hospitals or capital city. A number of municipal variables were tested, and only travel time was associated with higher costs and then only in Norway.

Regarding the cost–quality trade-off, there is a statistical significant negative relationship between hospital-level productivity and mortality within 30 days. Assuming that the case-mix adjustment is adequate, the driving mechanism seems to be that treating dying patients is costly for the hospitals, and the efficient hospitals are those that are better at preventing mortality. This relationship is valid for all countries, but in the analysis for each country, it is only significant for Denmark. Hospitals with higher inpatient readmissions within 30 days have a tendency for also having higher productivity. This relationship is significant for Finland and Norway indicating a productivity–quality trade-off.

The conclusion from the review by Hussey *et al.* (2013) was that evidence of the direct association between productivity and quality is inconsistent but that the association is small to moderate. Of these, six studies used similar outcome indicators as in our study (Barnato *et al.*, 2010; Bradbury *et al.*, 1994; Carey and Burgess, 1999; Deily and McKay 2006; McKay and Deily, 2008; Fleming, 1991). Among these studies, three indicate a clear negative relationship. Compared with our study, these studies include several limitations. Outcome is often measured using in-hospital mortality, or data are restricted to Medicare patients (over 65). In addition, the cost measures in the US studies usually exclude the costs for the physician, whereas in our study, the wages of physicians are included.

In some of the recent studies referenced in the introduction, the outcome measures are similar to the present study. Stukel *et al.* (2012) found that higher hospital spending intensity was associated with better survival and lower admission rates. Stargardt *et al.* (2014) confirmed the trade-off between costs and outcomes, estimating that an increase of costs by €100 leads to a reduction in mortality risk by 0.4%. Doyle *et al.* (2015) also found that patients brought to a higher cost hospital have lower mortality. These studies show a different result to ours by finding a productivity–quality trade-off. The US study at hospital level by Birkmeyer *et al.* (2012) came to a somewhat different conclusion. The study found a strong positive correlation between complication rates and episode payments, indicating that efforts to improve surgical quality may reduce costs. This study differed in the cost estimation as the time window was extended to 30 days after discharge, which could explain a different correlation pattern where low quality leads to high costs after discharge.

Controlling sufficiently for patient–case mix is a major concern and may be a limitation of our study. The most important variables included are the DRGs formed with the common Nordic grouper, age, gender, hospital transfers, and comorbidities. Still, as coding practices differ across the countries (especially in the reporting of secondary diagnoses), true differences in risk factors at the patient level may not be sufficiently captured. There is a need for improvement in harmonization in coding across countries.

The use of average cost to weigh the patient discharges in different DRGs does not necessarily reflect the social willingness to pay for the different treatment groups. Basing these weights on average costs in two Finnish municipalities poses additional problems if these weights then reflect costs or incentives that are particular for Finland. However, using calibrated Swedish weights showed results to be quite robust, and previous studies that exploit the variation in the use of activity-based financing in Nordic hospitals have found little effect on productivity (Kittelsen *et al.*, 2008). Still, there is a clear need for further research on the effect of different weight sets for the measurement of productivity.

DOI: 10.1002/hec

Another limitation of this study is that health outcomes and productivity measures are reported at the hospital level, which may conceal differences in outcomes across medical conditions. In a recent European study (Häkkinen *et al.*, 2015), the results indicated that there was no correlation either at the national or at the hospital level, between the quality in treatment of two different acute conditions (stroke and AMI). The results indicate that the quality of treatment for one specific health problem (disease) cannot be used as a proxy for hospital level overall quality of care.

Finland treats patients with incurable diseases to a larger extent outside acute hospitals, which may have an impact on hospital mortality figures. Still this does not explain the differences in-hospital mortality between Norway and Sweden, and some of these effects may have been controlled for by the use of age groups in the case-mix adjustments.

Finally, it is an important limitation that causal inferences cannot be drawn from the regressions; instead, these indicate the strength of statistical associations. In particular, the productivity–quality trade-off regressions are not set in a behavioral model with adequately specified econometric structure.

