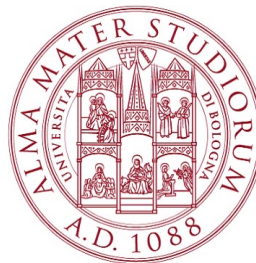# Cost and quality in the Norwegian hospital setting

Veerle Louise Bun

Supervised by Prof. Dr. Sverre Kittelsen (University of Oslo, Frisch Center)

Master Thesis

European Master in Health Economics and Management

Specialization track 'Health Economics and Policy'

University of Oslo and University of Bologna

Written and supervised at:

Faculty of Medicine

Department of Health Management and Health Economics (HELED)

University of Oslo

June 30, 2020

Cost and quality in the Norwegian hospital setting

Veerle Louise Bun

Supervised by Prof. Dr. Sverre Kittelsen (University of Oslo, Frisch Center)

Student number 620463 (UiO), 0900064253 (UniBo), 406658 (EUR)

Abstract

Introduction   The pressure on health care resources is rising. Therefore, creating a sustainable health care system is necessary. Cost minimization and productivity growth is demanded. An undesirable side effect of cost minimizing behaviour could be a reduction in quality. The current evidence on the relationship between cost and quality is ambiguous. The relevance of this research is to find evidence on the correlation between the quality of health care provided by hospitals and their costs in order to figure out what part of costs and cost inefficiency can be explained by quality.

Method   Operating costs, quantitative and case-mix adjusted qualitative (performance indicators) output variables based on data of 22 Norwegian hospitals between 2008-2014 were used to estimate the hospitals' cost functions and cost efficiencies using Stochastic Frontier Analysis. A true fixed effects method was applied to respect the longitudinal aspect of the data and quadratic terms of the qualitative output variables were added in order to allow for a flexible relationship between cost and quality.

Results   Sign, level and significance of estimated coefficients are inconsistent over the cross-section and panel data models and the models with or without quadratic terms. The most consistent result is a negative correlation between cost and quality for the patient safety indicator pulmonary embolism/deep venous thrombosis. Furthermore, negative correlations between cost and quality based on the patient safety indicators sepsis and obstetric trauma are estimated in two models. A positive, but inconsistent correlation between cost and quality based on 90-day mortality and quality based on bed sores was found. A U-shaped relationship between cost and quality was found for the quality indicators of 90-day mortality and an inverted U-shaped relationship was found for the patient safety indicator pulmonary embolism/deep venous thrombosis and the patient safety indicator obstetric trauma. Based on the most basic cross-section model, the average cost efficiency is 89.7%. There is no significant inefficiency term found in the more complex models.

i

Conclusion   The results are inconsistent leading to a remaining uncertainty in the evidence on the cost/quality trade-off. Unobserved heterogeneity and model specification might be important factors of influence in this field of research. The finding of a negative cost/quality relationship indicates that health care organizations could possibly aim for cost minimization by improving quality based on patient safety indicators. There is weak evidence for the existence of a cost/quality trade-off for the quality indicator 90-day mortality and bed sores. Practical implications for health care authorities are that unobserved heterogeneity should be considered when ranking and rewarding hospitals based on their activity and quality.

Preface

This master thesis is written as part of the European Master in Health Economics and Management (Eu-HEM). The relationship between cost and quality in the Norwegian hospital setting is explored. The pressure on health care resources is rising and it is important to know how costs and quality are related in order to understand the dynamics of a sustainable health care system including cost containment and cost minimization policies and their (un)desirable side effects. The project is part of the NFR-financed project 'The effects of DRG-based financing on hospital performance: productivity, quality and patient selection' and falls within the project plan section 2.3.1. 'Multi-level analysis of productivity and quality' (NFR : 214338, REK midt:2012/1887). The project is a collaboration between NTNU and Frisch Center.

Declaration of Oath

V.L. Bun
Date: 29.06.2020

S.A.C. Kittelsen
Date: 30.06.2020

---

[1] Quoted from the Eu-HEM Thesis Guidelines 2018-2020.

Table of Contents

List of Tables

List of Figures

1   Introduction

Health care expenditure in Norway has been increasing over the past years (World Health Organization, 2018; OECD, 2019). This trend has led to the tendency to focus on cost containment in national health care policy in order to achieve efficient allocation of scarce resources in health care. Activity-based financing (ABF) of hospitals might incentivize an increased effort to reduce costs per patient and has been proven effective to achieve higher efficiency in hospitals in Norway (Biørn et al., 2003). However, activity-based financing of hospitals might also incentivize hospitals to increase cost efficiency independent of the consequences of such a policy on the quality of provided health care. In addition to cost containment, quality of healthcare is an important aim of health care policy. Therefore, the incentivized aim for cost minimization and cost efficiency raises the important question about what (unwanted) side effects might occur with respect to the quality aspect of the medical care in hospitals.

In this introduction chapter the Norwegian health care system and the current evidence on the relationship between cost and quality is discussed. Then, the research objective is described. Finally, the relevance of the study is addressed.

1.1   The Norwegian health care system

1.1.1   Contextualisation of the Norwegian healthcare expenditure

According to the report 'Health at Glance' from the Organisation for Economic Co-operation and Development (OECD), Norway has the third highest increase in per capita health spending worldwide over the time period 2007-2018 (OECD, 2019). Over this period Norway's per capital purchasing power parity (PPP) health spending increased by 43%. In 2018, Norway spent 6187 US dollar (USD) PPP per capita; this amount puts Norway on a high ranking in the OECD health care expenditure list (OECD, 2019). Norway's health expenditure in 2018 was 10.2% of its gross domestic product (GDP). Predictions from the 'OECD division projection 2019' which are based on 2015 health expenditure as share of GDP, show that the expected share of health expenditure of GDP in Norway will go towards 12.2% by 2030 (OECD, 2019).

In regard to the question of whether this big spending on health care is paying off by showing results, the report shows that Norway spends more than average on health care while at the same time achieving a lower than average avoidable mortality rate (OECD, 2019). The same result is found for breast cancer; they invest more capital in health care which is paid back by the result that the 5 year survival of breast cancer in Norway is higher than average (OECD, 2019). This shows that investments are paying off.

