

Scale and quality in Nordic hospitals

SVERRE A.C. KITTELSEN^{1,*}

KJARTAN S. ANTHUN^{2,3}

UNTO HÄKKINEN⁴

MARIE KRUSE⁵

CLAS REHNBERG⁶

¹ Frisch Centre, Oslo, Norway

² SINTEF Health Research, Trondheim, Norway

³ NTNU Department of Public Health and Nursing, Trondheim, Norway

⁴ National Institute for Health and Welfare, Centre for Health and Social Economics CHESS, Helsinki, Finland

⁵ University of Southern Denmark, Odense, Denmark

⁶ Medical Management Centre, Karolinska Institutet, Stockholm, Sweden
with the Nordic Hospital Comparison Study Group (NHCSG)

Abstract: Empirical analysis of hospitals in production economics often find little or no evidence of scale economies and quite small optimal sizes. Medical literature on the other hand provides evidence of better results for hospitals with a large volume of similar procedures. Based on a sample of Nordic hospitals and patients, we have examined whether the inclusion of quality variables in the production models changes estimates of scale elasticity. A sample of 58 million patient records from 2008 and 2009 in 149 hospitals in Denmark, Finland, Norway and Sweden were collected. Patient data DRG-points were aggregated into 3 outputs (medical inpatients, surgical inpatients and outpatients) and linked to operating costs for 292 observations. The patient data were used to calculate quality indicators on emergency readmissions and mortality within 30 days, adjusted for age, gender, comorbidities, hospital transfers and DRG using DRG-specific logistic regressions. The hypothesis that the elasticity of scale increases when quality variables are included was tested against the null hypothesis of no change in the scale elasticity. The observations were used to estimate a cost function using Stochastic Frontier Analysis (SFA). Country dummies as well as dummies for University hospitals, capital city hospitals and the average travelling time for the patients were included as environmental variables. The estimated scale elasticities did not change with the inclusion of quality indicators in any of the tested models. This may be because medical volume effects are confined to few patient groups or possibly even offset by effects on other groups, where quality is reduced by volume. In one model, the scale elasticity was significantly larger than 1.0, a result that contradicts previous studies which have found decreasing returns.

JEL classification: C14, I12

Key words: hospitals, international comparisons, quality, performance, SFA, economies of scale

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

1 Introduction

Considerations on scale economies are important in determining the optimal hospital structure within a country or region. If there are economies of scale in hospital production, larger hospitals would have lower average costs than smaller hospitals, and hence a hospital structure with few large hospitals would be cost saving for the health sector. If travelling costs and medical outcomes are not adversely affected by a centralised hospital structure, larger hospitals may also be socially optimal.

There have been hospital mergers and increased centralisation of hospital services in many countries. However, empirical analysis of hospitals in production economics often find little or no evidence of scale economies and quite small optimal sizes. Aletras et al. (1997) reported a survey of a large number of empirical studies of economies of scale in hospital services production, with a view to giving recommendations to the British National Health Service (NHS) on the desirability of hospital mergers. Most of the included studies indicated that there are few economies of scale in hospitals beyond 200-300 beds. These findings were unrelated to the analysis being based on flexible cost functions, flexible production functions, Data Envelopment Analysis, survival analysis, studies of multi-hospital firms or more ad hoc studies. However, a Canadian study that estimated optimal hospital size to be 179 beds found that the statistical models were not optimal for special or very large hospitals (Preyra and Pink 2006), and another recent Canadian study demonstrated that optimal scale varied across locations (Asmild et al. 2013).

Studies on Nordic data have usually come to the same result. In e.g. Kittelsen et al. (2015b) there was evidence of diseconomies of scale in Denmark, Finland and Norway, while Sweden showed economies of scale, but these findings may be because the units of observation was mainly physical hospitals in the first three countries while Sweden only had data at the county level ("Landsting"). A long data series of Norwegian hospitals 1999-2014 also found the optimal scale to be quite small (Anthun et al. 2017).

On the other hand, a Danish study identified moderate-to-significant economies of scale and scope (Kristensen et al. 2012). In addition, medical literature provides evidence of better results for hospitals with a large volume of similar procedures. Birkmeyer et al. (2002) examined the relationship between mortality and volume for six different types of cardiovascular procedures and eight types of major cancer resections between 1994 and 1999 in the US, and found evidence of lower mortality with increased volume for all procedures, although the strength of the relationship varied. In a follow-up study, Reames et al. (2014) found that the relationship had strengthened over a 10-year period for five out of eight procedures examined and only weakened in one. Similar results were found for six surgical procedures in Tsai et al. (2013). In a Finnish context, Makela et al. (2011) found evidence that a high volume of total hip replacements in primary osteoarthritis patients reduced hospital costs through reduced length of stay, and may also have increased quality because the hip dislocation rate was reduced. An official report based on Swedish data gave some support for concentrating selected procedures to gain higher quality (SOU 2015:98).