6. CONCLUSION

The results show that there are significant differences between countries on most measured quality indicators. Case-mix adjustments are necessary but explain only a minor portion of quality variation. There are significant differences also between hospitals within countries, but only the readmission and mortality measures show enough differences to rank the majority of hospitals. For PSIs, the confidence intervals overlap too much for rankings to be meaningful. The PSI events are too infrequent in the Nordic countries to discriminate between chance and true hospital or country differences, and are generally prone to be invalid for country comparisons due to differences in coding practices. This highlights the need for continuous improvements in the harmonization of coding systems and patient registry information. At this point in time, only some patient registry-based quality indicators are useable for international comparisons, especially if one looks beyond the Nordic countries. If individual hospital managers are to learn from other hospitals, and national policy makers are to learn from other countries, comparable data must be provided. This does not necessarily imply use of common DRG systems or incentives but that the underlying diagnosis, procedure, and case-mix adjustment codes have the same content.

While previous findings on the relative productivity of the hospitals in the Nordic countries are confirmed, there is no clear pattern that any country has higher or lower quality on all measures. This may be due to the limitations of the available data as discussed earlier. This may also be due to that the treatment patterns and practices vary a lot between countries, even for countries that are as similar as Denmark, Finland, Norway, and Sweden. This is consistent with previous findings that efficiencies are similar across countries but that there are country-specific factors that make the production possibilities significantly different (Kittelsen *et al.*, 2015). Unfortunately, statistical methods have difficulty in identifying the effect of country-specific factors with only four countries. Again, the use of data from countries outside the Nordic region could give a better foundation for general results.

The evidence for a trade-off or a positive association between quality and productivity varies between the different performance measures. There seems to be a trade-off between productivity and better (lower) inpatient readmission rates, but high productivity is associated with lower mortality rates. Policies aimed at reducing readmission rates may be costly. There may be differences between emergency and planned readmissions in this regard, but it is important to look into how incentives for readmissions vary between countries. Policies aimed at decreasing mortality rates may reduce costs and increase productivity at the hospital level. For mortality at least, there seems to be a possibility of improving both quality and productivity.

DOI: 10.1002/hec

APPENDIX I

Table A.I R-squared for case-mix adjustment regressions of quality indicators

Model	0	1	2	3	4	5
Cummulative included independent variables	DRGs	+Patient characte-ristics	+Treatment variables	+Length of stay	+Municipal variables	+Travel time
Dependent variable						
Readm30_Emergency	0.05***	0.05***	0.08***	0.08***	0.09***	0.09***
Readm30_Inpatient	0.08***	0.09***	0.11***	0.11***	0.15***	0.15 ***
Mort30_LastAdmittance	0.09***	0.12***	0.14***	0.15***	0.16***	0.16***
Mort90_LastAdmittance	0.09***	0.12***	0.15***	0.16***	0.17***	0.17***
Mort180_LastAdmittance	0.09***	0.12***	0.15***	0.16***	0.16***	0.16***
Mort365_LastAdmittance	0.08***	0.11***	0.14***	0.14***	0.15***	0.15***
PSI12_vt_pe	0.01***	0.01***	0.01***	0.01***	0.02***	0.02***
PSI13_Sepsis	0.02***	0.03***	0.06***	0.07***	0.08***	0.08***
PSI15_AccidCutPunc	0.01***	0.02***	0.04***	0.04***	0.06***	0.07***
PSI18_ObstTrauma	0.10***	0.10***	0.11***	0.11***	0.11***	0.11***
BedSores	0.00***	0.01***	0.02***	0.02***	0.04***	0.04***

Model *m* includes all variables from model *m-1*. R^2 estimates for model *m* that are significantly higher than that of model *m-1* are marked at * 0.10; ** 0.05; *** 0.01 level. Calculations have been performed using Stata 13 matrix commands. There are 58 158 847 observations except for Readm30_Emergency which is not registered in Sweden and therefore has only 39 274 414 observations.

Table A.II Normalized Pearson chi-squared statistics (Z-values) for case-mix adjustment regressions of quality indicators

Model	0	1	2	3	4	5
Cummulative included independent variables	DRGs	+Patient characte-ristics	+Treatment variables	+Length of stay	+Municipal variables	+Travel time
Dependent variable						
Readm30_Emergency	0.03	2.093E+11***	0.00	0.00	0.00	0.00
Readm30_Inpatient	0.03	234.34***	0.00	0.00	0.00	0.00
Mort30_LastAdmittance	0.03	1.14	0.00	0.00	0.00	0.00
Mort90_LastAdmittance	0.03	1.30	0.00	0.00	0.00	0.00
Mort180_LastAdmittance	0.03	2.27**	0.00	0.00	0.00	0.00
Mort365_LastAdmittance	0.03	12.32***	0.00	0.00	0.00	0.00
PSI12_vt_pe	0.03	1.90*	0.00	0.00	0.00	0.00
PSI13_Sepsis	0.03	1.38	0.00	0.00	0.00	0.00
PSI15_AccidCutPunc	0.03	−2.13**	0.00	0.00	0.00	0.00
PSI18_ObstTrauma	0.03	−22.80***	−13.18***	−0.31	0.00	0.00
BedSores	0.03	−0.71	0.00	0.00	0.00	0.00