In general, Norway spends relatively more on inpatient and long-term care, and less on medical goods as share of their total expenditure than OECD average. For example; OECD average is that 38% of the total health expenditure is spent by the hospital provider, while in Norway this is 40% (OECD, 2019). As stated in this report, sufficient amount of resources are necessary to achieve a health system that functions properly (OECD, 2019). However, it does not necessarily mean that by having the highest expenditure you will keep achieving a higher health status in your country. This can be explained by the diminishing returns to scale of health expenditure on health status, explored in the model by Grossman (Grossman, 2000; Zweifel et al., 2009). Therefore, in addition to focusing on the absolute amount of resources that is spent on health care in order to improve health status, it is important to focus on whether this amount is well-spent. In other words: Is there an efficient way of producing health from the provided resources?

1.1.2   Financing health care in Norway

Norway has a publicly funded health system. There is a very limited use of private health insurance. Almost all hospitals in Norway are owned and financed by the state (Statens legemiddelverket, 2016).

In 2002 a reform of the ownership and finance structure of hospital care in Norway was executed (Hagen & Kaarboe, 2006). Ever since, regional health authorities have been given more responsibility: They receive a certain sum of money from the central government and have the responsibility to divide it within their region. This is also called the base payment or basic funding. In addition, there also still is a system of

activity-based financing (ABF) for somatic health care, called 'innsatsstyrt finansier-ing' (ISF) (Helsedirektoratet, 2019a). This finance system is based on the activity of a hospital.

The hospital activity is measured and weighted by the use of diagnosed-related groups (DRGs) (Helsedirektoratet, 2019a). DRGs classify patients into certain disease groups. Patients within these groups are of comparable medical disorder and use of resource. In this way, the finance does not only depend on the number of patients, but it also takes into account the complexity of the different diseases and treatments. The weight of a DRG is based on the average costs nationally. They include all costs that are made by the hospital to help a patient with a certain disease.

Regional health authorities are financed by the central government through the activity-based-payment system and base payments. In 2020, fifty percent of the payment is activity-based and fifty percent is a base payment for somatic health care.

Activity-based financing (ABF) is proven effective in order to achieve higher efficiency in hospitals in Norway (Biørn et al., 2003). Furthermore, the productivity of Norwegian hospitals has increased in the time period 1999 to 2014 (Anthun et al., 2017).

1.1.3   Health care policy in Norway.

In January 2020, the new strategy of the Norwegian health directorate was published. In this strategy it is expressed that the focus of the health policy will be on sustainability of the system (Helsedirektoratet, 2020). How this new policy will be executed in the future is not explicitly stated in the document.

1.2   Cost, cost efficiency and quality

This section reflects on the most important results from current literature on the relationship between quality of health care and cost efficiency, or quality of health care and costs at the aggregate, hospital level. I have found the literature through a semi-structured search on PubMed and through a snowballing approach. Firstly, the international evidence from the literature is discussed. Then, the focus will be on the

evidence from the Norwegian hospital setting.

A short note on definitions will be provided first. The cost/quality trade-off is defined as a situation where you have to sacrifice a certain level of quality when decreasing costs. Quality of care will be decreased, or 'traded-off', in order to minimize costs. In other words, a cost/quality trade-off is found if there is a positive relationship between cost and quality.

For a more extensive background on the definitions of cost, technical, and allocative efficiency see section 2.2. 'Efficiency and costs in health care'. The definition of quality of care is further elaborated in section 2.1. 'The quality of health care'. For more information on the method see section 4.1.1. 'SFA and DEA'.

1.2.1   Literature background

Evidence on the relationship between costs and quality at the hospital level is ambiguous. There is no use of a general method in the literature, neither for cost efficiency calculations nor for the measurement, or inclusion and use of quality indicators. This causes difficulties in the comparison and general interpretation of different study outcomes. Therefore, drawing a conclusion is challenging.

Deily and McKay analyzed the relationship between cost inefficiency and mortality rates of hospitals in Florida, USA (Deily & McKay, 2006). Their article focuses on the portion of the cost that can be defined as inefficiency. According to their article it is important to distinguish between the relationship between costs and quality, and the relationship between cost efficiency and quality (Deily & McKay, 2006). Part of their hypothesis is that if input and process quality is constant, you would not expect to find an association between cost inefficiency and risk-adjusted mortality rates. However, if input quality is constant, but process quality is varying between hospitals, then there are some hospitals that possibly have improved performance with respect to cost inefficiency as well as risk-adjusted mortality rates. So one would expect lower cost inefficiency to be associated with lower risk-adjusted mortality rate. They conclude that a 1%-point reduction in cost inefficiency is associated with 1 fewer in-hospital death per 10.000 discharges. This would indicate that being more cost efficient would

be associated with a lower mortality rate (Deily & McKay, 2006). They performed a Stochastic frontier analysis (SFA) to estimate the cost efficiency scores. However, they only included quantitative output in the SFA model. Their quality parameter 'in-hospital death per 10.000 discharges, corrected for patient risks and other factors', is used as an output in a second model. They used cost inefficiency residuals as data input for an independent variable for a second model, where they modeled their quality parameter as outcome and use cost inefficiency as regressor, together with other variables to risk adjust and case-mix. A follow-up study by McKay and Deily from 2008 focuses again on the relationship between costs due to inefficiency and quality outcome indicators (mortality and complication rates) (McKay & Deily, 2008). The study used data on US hospitals over the same time span, 1999 to 2001, as the previous described study on the hospitals in Florida. In this study they used the same kind of method as in the previous study; a second model regression analysis with cost inefficiency as independent variable of interest, estimated by the first SFA-model, and patient health outcomes mortality and complication rates as dependent variables of interest (McKay & Deily, 2008). The study found no evidence of existence of an association between cost inefficiency and mortality and complication rates in the US over the period 1999-2001. In addition, a different cross-sectional study from 1996 also finds no significant association between mortality ratio and cost efficiency (Ferrier & Valdmanis, 1996).

Carey & Stefos (2011) analyzes the cost-quality relationship using adverse patient safety events as quality indicators and hospital costs. Their most important finding is that adverse patient safety events are associated with higher hospitals costs. Adverse patient safety events are a sign of worse quality of care. This implies that quality and costs are negatively correlated. An interesting question that is raised in the discussion of this article is if the costs of prevention of adverse patient safety events exceed the costs of these adverse events. In conclusion, the results of this article also plead against the existence of a cost-quality trade-off as they find a negative association between cost and quality.