It may be that the quality of hospital services increases with volume, without there being any cost saving associated with the increased quality. In order to examine how volume interacts with costs and quality, data on both indicators at a disaggregated level are needed. Until recently, mortality information has only been available at the level of procedures or diagnosis, while cost information is generally only consistent at the hospital level. As part of the EuroHOPE project (Heijink et al. 2015), a recent study of Nordic hospitals has used linked patient-level and hospital-level data to examine the relationship between productivity/costs and quality (Kittelsen et al. 2015a).

The present study builds on the data collected for the Kittelsen et al. (2015a) article, and on the case-mix corrected quality indicators developed there. That article applied data envelopment analysis (DEA) and found statistically significant quality differences between countries and hospitals for a variety of case-mix adjusted performance measures, but only weak relationships between productivity and quality. As explained in section 3 below, the quality indicators were case-mix adjusted using logistic regression on patient characteristics and diagnosis information. They considered 11 quality indicators, including two measures of readmissions (emergency and inpatient), four mortality rates (within 30, 90, 180 and 365 days) and five patient safety indices (PSIs). The PSIs were only relevant to a very small share of patients, and the readmission and mortality indices respectively were highly correlated. The authors found a statistically significant association between low mortality and high productivity (low costs), so that hospitals could become more efficient by decreasing mortality and costs at the same time. However, the study found an association between low readmission rates and high costs, indicating the opposite, that quality (fewer readmissions) is, in fact, costly.

The theory of scope implies that there is a close relationship between the scale properties and the substitution between quality and quantity, since these are two aspects of the output. The returns to scale could in general be a function of the relative mix of output aspects. If an estimate of the cost function (cost frontier) is controlled for quality, then it is possible that for constant quality there could be more cost savings with increased volume. It is the main aim of this present study to test if quality-related economies of scale existed in Nordic hospital production. The Nordic EuroHOPE data is re-used for testing whether controlling for quality increases estimates of the elasticity of scale.

There could be important policy implications if increased volume allows for reduced unit costs with the same quality, or alternatively allows for increased quality at unchanged unit costs.

2 Data

To perform the analysis in this study we used data on hospital input and both outputs and outcomes. The productivity analysis utilized a single input of hospital costs, and three weighted outputs (medical inpatients, surgical inpatients, outpatient visits) based upon patient level discharge register data from 2008 and 2009. To account for case-mix patient discharges were grouped into Diagnosis Related Groups (DRGs) using a common Nordic grouper, as explained below. To calculate 30 days mortality, demographic data have been collected for 2010 as well. More details are available in Medin *et al.* (2013) and Anthun *et al.* (2013). All somatic hospitals with a 24 hours emergency department or at least two medical or surgical specialities were included, except in Sweden due to problems with the cost data as explained in the next section. Table 1 gives descriptive statistics and summary definitions for the variables in the analysis.

Table 1: Descriptive statistics and summary definitions for key variables.

	Denmark		Finland		Norway		Sweden		All		Definition
	Mean		Mean		Mean		Mean		Mean	Std. Dev	
Patient level raw quality indicators											
Mort30_LastDischarge	0.44 %		0.34 %		0.41 %		0.51 %		0.43 %	6.58 %	Out of hospital mortality rate from any cause within 30 days of admission, only for last admission
Readm30_Emergency	4.76 %		5.52 %		6.96 %		-		5.62 %	23.04 %	Share of patients admitted acutely to inpatient care in hospital within 30 days of the discharge
Hospital level Cost function input and output variables											
Costs in kEUR	183 778		118 682		134 550		168 804		152 948	161 092	Deflated real operating costs in common currency, corrected for differences in input price level between countries and years.
DRG-points Surgical inpatients	66 804		56 178		59 928		55 550		58 835	67 731	Number of department/speciality admissions weighted by DRG-weights calculated from average costs in Helsinki and Uusimaa district
DRG-points Medical inpatients	126 116		58 333		78 822		69 611		80 057	77 193	
DRG-points Outpatients	115 661		78 565		64 921		67 999		78 760	76 230	
Hospital level environmental variables											
CapitalCity, dummy variable	0.286		0.031		0.152		0.085		0.127	0.333	Dummy if hospital is located in the capital of each country
Mean traveltime in hours	0.403		0.424		0.793		0.461		0.517	0.312	Mean travelling time between residence and hospital
UniversityHospital, dummy variable	0.357		0.156		0.242		0.151		0.212	0.410	Dummy if hospital is a teaching or university hospital
Hospital level case mix adjusted performance measures											
Mort30_LastDischarge	1.001		1.041		0.784		1.042		0.976	0.215	Indirectly standardised (observed to predicted) performance indicator case-mix adjusted by age-group, gender, transfer to/from home or hospital, Charlson comorbidity index and number of secondary diagnoses. Numbers less than 1 are better.
Readm30_Emergency	0.910		0.995		1.087		-		1.002	0.201	
Patient observations	15 753 686		12 395 963		11 124 765		18 884 433		58 158 847		Number of admissions
Hospital observations											
in 2008	28		32		37		52		149		All somatic hospitals with a 24-hour emergency department or at least two medical or surgical specialities. In Sweden, six counties were aggregated to the county level and 3 counties were excluded due to missing cost information.
in 2009	28		32		29		54		143		
total	56		64		66		106		292		

2.1 Cost data

The hospitals costs included all production-related costs from somatic hospitals. Costs were harmonized across countries by excluding costs for ambulances, VAT, capital costs, purchased care and costs for teaching and research.