Model *m* includes all variables from model *m-1*. Z-values for model *m* that are significantly different from zero are marked at * 0.10; ** 0.05; *** 0.01 level. Calculations have been performed using Stata 13 matrix commands. There are 58 158 847 observations except for Readm30_Emergency which is not registered in Sweden and therefore has only 39 274 414 observations.

Table A.III Hospital productivity model - Number of observations, mean input and outputs and bootstrapped DEA productivity estimates with 95% confidence intervals

	Denmark	Finland	Norway	Sweden	Total
Observations					
2008	28	32	37	52	149
2009	28	32	29	54	143
Input					
Real costs in KEUR	183 778	118 682	134 550	168 804	152 948
Outputs					
Medical inpatients	126 116	58 333	78 822	69 611	80 057
Surgical inpatients	66 804	56 178	59 928	55 550	58 835
Outpatients	115 661	78 565	64 921	67 999	78 760
Productivity estimates					
Mean	0.791	0.805	0.746	0.566	0.702
95% CI	(0.773 - 0.805)	(0.787 - 0.818)	(0.732 - 0.756)	(0.556 - 0.574)	(0.691 - 0.712)

DOI: 10.1002/hec

Table A.IV Hospital level pairwise Pearson correlation coefficients (first part)

	Average Cost (Operating costs/ DRG-points)	Productivity estimate	Performance measures Readm30_ Emergency	Readm30_ Inpatient	Mort30_ LastDischarge	Mort90_ LastDischarge	Mort180_ LastDischarge	Mort365_ LastDischarge
Average Cost (Operating costs/ DRG-points)	1.00							
Productivity estimate	−0.89*	1.00						
Performance measures								
Readm30_Emergency	0.10	−0.15*	1.00					
Readm30_Inpatient	−0.04	0.12*	0.35*	1.00				
Mort30_LastDischarge	0.13*	−0.23*	0.08	−0.23*	1.00			
Mort90_LastDischarge	0.17*	−0.28*	0.10	−0.18*	0.97*	1.00		
Mort180_LastDischarge	0.24*	−0.34*	0.12*	−0.19*	0.89*	0.96*	1.00	
Mort365_LastDischarge	0.30*	−0.37*	0.11	−0.22*	0.63*	0.75*	0.90*	1.00
Hospital variables								
Number of patients	−0.06	0.09	−0.05	−0.04	−0.14*	−0.17*	−0.16*	−0.12*
Case-Mix Index (CMI)	0.35*	−0.05	−0.05	0.06	−0.19*	−0.19*	−0.16*	−0.06
LOS deviation	0.49*	−0.50*	0.17*	0.10	0.30*	0.33*	0.30*	0.21*
Outpatient share	−0.36*	0.03	−0.10	−0.17*	0.26*	0.24*	0.18*	0.03
UniversityHospital	0.02	0.13*	0.14*	−0.02	−0.28*	−0.31*	−0.31*	−0.23*
CapitalCity	−0.08	0.10	0.09	−0.08	−0.27*	−0.31*	−0.30*	−0.22*
Hospital average of municipal variables								
Population	0.01	0.06	−0.02	−0.01	−0.23*	−0.25*	−0.23*	−0.15*
Unemployment	0.10	−0.12*	0.00	0.10	0.38*	0.44*	0.43*	0.34*
SocialAssist	0.10	−0.11	0.12*	0.42*	0.22*	0.29*	0.25*	0.10
SingleFamilies	0.37*	−0.36*	0.33*	0.58*	−0.06	0.02	0.04	0.04
Foreign	0.18*	−0.17*	0.09	−0.22*	−0.26*	−0.28*	−0.21*	−0.03
Traveltime	0.14*	−0.08	0.04	0.18*	−0.31*	−0.27*	−0.21*	−0.12*

* Significant at the 0.05 level.

CONFLICT OF INTEREST

The authors have no conflict of interest. None of the authors have received grants, speakers fees, etc., from any relevant commercial body within the past 2 years.