Carey (2003) also focuses on the cost/quality trade-off in the hospital setting. A mortality index is used as quality indicator. The method of use is Stochastic Frontier

Analysis. Costs are lower when there is a higher mortality index (Carey, 2003). This implies a negative association between costs and adverse health events, and a positive association between costs and quality, which indicates a cost/quality trade-off. In this article it is argued that small number of hospitals control on costs by trading it off with quality of service. However, in addition the article elaborates on an alternative hypothesis that some hospitals do not try to control the costs, but they are just too small to have the buying power to invest in the newest medical technology to provide the best possible quality of care (Carey, 2003). Another study on mortality and hospitals costs provides a similar result.

Morey et al. (1992) found evidence that when mortality goes down, the hospital costs tend to increase. This also indicates a possible trade-off between cost and quality. They perform a linear regression model including output of the quality indicator, measured as mortality ratio. They emphasize in their discussion that a limitation is that their quality of care is just measured in mortality ratio and they recognize that the quality of care should be captured in multiple aspects and not just in the sole measurement of mortality (Morey et al., 1992).

Due to difference in analytical tools of use (Data Envelopment analysis, Stochastic frontier analysis and other methods), methods of measuring quality of care (mortality, patient safety parameters, readmission rates) and what type of efficiency or output is estimated (technical efficiency, allocative efficiency, cost efficiency, costs, productivity), it is difficult to derive an overall conclusion on the existence of correlation between costs and quality of care in the literature.

Deily & McKay (2006) and Carey & Stefos (2011) both show findings that their used quality indicator shows a lower quality to be associated with higher cost or higher cost inefficiency. On the contrary, Carey (2003) and Morey et al. (1992) display evidence that lower mortality rates, so higher quality, could be associated with higher costs. These findings are contradictive, and it is unsure to what extend their conclusions could be a result of their methods of use or other limitations like the inability to test if case-mix adjustments are adequate.

This conclusion is in line with the conclusion from a systematic review by Hussey

et al. (2013) on the evidence of the relationship between cost and quality. They provide an insightful overview of the studies on the topic, and derive the conclusion that both positive and negative associations are found, although small. Most studies included in their review had the hospital level as level of analysis and most studies used outcome parameters to measure quality. They suggest that follow-up studies focus on what part of the expenditure is real inefficiency and what part is contributing to quality (Hussey et al., 2013). However, an important limitation of the study is that they excluded all studies that used sources from outside of the USA (Hussey et al., 2013). Therefore, the evidence they provide is evidence of cost and quality trade-off in the USA healthcare setting.

There are other studies focusing on other types of efficiency than costs or cost efficiency, that could possibly provide insight in what way quality of hospital care can influence technical, scale and allocative efficiency, or vice versa. For example Mutter et al. (2010) shows that hospitals operating at a high level of quality of care are more likely to operate technically efficient than hospitals that operate with a lower quality of care. More studies support this evidence. Clement et al. (2008) studies technical efficiency through Data Envelopment Analysis (DEA), including risk-adjusted mortality in the model. They provide evidence that technical inefficiency is associated with lower quality. Furthermore, Ferrier & Valdmanis (1996) find evidence that lower mortality rate is positively related with technical efficiency. In this same study they also find a negative association between quality and scale as well as allocative efficiency. An important interpretation from this result is that in order to achieve an increase in quality, it might be necessarily to trade off quality with the most efficient size and input mix (Ferrier & Valdmanis, 1996).

Nayar & Ozcan (2008) find that there is no difference between calculations of efficiency and performance in hospitals when including or excluding quality indicators in the models. They use Data Envelopment Analysis. They use the quality indicators 'percent of pneumonia patients receiving oxygenation assessment', 'initial antibiotic timing' and 'pneumococcal vaccination'. This result could imply that there is no trade-off between technical efficiency and quality of hospital care (Nayar & Ozcan,

2008).

Valdmanis et al. (2008) provides evidence of the association between total efficiency and patient safety indicators in the hospital setting. Through the DEA method they derived an input efficiency frontier in which they corrected for patient safety indicators. Their aim was to find if certain hospital characteristics are associated with inefficiency and quality. They found that high-quality hospitals tend to be inefficient due to too high labour inputs, having a slack on personnel. Low-quality hospitals have too low labour input (Valdmanis et al., 2008). This indicates that input mixes can be changed in order to decrease inefficiency and increase quality. They also provide evidence that high quality hospitals seem to have a higher overall efficiency. They conclude that in policy perspective, higher overall efficiency could be achieved by more efficient use of resources at hand (Valdmanis et al., 2008).

In conclusion, evidence on the existence of a cost-quality trade-off is ambiguous. The evidence of the association between technical efficiency and quality is more consistent. Furthermore, multiple articles state that results might depend on the choice of method of analytical tool, choice of quality indicator, and whether there was an adequate correction for case-mix (Ferrier & Trivitt, 2013; Nayar & Ozcan, 2008; Morey et al., 1992). Also, it appears that most studies have measured a form of a health outcome parameter to detect the quality of care.

1.2.2   Norwegian literature on cost, cost efficiency and quality

Although there are studies on the efficiency and quality of the hospitals in Nordic countries over the past years, evidence in the literature on the relationship between costs and quality in the hospital setting in Norway is ambiguous. Almost all of the evidence on hospital efficiency in Norway is part of a bigger project called EuroHOPE which is doing a cross-country comparison in the Nordic countries (Kittelsen et al., 2015, 2018).

A 16-year study of the Norwegian hospital sector shows an increase in productivity over the period 1999 to 2014 (Anthun et al., 2017). They point out that further research on possible quality changes alongside the productivity change in this period is required

as quality of healthcare might have changed over time as well.