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. Some counties couldn't allocate all cost for acute hospitals to individual hospital units due to either accounting principles or organisational merges of hospitals. For these counties the entire county was considered as one unit of analysis (Blekinge, Dalarna, Gotland, Gävleborg, Kronoberg and Sörmland). In addition, three counties couldn't deliver the costs for acute hospitals (Värmland, Västmanland and Jämtland) and were excluded from the analysis. As can be seen from Table 1, the total number of units for Sweden was $n=52$ (year 2008) and $n=54$ (year 2009). There were 46 individual Swedish hospitals in year 2008 and 48 hospitals in year 2009, together with six counties for each year. The number of missing hospitals was 6 units both years.

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

2.2 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 input prices between the countries, and to allow for comparison between countries the cost level had to be harmonized. For the period 2008-2009 hospital cost deflators were not available from OECD or EuroStat, so wage indices were calculated for the nine most important personnel groups. The wage indices were based on official wage data for the nine separate groups and comprised all personnel costs including wage taxes and pension contributions (Anthun *et al.* 2013; Kittelsen *et al.* 2009; Medin *et al.* 2013). Personnel costs were the most important cost component with about 60 % of total hospital costs. For the other costs we used the Purchasing Power Parity adjusted GDP price index from OECD. To form the aggregate cost level deflator, input price indices should be weighted by fixed cost shares. The nine personnel group indices and the index for other costs were weighed with fixed Norwegian personnel cost shares for 2008, as personnel cost shares were not available for the other countries.

2.3 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 registers. In Denmark and Norway, the date of death was provided by the respective population registries. In Sweden and Finland, the time to death was collected by linking with the cause of death registries.

2.4 DRG grouping and weights

Diagnosis related groups (DRG) is a system of classifying treatments into groups that are more homogenous than all possible combinations of diagnoses and procedures. It is much adopted in the western world as a system for reimbursement, in which each specific DRG group is assigned a specific price or weight, usually based on the average costs estimated for individual admissions in that DRG.

Finland, Norway and Sweden each have a national version of a common grouping system for the hospital visits, NordDRG, developed at the Nordic Casemix Centre. Denmark used to be part of NordDRG but developed a national adaptation, DkDRG, and started using this in 2002 (Ankjær-Jensen *et al.* 2006). The NordDRG grouping as such is not used in practice in any Nordic country. A common grouping is desirable in order to enhance the comparability of the output measures and quality indices, and to remove the impact of some of the idiosyncrasies of the different health systems. All four countries have nation-wide patient registers, and all four countries use the same diagnosis and procedural classification systems: ICD10 and the Nomesco Classification of Surgical Procedures. For DRG-grouping, Datawell Oy Finland has developed a common Nordic grouper based on definitions from the Nordic Casemix Centre. This grouper allows similar grouper logic to be applied in all four Nordic countries, and all patient discharges were grouped in this common Nordic DRG grouper (see Kittelsen *et al.* (2015a); Kittelsen *et al.* (2015b) for similar use of fixed grouper).

To compare the outputs across countries, it is also necessary to apply identical weights for all countries. Such a common weight set does not exist, and the national weights are not directly applicable, since there are some differences between the national DRGs. Instead, a set of Finnish weights were developed: 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. The hospital district is the largest in Finland and its share of total cost of acute somatic care is about 25 %. During the years 2008-2009 the districts had an advanced cost accounting system. Finnish cost weights were used since cost per patient data were not available for the other countries. The weights are normalised so that the average weight of treatments across all DRGs was 1.0.

Basing these weights on average costs in a Finnish hospital district poses additional problems if these weights then reflect costs or incentives that are particular for Finland. However, using calibrated Swedish weights in the Kittelsen *et al.* (2015a) study 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).

In the productivity analysis, the Finnish cost weights were used to aggregate the patient discharges into each of three outputs. Earlier studies have attempted a varied number of aggregation groups, with different results (Magnussen 1996). For the present study, we have chosen to aggregate the patient discharges into each of three outputs to capture important difference among hospitals while not being too specific. Using too many output groups would cause more hospitals to be only comparable to themselves, exhausting the degrees of freedom. The three outputs are DRG-weighted surgical inpatients, medical inpatients and outpatients as described in Table 1.

2.5 Quality indicators

Of the 11 performance indicators calculated in Kittelsen *et al.* (2015a), this study has used only two: i) Emergency readmissions within 30 days and ii) out of hospital mortality within 30 days. The remaining available performance indicators were either variants of these with

different time frames or emergency status, or were patient safety indicators that were relevant for far less than one percent of the patients.