ETHICAL STATEMENT

Permission to use patient data from Norwegian Regional Ethics Committee (Ref: 2011/930/REK), from the Norwegian Data Protection Authority (Ref: 11/01210-3/THE), from the Regional ethical review board in Stockholm (Dnr: 2011/213-31/1), and from the Danish Data Protection Agency.

ACKNOWLEDGEMENTS

We acknowledge the contribution of other participants in the Nordic Hospital Comparison Study Group (http://www.thl.fi/nhcsg/) and the EuroHOPE project (http://www.eurohope.info/) in the collection of data and discussion of study design and results. We thank the European Union (7FP grant agreement no. 241721), the Research Council of Norway (grant 214338/H10), as well as our respective employers, for financial contributions, and we also thank the participants of various seminars for helpful comments. We finally thank two anonymous referees for very insightful and useful comments.

REFERENCES

Anthun KS, Goude F, Häkkinen U, Kittelsen SAC, Kruse M, Medin E, Rehnberg C, Rättö H. 2013. Eurohope hospital level analysis: material, methods and indicators. *Eurohope discussion papers* Helsinki, THL.

Ash AS, Shwartz M, Peköz EA. 2003. Comparing outcomes across providers. In *Risk Adjustment for Measuring Health Care Outcomes*, LI Iezzoni.Health Administration Press: Chicago; 297–333.

Table A.IV Hospital level pairwise Pearson correlation coefficients (second part)

	Hospital variables						Hospital average of municipal variables					
	Number of patients	Case-Mix Index (CMI)	LOS deviation	Outpatient share	University Hospital	Capital City	Population	Unemployment	Social Assist	Single Families	Foreign	Traveltime
Average Cost (Operating costs/ DRG-points)												
Productivity estimate												
Performance measures												
Readm30_Emergency												
Readm30_Inpatient												
Mort30_LastDischarge												
Mort90_LastDischarge												
Mort180_LastDischarge												
Mort365_LastDischarge												
Hospital variables												
Number of patients	1.00											
Case-Mix Index (CMI)	−0.04	1.00										
LOS deviation	−0.29*	0.10	1.00									
Outpatient share	−0.07	−0.77*	−0.02	1.00								
UniversityHospital	0.54*	0.08	−0.16*	−0.22*	1.00							
CapitalCity	0.29*	0.04	−0.24*	−0.08	0.46*	1.00						
Hospital average of municipal variables												
Population	0.36*	0.12*	−0.20*	−0.20*	0.39*	0.77*	1.00					
Unemployment	−0.08	−0.10	0.22*	0.29*	−0.10	−0.21*	−0.18*	1.00				
SocialAssist	−0.03	−0.07	0.27*	0.21*	−0.05	−0.15*	−0.11	0.81*	1.00			
SingleFamilies	−0.07	0.04	0.33*	−0.05	−0.03	0.04	0.10	0.28*	0.57*	1.00		
Foreign	0.22*	0.12*	−0.08	−0.20*	0.39*	0.68*	0.67*	−0.34*	−0.30*	0.06	1.00	
Traveltime	0.00	0.10	−0.02	−0.24*	0.06	−0.16*	−0.21*	−0.20*	−0.11*	0.15*	−0.13*	1.00

Asmild M, Tam F. 2007. Estimating global frontier shifts and global Malmquist indices. *Journal of Productivity Analysis* **27**: 137–148.

Barnato AE, Chang C-CH, Farrell MH, Lave JR, Roberts MS, Angus DC. 2010. Is survival better at hospitals with higher "end-of-life" treatment intensity. *Medical Care* **48**(2): 125–132.

Birkmeyer JD, Gust C, Dimick JB, Birkmeyer NJO, Skinner JS. 2012. Hospital quality and the cost of inpatient surgery in the United States. *Annals of Surgery* **255**(1): 1–5.

Bradbury RC, Golec JH, Steen PM. 1994. Relating hospital health outcomes and resource expenditures. *Inquiry* **31**(1): 56–65.

Carey K, Burgess JF. 1999. On measuring the hospital cost/quality trade-off. *Health Economics* **8**(6): 509–520.

Carey K, Stefos T. 2011. Measuring the cost of hospital adverse patient safety events. *Health Economics* **20**(12): 1417–1430.

Charlson ME, 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**(5): 373–383.

Coelli T, Rao DSP, O'donnell CJ, Battese GE. 2005. *An Introduction to Efficiency and Productivity Analysis*, Springer Verlag: New York.