Multiple studies were conducted on the comparison of cost efficiency between Nordic countries (Linna et al., 2006, 2010). The only study that addresses the association between costs and quality in Norway, tests the hypothesis that difference in productivity between Nordic countries could be caused by difference in health care quality (Kittelsen et al., 2015). This study makes use of panel data from 2008 and 2009. Technical productivity is estimated through DEA. They use a secondary regression model to calculate the effect of quality on technical productivity. The study finds a small but significant, positive relationship between productivity and inpatient readmissions within 30 days. Assuming that quality is lower if inpatient readmission within 30 days is higher, this result could indicate a productivity-quality trade-off. Furthermore, they found a possible sign that higher 30-day mortality could lead to higher costs, but no significant evidence was found on the existence of a trade-off between cost and quality (Kittelsen et al., 2015).

Recently, a master thesis was done on quality and cost trade-off in the Norwegian hospital setting. In this thesis, data from 2008 to 2014 was used as a pooled data set, doing a SFA analysis including quality as an outcome variable. The results of this thesis showed no significant association between costs and emergency readmission rate and a significant, negative association between costs and 30-day mortality. This indicates a trade-off between costs and quality between costs and 30-day mortality rate (Bakke, 2019). However, another important conclusion was that it is possible that these outcomes could have been found due to arbitrary choice of the analytical tool and methods. Furthermore, the data on quality that was included was limited to quality indicators mortality and readmission rates. In addition, the longitudinal aspect of the data was not respected, and therefore, there could be omitted variable bias due to confounding of unobserved heterogeneity between hospitals.

There is one recent study on the scale efficiency and quality in Nordic countries. This study found that costs could possibly be decreased by increasing size of the hospital while thereby increasing scale efficiency, without trading this off for quality parameters readmission and mortality rate (Kittelsen et al., 2018).

A study on comparison of cost efficiency at university hospital level in Nordic countries tried to find determinants of cost efficiency (Medin et al., 2011). It did not include quality indicators. They found that important explanatory variables are geographic location of the hospital and the share of discharges with a high case weight (Medin et al., 2011). However, it is concluded that a large share of the variation in cost efficiency in the university hospital level in this study remains unexplained (Medin et al., 2011).

1.3   Research objective

The major research objective is to expand the evidence on the relationship between cost and quality in the Norwegian hospital setting beyond what has been done before. The main research question is what the relationship between costs and quality is, and if there is a trade-off between cost and quality. The focus of the thesis is on how quality of health care is related to the total operating costs of hospitals and what part of the cost can be explained by quality of care provided in a hospital. The hypothesis is that providing 'high' quality of care is associated with higher operating costs. In conclusion, the focus is on what kind of relationship there is between quality of care and operating costs in Norwegian hospitals in the period between 2008-2014 and what quality indicators possibly play a role in this relationship.

Eventually, this could have policy implications for the main focus of the financing system of the Norwegian hospitals and in what manner cost efficiency and quality of care in hospitals are defined and rewarded in Norway or other countries.

1.4   Relevance

This introduction has briefly described the context of the financial situation of the Norwegian health system and the research objective of this study. Norway has a top ranking on the OECD health expenditure list and there is a prediction by the OECD that the Norwegian health expenditure as percentage of the GDP will keep increasing to 2030 (OECD, 2019). Simultaneously, the Norwegian health directorate expresses

their interest to aim for a sustainable health care system in the newest policy report of January 2020 (Helsedirektoratet, 2020). The relevance of this research is to find evidence on the correlation between the quality of health care provided by hospitals and incentives for 'sustainable healthcare system' by cost-containment and cost efficiency.

This master thesis is a part of the NFR-financed project 'The effects of DRG-based financing on hospital performance: productivity, quality and patient selection' of which the project leader is Jon Magnussen, NTNU [2]. NFR : 214338. REK midt: 2012/1887. The project is a collaboration between NTNU and Frisch Center. This thesis is within the project plan section 2.3.1. 'Multilevel analysis of productivity and quality', with prof. dr. Sverre Kittelsen[3] as thesis supervisor.

Furthermore, this project has been granted a NORCHER[4] student stipend.

---

[2]NTNU is the Norwegian University of Science and Technology based in Trondheim, Gjøvik and Ålesund

[3]Professor II, Avdeling for helseledelse og helseøkonomi (HELED), Institutt for helse og samfunn, Universitetet i Oslo. Contact: sverre.kittelsen@frisch.uio.no

[4]The Norwegian Centre for Health Service Research (NORCHER) aims to provide innovative research on health care delivery models in Norway. More information on https://www.med.uio.no/helsam/forskning/prosjekter/NORCHER-helsetjenesteforskning/index.html

2   Theoretical Framework

2.1   The quality of health care

The academic conversation about the definition of quality of health care and methods of measuring quality of health care is ongoing. Given the difficulty of measuring quality of care and the multidimensional aspects of the topic, the topic is continuously evolving.

According to the model by Grossman, medical care is not demanded for its own sake per se. It is demanded by an individual with the purpose to increase health status (Jacobs et al., 2006; Grossman, 1972). It is important to note that medical care is not the same as health care. Medical care is the provision of a treatment by a health care professional and is part of the bigger concept health care. The concept health care includes, besides medical care, the quality of the service and administration around the medical care that is provided.

The model on quality of health care by Donabedian describes that quality of health care can be defined by using a combination of three main pillars: process, structure and outcome indicators (Donabedian, 1978, 2005).

The *process* of health care can be assessed, for example, in patient record analysis (Donabedian, 1978). This kind of analysis includes a review on the information that is registered in the patient record care by another health care professional. This type of peer review is a review of the medical care or service that is provided based on the available patient information from the patient record. Based on this patient record administration, another health care professional could assess the quality of care that this patient has received while he or she was under treatment. This is a very thorough and time demanding way of analysing health care quality. Even still, this type of analysis leaves a large part of the performance unexplained. It is concluded that the process of health care is hard to observe (Donabedian, 1978, 2005). Since you would have to be an invisible fly on the wall in order to observe process indicators objectively, it is very resource demanding and technically difficult to execute such an analysis as a method to measure quality of care in practice (Donabedian, 1978).

Structural indicators observe a combination of the process and health outcomes, tracking a patient through a trajectory. Health outcomes are variables that display part of the effect of the medical care on the health status.

Outcome indicators cover the improvement in health status of an individual as a result of the treatment. Initially an increase in health status is why health care is demanded. However, medical care only partly explains a change or a level of health status, next to the many genetic, social, and environmental factors that influence health status (Donabedian, 1978). Therefore, it is very hard to explicitly measure the effect of the treatment on health status without capturing effects of other observed or unobserved variables of influence.