Unlike planned readmissions, emergency readmissions as an inpatient within 30 days of a hospital discharge (but no sooner than the next day) are commonly interpreted as a signal of poor medical quality provided that proper case-mix adjustment has taken place (Leng *et al.* 1999; Morris *et al.* 2014; Tsai *et al.* 2013). Only readmissions as an inpatient were included in this indicator as coding practice for the emergency status of outpatients varied 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. A recent study found a strong association between post discharge complications and readmissions (Morris *et al.* 2014). We included emergency readmissions for any reason, since poor quality in the initial treatment (e.g. an operation) potentially can cause a readmission with another diagnosis (e.g. an infection). Country differences in the readmission rates (reported in Table 1) were considerable, with Denmark at less than 5 % and Norway at almost 7 %. The coverage, and thus quality, in Sweden of the variable reflecting whether the admission is acute or planned, was poor, and therefore we performed separate analyses with Sweden but without emergency readmissions.

The out of hospital mortality rate is probably the most widely accepted quality indicator. Ross *et al.* (2010) showed that higher-volume hospitals was associated with a reduction in 30-day mortality for major medical conditions. Even though a share of mortality is unavoidable, lowering mortality will always be an improvement. The use of mortality rates to punish or reward hospitals or wards in pay for performance schemes (P4P) is controversial (Lilford and Pronovost 2010; Nicholl 2007), but in this analysis we are only interested in the statistical association between mortality and costs. An additional advantage of mortality as quality indicator is the small measurement error.

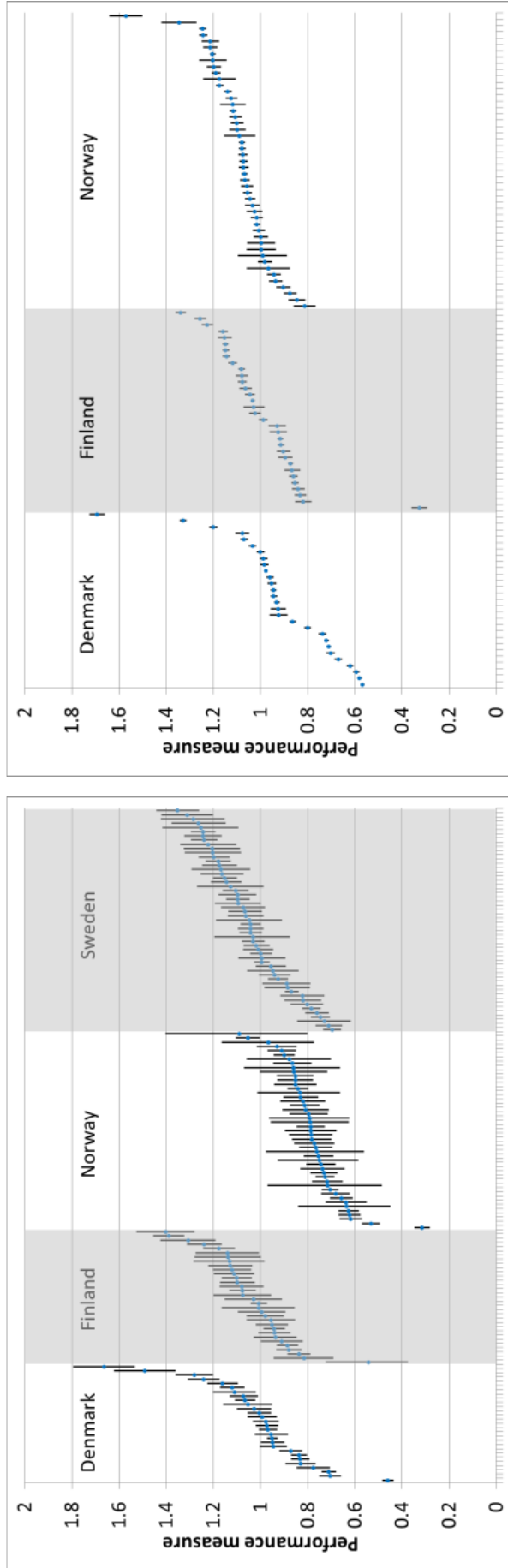
In this analysis, we included only death within 30 days of admission. However, a terminally ill patient could potentially have several hospital stays within the last days of life (with variance likely between type of illness, type of hospital and between countries), but to avoid exaggerating national differences in the treatment of terminal patients we have calculated a mortality dummy only for the last hospital visit before death.

2.6 Case mix adjusting variables

Case-mix adjustment used a logit regression model and variables available in the patient registers, resulting in a performance indicator for each hospital with a value above 1.0 if quality was lower than the Nordic average. The exact procedure is explained in Kittelsen *et al.* (2015a).