Deily ME, McKay NL. 2006. Cost inefficiency and mortality rates in Florida hospitals. *Health Economics* **15**: 419–431.

Donabedian A. 1966. Evaluating the quality of medical care. *The Milbank Memorial Fund Quarterly* **44**(3): 166–206.

Doyle JJJ, Graves JA, Gruber J, Kleiner SA. 2015. Measuring returns to hospital care: evidence from ambulance referral patterns. *Journal of Political Economy* **123**(1): 170–214.

Drösler S. 2008. Facilitating cross-national comparisons of indicators for patient safety at the health-system level in the OECD countries. *OECD HEALTH TECHNICAL PAPERS*. Paris, OECD.

Farrell MJ. 1957. The measurement of productive efficiency. *Journal of the Royal Statistical Society* **120**: 253–281.

Fleming ST. 1991. The relationship between quality and cost: pure and simple? *Inquiry* **28**: 29–38.

Førsund FR, Hjalmarsson L. 1987. *Analyses of Industrial Structure: A Putty-clay Approach*, Almqvist & Wiksell International: Stockholm.

Fried HO, Lovell CK, Schmidt SS. 2008. *The Measurement of Productive Efficiency and Productivity Growth*, Oxford University Press: Oxford.

Greene WH. 2000. *Econometric Analysis*, Prentice-Hall: New Jersey.

Grifell-Tatjé E, Lovell CK. 1995. A note on the Malmquist productivity index. *Economics Letters* **47**(2): 169–175.