In the hospital setting in health care it is hard to observe the counterfactual situation or to establish a counterfactual situation by setting up an randomized controlled trial. Hence, it is challenging to find the pure effect of medical care on health status in health care. Even though econometric models are able to correct for genetic, social, and environmental factors that influence health status, it is difficult to assess whether this correction is adequate. In conclusion, it is difficult to model an effect of medical care on health status that only displays the effect of the level of quality of medical care on health status.

Furthermore, quality of health care is much more than just the effectiveness of a treatment. Given the practice of evidence-based medicine, we assume that a certain treatment has the same effectiveness regardless of the personal characteristics of the physician that performs the treatment (Jacobs et al., 2006). If there are differences in health outcomes after the same treatment between different physicians or health care providers, while keeping patient characteristics and other influencing factors constant, a plausible explanation for this difference could be that there is a difference in the quality of the treatment between the different providers or a difference in the process that lead to the diagnosis and treatment (Jacobs et al., 2006). The difference in quality of treatment can be assessed if there is a adequate correction for the individual (risk) characteristics of the patient that also influence health status and a proper correction of the case-mix of the patient group as a whole. If risk adjustment and

case-mix adjustment is performed adequately, health outcome could be used to assess difference in quality of health care between different health institutions. Unfortunately, determining what part of the production process of medical care is leading to the lower quality is not possible, as more information on the relationship between process, structure, and outcome measurements of the quality of care is needed. For example: did a lack of educated health personnel, thus a lack of knowledge and skill, lead to a series of adverse events for a specific patient group? Then the outcome parameter, used as quality indicator, would possibly show the effect of a problem with a structural issue on health status.

In conclusion, it is very difficult to assess to what extent a variation in quality of care can cause a different outcome in health status and therefore it cannot be assumed that measures of health outcome cover all aspects of quality of care.

The OECD has defined a set of Quality Indicators (QIs) (Drösler et al., 2009). These QIs are standardized, evidence-based measures of health care quality that are used to measure performance and outcomes in the hospital setting. Norway utilizes a quality indicator system which is based on the OECD quality model: The health care quality indicator project (Helsedirektoratet, 2018). Their National Quality Indicator system includes 100 QIs. In this system, they use two groups that are also suggested by the quality model by Donabedian; outcome and process indicators. In addition, they also use patient satisfaction as an indicator. Over the last years they started the implementation of these QIs in the financial health care system in order to reward health care institutions for performance and in the future they will continue to implement this (Helsedirektoratet, 2018).

The national health authority in Norway is responsible for developing and sustaining national quality indicators for health care services. The goal is to make sure that there is access to health care of high quality (Helsedirektoratet, 2018). In order to achieve and maintain this goal the quality indicators can be used to compare health care organizations objectively and incentivize them to improve their health care service. In the Norwegian system a quality indicator is defined as "an indirect measure, a clue that says something about the quality of the area being measured"

(Helsedirektoratet, 2019b).

In this study indicators that display health outcome are used to measure quality of care. The outcome indicators indirectly measure the quality of hospital care.

2.2   Efficiency and costs in health care

Measuring the absolute value of input and output in the health care industry is challenging (Jacobs et al., 2006). The value of the input can be based on the value of the purchased inputs (labor, capital). It is, however, more difficult to value the level of output. Firstly, it can be debated what the health outputs produced are. As discussed before, health care is demanded for the sake of improvement of an individual's health status. Output of a health care organization could be measured in the improved health outcome. An alternative to using health outcome, could be using the quantity of different types of activities, since health care organizations do have information on the quantity of their activities. Unfortunately, these quantitative outputs do not display information on the quality of health care (Jacobs et al., 2006). Additionally, the output still needs to be valued.

The output of a health care organization is not valued by a price that is derived at the market equilibrium in a free market of supply and demand. The consumer, the patient, is protected from paying the full price for the health care service by health insurance or the national health services. In addition, there are more characteristics of the health care market that would lead to market failure in case of a free market setting, e.g. information asymmetry, internalities, and externalities. Thus, it is very challenging to value the outputs that health care organizations produce for several reasons. Prices of output are not set in a free market equilibrium, and the impact on health status of the level of output is non-observable. Additionally, there is a great variation by which a health gain is valued among different individuals.

A solution to valuing output is to benchmark the organizations with each other within a group of health care providers and/or organizations. You could for example figure out the best practice that can be observed within the group. The performance of hospitals within the benchmark group, is valued relatively to the best practice per-

formance that is observed within the benchmark group. In order to compare different health care entities and providers, efficiency analysis is used (Jacobs et al., 2006). A health care organization produces health care and therefore, indirectly, health. But, how health care and health is produced by this organization exactly, is often unknown, a so-called 'black box' (Jacobs et al., 2006). For instance, it is unclear why a certain level of output is generated from a certain level of input for one organization, while for another with this same level of input, a different output level is produced (Jacobs et al., 2006).

A production function describes this black box where input turns into output (Kumbhakar et al., 2015). Productivity is defined as 'the ratio of all valued outputs that an organization produces to all inputs used in the production process' (Jacobs et al., 2006). In order to describe and benchmark the relationship between cost and quality we also need to involve technological innovation and cost efficiency. When actual cost of production exceed the minimum necessary costs to produce the maximum achievable value of output, the organization is considered to be operating cost inefficient (Kumbhakar et al., 2015). When the level of output is lower than the highest possible output or when the level of input is higher than the minimum necessary input, an organization operates technically inefficient(Kumbhakar et al., 2015).

Output oriented technical efficiency (TE_{OO}) is the ratio of attained output with a given input to the best technically, highest possible output with a given input. Input oriented technical efficiency (TE_{IO}) is the ratio of attained input with a given output to the best technically, lowest possible input given output (Jacobs et al., 2006). In this case TE_{IO} could be greater than one. There is an ongoing debate whether it is sensible to estimate an efficiency score of more than 100%. In order for efficiency scores to be between zero and one, one could take the inverse of the ratio of attained input with a given output to the best technically and lowest possible input given output. Using the inverse calculation of input oriented technical efficiency (TE_{IO}) would be a more logical calculation.