Ideally, the adjusting variables should capture characteristics of the patients and their illnesses that possibly influence the outcome, regardless of the treatment given by the hospital. The main risk adjustor used was the DRG group assigned by the common Nordic grouper. Since the division into the more than 700 DRGs was 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. In addition to the DRG adjustment, data was also adjusted for gender as well as age in 10-year groups. The age group from 0 through 9 was divided into two groups; one for infants (age 0) and one for children (age 1-9). For data privacy reasons, the exact age was not available in the pooled cross-country dataset. Although partly endogenous, selected treatment variables were also allowed to adjust for risk, since these may reflect severity. For describing patient transfers in and out of hospital or department variables indicating where the patients came from and where they went, were selected. These variables did not, however, distinguish between

Figure 1: Case-mix adjusted performance measures for hospitals sorted by country, with 99% confidence intervals. Left panel is for mortality within 30 days (Mort30_LastDischarge). Right panel is for emergency readmissions within 30 days (Readm30_Emergency), for which Sweden does not have data. Lower numbers indicate better quality.



transfer to/from home, a (non-hospital) health clinic or a nursing home in all four countries. Comorbidity was included both as the number of secondary diagnoses, and as the Charlson comorbidity index which is based on the severity of secondary diagnoses (Charlson *et al.* 1987).

The case-mix adjustment corresponded to model 2 in Kittelsen et al. (2015a), and excluded length of stay, as this may be partially endogenous; as well as various characteristics of the residential municipality of the patient, as these were found to have little or no impact on quality indicators. The case-mix adjustment was performed by indirect standardisation through a logit estimation on the adjusting variables within each DRG, and the performance indicator was then calculated as the observed divided by the predicted quality level given the patient mix of each hospital. This indirect standardisation ensures that e.g. having many patients in a DRG with high mortality only penalises hospitals that have higher than average mortality in that DRG, nor will having many patients in DRGs with high readmission rates by itself increase the readmission performance measure. Heterogeneity in observable patient mix variables does not then bias the performance measures. The performance indicators listed in table 1 are normalised to 1.0 for the average patient, although this would not be the average across hospitals. Higher numbers indicate worse quality; Figure 1 shows the hospital mean performance indicators together with the 95 % confidence intervals for each hospital. Even though the countries performed significantly different, there were many hospital confidence intervals that overlap, and a ranking of hospital performance can only be partial.

3 Method

The hypothesis that the elasticity of scale increases when quality variables are included was tested against the null hypothesis of no change in the scale elasticity. Earlier evidence (Kittelsen *et al.* 2015b) suggested substantial inefficiency in the Nordic hospital sector, so the observations were used to estimate a cost function using Stochastic Frontier Analysis (SFA) which allows for the separate estimation of inefficiency and stochastic noise.

As discussed, the scale elasticity may well vary with output mix and size. Here we are interested in the simplest empirical specification that allows us to test our hypothesis. The Cobb-Douglas functional form was chosen since the estimated scale elasticity is then a constant in the domain of the function and thus provides a single statistic that can be used to test the hypothesis. The Cobb-Douglas¹ cost function has the form:

$$C = \alpha \prod_i y_i^{\beta_i} \quad (1)$$

Where C is the necessary (minimum) costs for producing the output vector y , and α, β are parameters. Input prices also enter the theoretical cost function, but it has not been possible to distinguish between different inputs at the hospital level, and the single input price was normalised to 1.0. The output coefficients β have the interpretation of cost elasticities and are the percentage increase in costs with a one per cent increase in each output. The scale elasticity is defined as the increase in production resulting from a proportionate increase in inputs, which for the cost function translates to the inverse of the increase in costs resulting from a proportionate increase in output:

¹ The original Cobb & Douglas was defined with constant returns to scale by imposing $\sum_i \beta_i = 1$ but the general form used in the literature does not impose this restriction.

$$\varepsilon_s = \frac{1}{\sum_j \frac{dC}{dy_j} \frac{y_j}{C}} = \frac{1}{\sum_j \beta_j} \quad (2)$$

The empirical specification, in addition to the quantitative output variables y (the three DRG-weighted patient aggregates surgical inpatients, medical inpatients and outpatients) included a vector z of country dummies as well as dummies for University hospitals, capital city hospitals and the average travelling time for the patients as environmental variables or cost shifters. Thus, the model allowed for e.g. the cost in one country to be higher than in another by a proportionality factor. Finally, the quality performance indicators q were controlled for in the estimated equation model:

$$\ln C_{jt} = \ln \alpha + \sum_i \beta_i \ln y_{ijt} + \sum_k \phi_k z_{kjt} + \sum_l \delta_l q_{ljt} + u_j + v_{jt} \quad (3)$$