DOI: 10.1002/hec

Gryna FM. 1999. Quality and costs. In *Juran's Quality Handbook*, Juran JM, Godfrey AB (eds.), McGraw-Hill: New York.
Gutacker N, Bojke C, Daidone S, Devlin NJ, Parkin D, Street A. 2013. Truly inefficient or providing better quality of care? Analysing the relationship between risk-adjusted hospital costs and patients' health outcomes. *Health Economics* **22**(8): 931–947.
Hagen T, Häkkinen U, Belicza E, Fattore G, Gaude F. 2015. Acute myocardial infarction, use of percutaneous coronary intervention, and mortality: a comparative effectiveness analysis covering seven European countries. *Health Economics* **24**(Suppl. 2): 88–101.
Häkkinen U, Rosenqvist G, Iversen T, Rehnberg C, Seppälä T, Kohavakka R. 2015. Outcome, cost and their relationship in the treatment of ami, stroke and hip fracture in European hospitals. *Health Economics* **24**(Suppl. 2): 116–139.
Heijink R, Engelfriet P, Rehnberg C, Kittelsen SAC, Häkkinen U. 2015. A window on regional variation in healthcare: insights from EuroHOPE. *Health Economics* **24**(Suppl. 2): 164–177.
Hosmer DW, Lemeshow S. 1980. Goodness-of-fit tests for the multiple logistic regression model. *Communications in Statistics - Theory and Methods* **9**(10): 1043–1069.
Hosmer DW, Lemeshow S, Sturdivant RX. 2013. *Applied Logistic Regression*, John Wiley and Sons: Hoboken, New Jersey.
Hussey PS, Wertheimer S, Mehrotra A. 2013. The association between health care quality and cost: a systematic review. *Annals of Internal Medicine* **158**(1): 27–34.
Hvenegaard A, Arendt JN, Street A, Gyrd-Hansen D. 2011. Exploring the relationship between costs and quality: does the joint evaluation of costs and quality alter the ranking of Danish hospital departments? *European Journal of Health Economics* **12**: 541–551.
Hvenegaard A, Street A, Sørensen TH, Gyrd-Hansen D. 2009. Comparing hospital costs: what is gained by accounting for more than a case-mix index? *Social Science & Medicine* **69**(4): 640–647.
Iversen T, Aas E, Rosenqvist G, Häkkinen U. 2015. Comparative analysis of costs in EuroHOPE. *Health Economics* **24**(Suppl. 2): 5–22.
Kalseth B, Anthun KS, Hope Ø, Kittelsen SAC, Persson B. 2011. Spesialisthelsetjenesten i norden. Sykehusstruktur, styringsstruktur og lokal arbeidsorganisering som mulig forklaring på kostnadsforskjeller mellom landene. *Rapport* SINTEF Health Services Research.
Kittelsen SAC, Anthun KS, Kalseth B, Halsteinli V, Magnussen J. 2009. En komparativ analyse av spesialisthelsetjenesten i finland, sverige, danmark og norge: aktivitet, ressursbruk og produktivitet 2005–2007. *Rapport* SINTEF Health Services Research.
Kittelsen SAC, Magnussen J, Anthun KS, Häkkinen U, Linna M, Medin E, Olsen K, Rehnberg C. 2008. Hospital productivity and the Norwegian ownership reform—a Nordic comparative study. *STAKES discussion paper* STAKES.
Kittelsen SAC, Persson BA, Anthun KS, Goude F, Hope Ø, Häkkinen U, Kalseth B, Kilsmark J, Medin E, Rehnberg C, Rättö H. 2015. Decomposing the productivity differences between hospitals in the Nordic countries. *Journal of Productivity Analysis* **43**(3): 281–293.
Kruse M, Christensen J. 2013. Is quality costly? Patient and hospital cost drivers in vascular surgery. *Health Economics Review* **3**(22).
Leng GC, Walsh D, Fowkes FG, Swainson CP. 1999. Is the emergency readmission rate a valid outcome indicator? *Quality in Health Care* **8**: 234–238.
Linna M, Hakkinen U, Peltola M, Magnussen J, Anthun KS, Kittelsen S, Roed A, Olsen K, Medin E, Rehnberg C. 2010. Measuring cost efficiency in the Nordic hospitals—a cross-sectional comparison of public hospitals in 2002. *Health Care Management Science* **13**(4): 346–357.
McKay NL, Deily ME. 2008. Cost inefficiency and hospital health outcomes. *Health Economics* **17**: 833–848.
Medin E, Anthun KS, Hakkinen U, Kittelsen SAC, Linna M, Magnussen J, Olsen K, Rehnberg C. 2011. Cost efficiency of university hospitals in the Nordic countries: a cross-country analysis. *European Journal of Health Economics* **12**(6): 509–519.
Medin E, Goude F, Melberg H, Tediosi F, Belicza E, Peltola M. 2015. European regional differences in all-cause mortality and length of stay for hip fracture patients. *Health Economics* **24**(Suppl. 2): 53–64.
Medin E, Häkkinen U, Linna M, Anthun KS, Kittelsen SAC, Rehnberg C. 2013. International hospital productivity comparison: experiences from the Nordic countries. *Health Policy* **112**(1): 80–87.
Moger TA, Peltola M. 2014. Risk adjustment of health care performance measures in a multinational register-based study—a pragmatic approach to a complicated topic. *SAGE Open Medicine* **2**.
Morey RC, Fine DJ, Loree SW, Retzlaff-Roberts DL, Tsubakitani S. 1992. The trade-off between hospital cost and quality of care. An exploratory empirical analysis. *Medical Care* **30**: 677–698.
Mukamel DB, Zwanziger J, Tomaszewski KJ. 2001. HMO penetration, competition, and risk-adjusted hospital mortality. *Health Services Research* **36**: 1019–1035.
OECD. 2009. Health care quality indicators project: patient safety indicators report 2009. *OECD Health Working Papers*.
Osius G, Rojek D. 1992. Normal goodness-of-fit tests for multinomial models with large degrees of freedom. *Journal of the American Statistical Association* **87**(420): 1145–1152.

Ozimek A, Miles D. 2011. Stata utilities for geocoding and generating travel time and travel distance information. *Stata Journal* **11**(1): 106–119.

Peltola M, Seppälä TT, Malmivaaraa A, Belicza E, Numerato D, Goude F, Fletcher E, Heijink R. 2015. Individual and regional-level factors contributing to variation in length of stay after cerebral infarction in six European countries. *Health Economics* **24**(Suppl. 2): 38–52.

Simar L, Wilson PW. 1998. Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Management Science* **44**: 49–61.

Stargardt T, Schreyögg J, Kondofersky I. 2014. Measuring the relationship between costs and outcomes: the example of acute myocardial infarction in German hospitals. *Health Economics* **23**(6): 653–669.

Stukel TA, Fisher ES, Alter DA, Guttman A, Ko DT, Fung K, Wodchis WP, Baxter NN, Earle CC, Lee DS. 2012. Association of hospital spending intensity with mortality and readmission rates in Ontario hospitals. *Journal of the American Medical Association* **307**(10): 1037–1045.

Varabyova Y, Schreyögg J. 2013. International comparisons of the technical efficiency of the hospital sector: panel data analysis of OECD countries using parametric and non-parametric approaches. *Health Policy* **112**(1-2): 70–79.

DOI: 10.1002/hec