In case technical and cost efficiency do not coincide, this can be explained by the existence of allocative inefficiency. When the value of the used combination of inputs is

higher than minimum necessary value of a combination of inputs given that technical efficiency is in place, an organization operates allocatively inefficient (Kumbhakar et al., 2015). Technical efficiency coincides with cost efficiency when there is no allocative inefficiency. The level of cost efficiency will be between zero and one, where an organization with an efficiency score of one operates fully cost efficient. Organizations with an efficiency score close to one are considered efficient. Efficiency and productivity in practice is often depending on the scale of productivity that an organization operates on. This is called economies of scale(Jacobs et al., 2006). The difference in economies of scale between organizations within a group that is compared with each other, might be a confounding factor in a cost efficiency analysis. In order to calculate cost efficiency, a cost function needs to be estimated first. If a cost function is estimated, the cost inefficiency can be defined as 'the extent to which the costs of an organization exceed the predicted costs by the cost function' (Jacobs et al., 2006).

There are two main type of models that can be used to estimate cost efficiency for (health care) organizations: parametric and non-parametric models. Which method is used depends on the relationship between input and output, what amount of input and output variables need to be estimated simultaneously, and whether the researcher needs to take random error from input or output variables into account. Section 4.1.1. 'SFA and DEA' elaborates more extensively on the differences between the two types of analysis and compares strengths and weaknesses of both methods.

Before deciding whether to use a non-parametric model or a parametric model, the possible relationship between cost and quality is studied first. Depending on what relationship is expected between cost and quality, a model is chosen that will fit the expected relationship between the variables of interest best.

2.3   Cost and Quality

As discussed in the introduction section, literature on cost and quality relationship in health care is ambiguous. However, there are a some theoretical frameworks deduced from empirical literature. Considering mortality and readmission rates as quality indicators, it appears that the relationship between cost and the quality indicators is

complex (Fleming, 1991). Fleming finds a cubic relationship between quality and cost where at very low levels of quality, a positive relationship between cost and quality is found. In other words, quality increases when costs increase. But, at some point the positive sign of the derivative changes to a negative sign. There, the marginal costs are negative. Marginal costs in this case are the change in total costs in case of a one unit increase of quality. Fleming calls the area of the curve where the derivative and thus marginal costs are negative, the region of marginal savings (Fleming, 1991). At the highest levels of quality, the derivative sign of the cost function to quality changes again to positive. So, for the highest level of quality, the marginal costs of one extra level of quality of care are positive again (Fleming, 1991). Fleming comments on the negative cost/quality relationship by introducing the concept of 'learning by experience'. More volume could lead to more cost efficiency since they learn to work with less mistakes and more precise for each case (Fleming, 1991). The "learning by experience" increases the ability to decide on the most cost efficient production of health care (Fleming, 1991). Decreasing costs could improve quality in this way. A positive relationship between cost and quality is explained by the expectation that decreasing costs does not only cuts unnecessary costs of the production of health care (Fleming, 1991). However, Fleming does not describe a full hypothesis on the underlying mechanism of the cubic relationship. Figure 1-(a) displays the result of the relationship between quality and cost from Fleming (1991). Area C is the so-called area of marginal savings.

Additional to Fleming's relatively old article (1991), the theory of a U-shaped relationship has come up in the more recent Nordic literature (Hvenegaard et al., 2011). In the study they focus on the relationship between cost and quality in hospital departments and whether implementation of quality indicators changes the ranking in benchmarking departments. The U-shaped relationship is theorized, but not proven by Hvenegaard et al. (2011). This theory on the cost quality relationship suggests that dependent on the level of quality, there are positive or negative marginal costs for the one unit increase of quality. The theory is based on the hypotheses that adverse health events due to worse quality of care can be correlated with higher costs
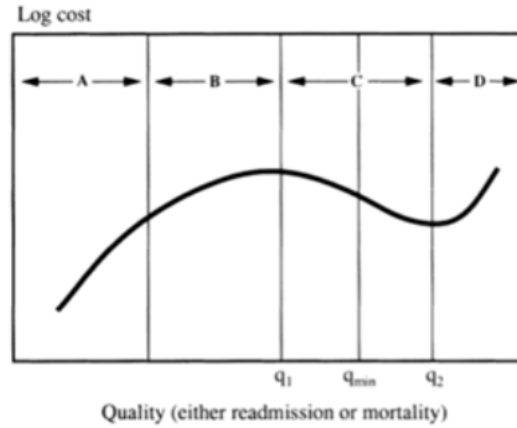
(a) Cubic relationship by Fleming
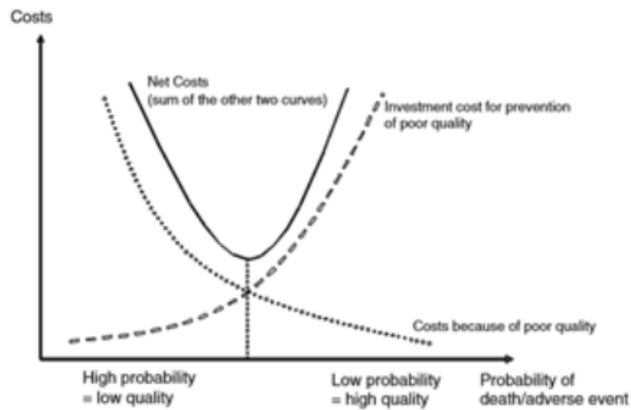
(b) U-shaped relationship by Hvenegaard et al.

Figure 1: Cost/Quality relationship

while at the same time investment in quality of care, i.e. prevention of adverse health events can be correlated with higher costs. Then, costs due to adverse health events decrease if quality improves, and costs of prevention of adverse events increase in order for quality to improve (See Figure 1-(b)). If the marginal costs are positive, cost and quality are positively correlated. This indicates a cost/quality trade-off. The theoretical framework from Hvenegaard et al. (2011) is illustrated in figure 1-(b).

Based on these theoretical frameworks there is no reason to expect a constant trade-off between cost and quality. The theory indicates that the relationship between

cost and quality is complex and U-shaped or cubic. In order to analyze the relationship between cost and quality adequately, the method of use needs to allow for changing marginal costs or trade-offs at different levels of quality. Therefore, a flexible functional form should be considered.

3   Data

This study uses two different sources of data. The first source is patient level data and the second source is data at the hospital level.