The estimation procedure used the standard stochastic frontier analysis (SFA) specification of decomposing the error term into an inefficiency term u and a stochastic error v (See e.g. Fried *et al.* 2008). Since data constituted an unbalanced panel of observations for two years, a panel frontier estimator for the separate distributions of the inefficiency and stochastic error terms was used. The assumption is that the total variation of observed costs around the predicted costs can be partly due to a one-sided inefficiency (always positive) and partly due to a symmetric stochastic error, here assumed to be normally distributed. In many panel models, inefficiency is modelled as a fixed effect for each hospital, but with only two years a fixed effects model is precluded by a lack of degrees of freedom. Following Battese and Coelli (1988) the model included a time-invariant inefficiency term u_i that did not depend on year t , but instead by assumption has a truncated normal distribution²:

$$u_j \sim iidN^+(\mu, \sigma_u^2), v_{jt} \sim iidN(0, \sigma_v^2), \gamma = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2} \quad (4)$$

where the statistic γ is the share of the inefficiency variance in the total error variance, i.e. the importance of inefficiency in the estimated model. The tested hypothesis was that the scale elasticity estimate ε was larger with quality variables included than the estimate without quality variables. If higher quality is associated with a larger optimal scale, then one should expect the estimated elasticity of scale to be larger (and thus by (2) the sum of the cost elasticity estimates $\sum_j \beta_j$ to be smaller) in the model with quality variables than in the model without quality variables. The hypothesis was tested by computing ε with and without the quality term in equation (3) and whether the estimates were significantly different at the 95% level.

4 Results

The results of the regressions with and without the quality term are tabulated in Table 2. All models estimated had a very high goodness of fit, and the coefficients on the three output variables were all reasonable and highly significant. The output coefficients in a Cobb-Douglas cost function have the interpretation of cost elasticities, so that e.g. a β coefficient of 0.194 for the surgical inpatients in the first model implies that a 1% increase in the DRG-weighted number of surgical inpatients leads to a predicted 0.194% increase in total costs.

² The estimation is performed by the panel maximum likelihood procedure `xtfrontier` in STATA 13.

Cost elasticities for surgical inpatients were stable across model specifications, but varied more for medical inpatients and outpatients, probably reflecting different criteria for choosing outpatient treatment in Sweden than in the other countries. In the models without Sweden, real costs were about 2 % higher in 2009 than in the reference year 2008 and significantly higher in Norway than in Finland, where Finland is the reference case. Travel time and University hospital significantly increased costs. University hospitals had an estimated 13 % higher cost than non-university hospitals. Location in the capital of each country did not seem to affect costs.

With Sweden included, this country had higher costs than Finland, while the 2009 dummy was insignificant. Other coefficients were essentially unchanged. All models estimated that the larger part of the variation of observed costs around predicted costs was due to inefficiency rather than stochastic noise, since the inefficiency share (γ) was estimated between 0.82 - 0.91.

As for the main hypothesis to be tested, the estimated scale elasticities did not change with the inclusion of quality indicators. With both quality indicators (mortality and emergency readmissions) included and Sweden excluded the point estimates of the scale elasticity changed from 1.041 to 1.039. With only the mortality indicator included the estimates changed from 1.049 to 1.048. Both changes were clearly insignificant. Thus we did not find support for our hypothesis that higher quality is associated with a larger optimal scale.

Perhaps surprisingly, contrary to previous literature estimates, the scale elasticity was estimated as larger than one in all models, significantly so if Sweden was included. Thus, all models indicated increasing returns to scale. While the significance levels of the scale elasticities being different from 1 changes when Sweden is included, the differences between the models (in the order of 0.006-0.009) are not significant.

Robustness has been examined (not reported in tables) by instead estimating a) an ordinary least squares (OLS) model, b) a SFA model that disregard the panel structure of the data, c) a translog cost function, d) a model where outpatients are counted instead of DRG-weighted, and e) a model without cost shifters/environmental variables such as Capital city, Travel time and University hospital. Since it has been suggested (Hvenegaard *et al.* 2011) that the quality effect on costs might be U-shaped, a sixth robustness check f) adding quadratic performance was performed. Results were essentially the same in all specifications, with no statistically significant changes in the estimated elasticity of scale when including quality variables.

Table 2: Results from Maximum likelihood panel estimation of stochastic frontier cost functions with time invariant inefficiency.

Dependent variable	Without Sweden				With Sweden				
	Without quality		With quality		Without quality		With quality		
	Coef.	z	Coef.	z	Coef.	z	Coef.	z	
Ln(Costs in kEUR)									
Constant	0.576	3.40 ***	0.481	2.68 ***	0.410	0.24	0.305	0.27	
Ln(DRG-points Surgical inpatients)	0.194	8.63 ***	0.195	8.64 ***	0.207	8.30 ***	0.212	8.53 ***	
Ln(DRG-points Medical inpatients)	0.429	9.70 ***	0.442	9.89 ***	0.684	19.97 ***	0.687	20.19 ***	
Ln(DRG-points Outpatients)	0.338	9.12 ***	0.326	8.80 ***	0.062	4.72 ***	0.055	4.13 ***	
2009	0.019	2.85 ***	0.021	3.17 ***	0.009	1.17	0.012	1.54	
Denmark	0.018	0.46	0.007	0.18	-0.057	-1.30	-0.057	-1.31	
Norway	0.083	2.11 **	0.102	2.55 **	-0.060	-1.55	-0.038	-0.96	
Sweden					0.296	9.02 ***	0.293	8.88 ***	
CapitalCity	0.036	0.77	0.048	1.05	-0.017	-0.42	-0.003	-0.07	
Traveltime	0.128	3.70 ***	0.132	3.77 ***	0.143	3.84 ***	0.149	4.00 ***	
UniversityHospital	0.095	2.39 **	0.105	2.60 ***	0.128	3.63 ***	0.135	3.80 ***	
Mort30_LastDischarge			0.101	1.96 *			0.112	2.40 **	
Readm30_Emergency			-0.035	-0.66					
gamma (inefficiency share)	0.899	39.63 ***	0.906	42.49 ***	0.816	27.96 ***	0.827	29.63 ***	
sigma_u^2	0.015	5.61 ***	0.015	5.72 ***	0.018	7.75 ***	0.019	7.78 ***	
sigma_v^2	0.002	6.20 ***	0.002	6.13 ***	0.004	7.90 ***	0.004	7.86 ***	
Elasticity of scale	1.041	2.33 **	1.039	2.23 **	1.049	2.96 ***	1.048	2.92 ***	
Change in Elasticity of scale			-0.002	-0.11			-0.001	-0.04	
Number of observations	186		186		292		292		
Number of groups	109		109		163		163		
Log likelihood	183.5		185.4		211.7		214.6		
Chi squared		7 959		7 898		7 598		7 543	

Significant coefficients are marked at 0.10 (*), 0.05 (**), and 0.01 (***) level.

5 Discussion

In this analysis, the inclusion of quality performance indicators in no way changed the scale properties. Thus, even if large hospital volume enhances quality, there seem to be no impact on costs when holding quality constant compared to the scenario when quality is disregarded. With the estimated scale properties, unit costs can be reduced by increasing volume without changing quality. This may be due to the absence of effects of quality on costs in general, but it may also be due to mechanisms that work in different directions. The

medical volume effects on quality seem to have been mostly studied for surgical procedures, and it may be that non-surgical hospital treatments show medical diseconomies of scale.

The applied measures of quality performance may be too restrictive, only covering some of the relevant patient groups. While mortality and emergency readmissions are important quality aspects that are clearly relevant, there are many patients that have low risk of either. Other aspects of treatment can influence health outcomes and patient satisfaction in particular groups and be major cost drivers for the hospitals. Unfortunately such aspects are either too disease specific (such as the patient safety indicators), or are not systematically available as data at the hospital level.

Since this study is based on aggregate hospital volume, it does not consider the volume per physician and the threshold effect linked to individual physicians' annual workload for a specific procedure (Ravi *et al.* 2014). The volume-outcome relationship might also differ across medical specialities and disappear on an aggregate level. The disease-specific quality instruments would be relevant in an analysis of costs and quality at the departmental or speciality level, but there is generally a lack of resource use data or cost data at this level.

Finally, the quality indicators may be insufficiently case-mix adjusted. If larger hospitals treat more severe patients, everything else being equal, then one should expect both mortality rates and emergency readmissions to be higher in large hospitals. The quality indicators used control for differences between DRGs, patient age, gender, comorbidities and transfer patterns. This analysis further controls for differences between countries (i.e. different standards or guidelines), between university and non-university hospitals and for capital cities hospitals. However, if the patients treated e.g. in a large university hospital are more severe than those treated in a small university hospital, within each country, for given DRG and patient characteristics, then there could remain unobservable quality-related economies of scale that are not captured.

Even without quality variables the cost function estimates were characterised by increasing returns to scale. Attempts at finding the reason for this result and its deviation from what has been found in previous literature have not been successful. Methodologically, a change from the specification of e.g. Kittelsen *et al.* (2015b), which uses data for the earlier period 2005-2007, is that outpatients are now DRG-weighted instead of just being counted. The robustness exercise d) mimicked the previous specification but did not change the scale properties. Some earlier studies may also have failed to take account of potential cost disadvantages that are correlated with hospital size, such as university hospital status, capital city location or travelling times, in essence disregarding the possibility that large hospitals have more severe patients or other tasks such as teaching and research. However, estimates on this dataset but excluding these environmental variables did not change the scale properties. A remaining possibility is that recent developments in hospital treatment technology have enhanced returns to scale through e.g. more specialised procedures or machinery.

However, there seems to be no evidence in this dataset that medical returns to scale measured at the hospital level provides any additional justification for larger hospitals over and above any economic returns to scale.

6 Conclusion

The analysis did not support the existence of medical volume effects on key quality indicators of a sufficient strength to increase the scale elasticity at the hospital levels. This may be because medical volume effects are confined to few patient groups or possibly even offset by other groups where quality could be reduced by volume.

The results indicate that there is a potential to reduce costs per treatment by increasing hospital size, without sacrificing or enhancing quality as measured by mortality and readmission rates. Since the scale properties are in contrast to findings in several previous studies, it might be premature to take this as an argument for bigger hospitals in general. In addition, a full analysis must take patient distance-related costs and health effects into account.

If the medical volume effects are very different between patient groups, there might instead be reasons to reorganise the division of functions between large centralised hospitals and smaller local hospitals. Further research into the mechanisms behind medical returns to scale in different patient groups and hospital-level measures of performance is therefore warranted.