3.1   Patient level data

The patient-level data is derived from the Norwegian Patient Registry. The available patient-level data from this registry for this study is data from 2008 to 2014. It includes 39,930,986 units of observations of hospital admissions in 22 hospitals in the years 2008 to 2014. It includes in- and out- patient hospital care.

Patient characteristics recorded in the data are age, sex, and region of residence.

Information on the treatment or so-called 'episode' includes referral date, admission and discharge date and time, to which hospital they are admitted, where they come from before admittance (home, nursing home, other hospital), with what urgency they are admitted to the hospital, and at what level they are treated.

For every admission, the classification of the 'episode' that caused the patient to be in contact with the health care provider, is registered. There is information on the primary diagnosis, secondary diagnosis and the DRG-group of both primary and secondary diagnosis. DRG in this database is the Nordic DRG[5]. ICD-10, the 10th revision of the International Statistical Classification of Diseases and Related Health Problems from the WHO, is also reported (World Health Organization, 2004).

Date of death is registered in this database. Out-of-hospital death is registered in the national population registry, which is automatically linked to the national patient registry.

From the above described information, additional information on episodes and quality indicators is calculated (length of stay, Charlson index, number of secondary diagnosis, readmission rate, referral time, mortality rate, patient safety indicators). The codes for the calculations of these indicators are developed by the EuroHOPE

---

[5]The Nordig DRG is defined by the Nordic Casemix Center in Finland. They are responsible for the definitions and updates of the DRG grouping for the Nordic DRG and the national DRG systems of each of the Nordic countries. More information: http://www.nordcase.org/eng/

Table 1: Patient safety indicators

(Anthun et al., 2012)	
Patient safety indicator	**OECD PSI Code**
Pulmonary embolism, deep vein thrombosis	*OECD PSI-12*
Sepsis	*OECD PSI-13*
Accidental cut, puncture or haemorrhage during medical care	*OECD PSI-15*
Obstetric trauma	*OECD PSI-18*
Bed sores	*N.A.*

project (Anthun et al., 2012).

3.1.1   Quality indicators

Based on the information from the patient-level data, quality indicators for each hospital are calculated. Quality indicators that are derived from the data available are out of hospital mortality within 30, 90, 180 and 365 days after last admission, emergency readmission rate within 30 days after discharge, several patient safety indicators, and referral time. Referral time is not used in this study.

Patient safety indicators are calculated based on the ICD-10. The way that they are defined and calculated is taken over from the EuroHOPE project (Anthun et al., 2012). They correspond to the patient safety indicators (PSIs) defined in OECD Health Care Quality Indicator data collection (Drösler et al., 2009). The PSIs that are calculated are shown in table 1. There is one PSI that is not defined by the OECD, but is used in this research: Bed sores. This quality indicator is also defined and calculated by the EuroHOPE project (Anthun et al., 2012).

3.2   Hospital level data

Hospital level data includes cost data and quantitative output data. The hospital costs reported in this database are operating costs, including all production-related costs in a hospital. Capital, teaching, and research costs are excluded from the operating costs (Anthun et al., 2012). The cost data is derived from the data called 'SAMDATA'. This is a data base for somatic care from the Norwegian Ministry of Health 'helsedi-

rektoratet'. There is no information on input prices. The cost data available is overall operating costs. The database includes information on the total number of patients and total number of DRG-points per group in every hospital. In table 12 in appendix A the 22 Norwegian hospitals that are included in the data, are summed.

3.3 Ethical implications

For the use and aggregation of the patient level data from the Norwegian Patient Registry, ethical permission was granted by REK [6].

---

[6]REK is the regional committee for medical and health reseach ethics.

4   Method

4.1   Efficiency analysis

The following section describes the method used to research the relationship between quality of care and operating costs in Norwegian hospitals in the period from 2008 to 2014 and what quality indicators possibly play a role. In the health care setting, cost function and efficiency estimation can be done in different ways. What method fits best to the research objective and data of use, is discussed in this chapter. The unit of analysis of this study is Norwegian hospitals. The hospitals are the decision-making units of interested. They are considered to have the same function, namely producing medical care. This study operates in the Norwegian hospital setting in which the hospitals are considered comparable.

4.1.1   SFA and DEA

There are two major analytical tools available to estimate efficiency and benchmark different organizations with each other: Data envelopment analysis (DEA) and Stochastic frontier analysis (SFA)(Jacobs et al., 2006). SFA utilizes a parametric model to estimate a production or a cost function(Jacobs et al., 2006). This production or cost function includes an error term that is expected to consist of both idiosyncratic error and inefficiency(Jacobs et al., 2006). Therefore, any deviation of an observation from the best practice on the frontier could be explained by both random noise and inefficiency. Since SFA makes use of a parametric model, an assumption has to be made on the functional form. Additionally, an assumption has to be made on the distribution of the error term, random noise and inefficiency term. The functional form puts a restriction on the estimates and what can be tested.

This would be a main argument to switch to the other method of analysis: Data Envelopment analysis (DEA)(Jacobs et al., 2006). DEA is a non-parametric model. In DEA it is assumed that all deviations from the frontier are solely caused by inefficiency(Jacobs et al., 2006). It is assumed that measurements are done without idiosyncratic error, and therefore there is no possibility to accept noise in the mea-

surements. Statistical tools are difficult to use in DEA(Jacobs et al., 2006). While parametric models directly estimate the cost function, the non-parametric model estimates the efficiency scores (Jacobs et al., 2006). This makes it easier to directly estimate individual efficiencies using the non-parametric model, while in parametric models individual efficiencies are calculated after the estimation of the parameters of the cost or production function.

The data consists of objective measurements of operating costs of hospitals, DRG-points and health outcomes. In general, the most used method in the healthcare sector is the DEA since in this method it is assumed that observations are done without random error. The measurement of quality, however, is considered to be done with error. Therefore, observation of quality might need allowance for random noise. Furthermore, there are theories on the association between cost and quality that show a cubic or U-shaped relationship between cost and quality (Hvenegaard et al., 2011; Fleming, 1991). This type of relationship could be fitted by a logarithmic model with quadratic terms of quality indicators, or a transcendental logarithmic model. Lastly and most importantly, DEA does not allow for situations where estimation of the qualitative and quantitative aspects is done simultaneously. It is not impossible and could be done by applying multiple bootstrap methods, but it is difficult. It is done by Simar & Wilson (2000). They describe a bootstrap method for non-parametric models.