Acknowledgements

We acknowledge the contribution of other participants in the Nordic Hospital Comparison Study Group (<http://www.thl.fi/nhcsrg/>) and the EuroHOPE project (<http://www.eurohope.info/>) in the collection of data and discussion of study design and results. Sverre Kittelsen has undertaken the analysis and written the first draft of the manuscript. All authors have contributed text and critically reviewed the manuscript. 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.

References

- Aletras VH, Jones A, Sheldon TA. (1997). *Economics of scale and scope. Concentration and choice in healthcare*. B Ferguson, Sheldon TA, Posnett J.RSM Publishing: London. 23-36.
- Ankjær-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 Management Science* 9 (3), 259-268.
- 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.
- Anthun KS, Kittelsen SAC, Magnussen J. (2017). Productivity growth, case mix and optimal size of hospitals. A 16-year study of the norwegian hospital sector. *Health Policy* 121(4), 418-425.
- Asmild M, Hollingsworth B, Birch S. (2013). The scale of hospital production in different settings: One size does not fit all. *Journal of Productivity Analysis* 40(2), 197-206.
- Battese GE, Coelli TJ. (1988). Prediction of firm-level technical efficiencies with a generalised frontier production function and panel data. *Journal of Econometrics* 38: 387-399.
- Birkmeyer JD, Siewers AE, Finlayson EVA, Stukel TA, Lucas FL, Batista I, Welch HG, Wennberg DE. (2002). Hospital volume and surgical mortality in the united states. *New England Journal of Medicine* 346 (15): 1128-1137.
- 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.

- Fried HO, Lovell CaK, Schmidt SS. (2008). *The measurement of productive efficiency and productivity growth*. Oxford University Press: Oxford.
- 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.
- 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.
- Kittelsen SAC, Anthun KS, Goude F, Huitfeldt IMS, Häkkinen U, Kruse M, Medin E, Rehnberg C, Rättö H. (2015a). Costs and quality at the hospital level in the nordic countries. *Health Economics* 24 (Suppl. 2): 140-163.
- 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, Winsnes BA, Anthun KS, Goude F, Hope Ø, Häkkinen U, Kalseth B, Kilsmark J, Medin E, Rehnberg C, Rättö H. (2015b). Decomposing the productivity differences between hospitals in the nordic countries. *Journal of Productivity Analysis* 43 (3): 281-293.
- Kristensen T, Olsen KR, Kilsmark J, Lauridsen JT, Pedersen KM. (2012). Economies of scale and scope in the danish hospital sector prior to radical restructuring plans. *Health Policy* 106: 120-126.
- 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.
- Lilford R, Pronovost P. (2010). Using hospital mortality rates to judge hospital performance: A bad idea that just won't go away. *Bmj* 340.
- Magnussen J. (1996). Efficiency measurement and the operationalization of hospital production. *Health Services Research* 31: 21-37.
- Makela KT, Hakkinen U, Peltola M, Linna M, Kroger H, Remes V. (2011). The effect of hospital volume on length of stay, re-admissions, and complications of total hip arthroplasty. *Acta Orthopaedica* 82 (1): 20-26.
- 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.
- Morris MS, Deierhoi RJ, Richman JS, Altom LK, Hawn MT. (2014). The relationship between timing of surgical complications and hospital readmission. *JAMA Surgery* 149 (4): 348-354.
- Nicholl J. (2007). Case-mix adjustment in non-randomised observational evaluations: The constant risk fallacy. *Journal of Epidemiology and Community Health* 61 (11): 1010-1013.
- Preyra C, Pink G. (2006). Scale and scope efficiencies through hospital consolidations. *Journal of Health Economics* 25 (6): 1049-1068.
- Ravi B, Jenkinson R, Austin PC, Croxford R, Wasserstein D, Escott B, Paterson JM, Kreder H, Hawker GA. (2014). Relation between surgeon volume and risk of complications after total hip arthroplasty: Propensity score matched cohort study. *BMJ : British Medical Journal* 348.
- Reames BN, Ghaferi AA, Birkmeyer JD, Dimick JB. (2014). Hospital volume and operative mortality in the modern era. *Annals of Surgery* 260 (2): 244-251.
- Ross JS, Normand S-LT, Wang Y, Ko DT, Chen J, Drye EE, Keenan PS, Lichtman JH, Bueno H, Schreiner GC, Krumholz HM. (2010). Hospital volume and 30-day mortality for three common medical conditions. *New England Journal of Medicine* 362 (12): 1110-1118.

Sou. (2015) Sou 2015:98. Träning ger färdighet. Koncentrera vården för patientens bästa. Betänkande av utredningen om högspecialiserad vård. Stockholm Statens Offentliga Utredningar 2015:98.

Tsai TC, Joynt KE, Orav EJ, Gawande AA, Jha AK. (2013). Variation in surgical-readmission rates and quality of hospital care. *New England Journal of Medicine* 369 (12): 1134-1142.

© 2018 by the author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).