The above discussed arguments would justify to use SFA as method of analysis to estimate the hospital cost function with quantitative and qualitative output and cost efficiency.

4.1.2   Cost function

A cost function is estimated. A cost function is a function of output of the decision-making unit of interest and input prices. This could for example be written as

$$C = f(\boldsymbol{Y}, \boldsymbol{P}), \tag{1}$$

(Jacobs et al., 2006).

where C is operating costs and f is the functional form of the cost function. $\boldsymbol{Y}$ is a vector of hospital output and $\boldsymbol{P}$ is a vector of input prices, which could for example consist of wages and rent (Jacobs et al., 2006). What functional form f will have is an arbitrary decision.

The cost function has several important properties (Coelli et al., 2005). Firstly, non-negativity of the dependent variable (Coelli et al., 2005). Secondly, it is assumed that if the level of the independent variable input price increases, the level of the dependent variable costs shall not decrease (Coelli et al., 2005). This is also assumed for the other independent variable output. If the level of the independent variable hospital output increases, the level of the dependent variable costs shall not decrease(Coelli et al., 2005). Next, homogeneity in price on cost is assumed(Coelli et al., 2005). Lastly, it is assumed that the cost function is concave on the independent variable input price (Coelli et al., 2005).

The application of the cost function in this thesis does not include the explanatory variable input prices because there is no information available on wages or rent. Section 4.4. 'Variables' includes more information on the explanatory variables of use.

The cost function could be written as

$$C = f(\boldsymbol{Y}), \tag{2}$$

where $\boldsymbol{Y}$ is a vector of quantitative hospital output k $(j = 1, ..k)$ for each observation n $(i = 1, ..n)$. Including qualitative output in the cost function, the function could be written as

$$C = f(\mathbf{Y}, \mathbf{Q}), \tag{3}$$

where $\boldsymbol{Q}$ is a vector of qualitative hospital output k $(g = 1, ..k)$ for each observation n $(i = 1, ..n)$. The cost function has a certain functional form f. Applying Cobb-Douglas as a functional form transforms the cost function into

$$C = \alpha \boldsymbol{Y}^{\boldsymbol{\beta}} \boldsymbol{Q}^{\boldsymbol{\gamma}}, \tag{4}$$

which is to be estimated as

$$\ln C = \alpha + \boldsymbol{\beta} \ln \boldsymbol{Y} + \boldsymbol{\gamma} \ln \boldsymbol{Q} + (u + \nu) \tag{5}$$

(Jacobs et al., 2006).

where $(u + \nu)$ is the error term ε which consists of u inefficiency and ν idiosyncratic error. $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are vectors.

In case of heteroscedasticity of one or more of the variables, transforming the function into logarithmic function normalizes the distribution of the variance of the residuals and hereby solves the heteroscedasticity problem (Jacobs et al., 2006). Considering the theoretical framework on the U-shaped relationship between cost and quality as described in section 2.3 'Cost and Quality', a flexible functional form needs to be considered to analyze the association correctly. Transcendental logarithmic is the most used flexible form in empirical research (Greene, 2018). Transforming the cost function including both quantitative and qualitative output into a translog function

leads to the following equation:

$$\ln C = \alpha + \boldsymbol{\beta_y} \ln \boldsymbol{Y} + \boldsymbol{\gamma_q} \ln \boldsymbol{Q} + \frac{1}{2}\boldsymbol{\delta_y}(\ln \boldsymbol{Y})^2 + \frac{1}{2}\boldsymbol{\delta_q}(\ln \boldsymbol{Q})^2 + \boldsymbol{\delta}\ln \boldsymbol{Y} \ln \boldsymbol{Q} + \varepsilon \qquad (6)$$

An important limitation of the use of a translog function is that the degrees of freedom decline with the number of parameters added to the function (Jacobs et al., 2006; Kumbhakar et al., 2015). Therefore, in case of a data set with a relatively small number of observations, instead of fully transforming the cost function into a transcendental logarithmic function, quadratic terms are added for the quality indicators in order to allow the cost function to be flexible in quality.

Transforming the Cobb-Douglas cost function into a cost function including quadratic terms for the qualitative variables, leads to the following equation:

$$\ln C = \alpha + \boldsymbol{\beta} \ln \boldsymbol{Y} + \boldsymbol{\gamma}\ln \boldsymbol{Q} + \boldsymbol{\delta}(\ln \boldsymbol{Q})^2 + \varepsilon \qquad (7)$$

Again, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are vectors. Also, δ is the vector of the coefficient of the quadratic terms of the qualitative variables. If $\delta < 0$ the the cost function is concave while if $\delta > 0$ the cost function is convex in Q.

In this SFA cost function model $\varepsilon = u + \nu$ is the error term, consisting of u inefficiency and ν idiosyncratic error. The value of the (full) error term is estimated through SFA. The value of the inefficiency term can be derived by making an assumption on the distribution of the error term and the distribution of the idiosyncratic error term. The standard assumption would be that the idiosyncratic error term ν is normally distributed with $\nu_i \sim N(0, \sigma_i^2)$.

The literature describes at least four different assumptions on the distributional properties of the inefficiency term u: half-normal, truncated normal, exponential and gamma (Jacobs et al., 2006). In this thesis the inefficiency term is assumed to be half-normal distributed. This is the underlying assumption in the coding of use. In addition, the 'true fixed effects' method by Greene (2004) assumes a half-normal distribution of the inefficiency term. The true fixed effects method is discussed in the next section (see 4.1.3 'Respecting the longitudinal aspect of the data'). Assuming a

half-normal distribution of the inefficiency term, the expected value of inefficiency of each individual unit of analysis can be calculated as:

$$\tilde{u}_i = E(u|\varepsilon_i) = \frac{\sigma\lambda}{(1+\lambda)^2}\left[\frac{\varphi(\frac{\varepsilon_i\lambda}{\sigma})}{\Phi(\frac{\varepsilon_i\lambda}{\sigma})} - \frac{\varepsilon_i\lambda}{\sigma}\right],$$ (8)

(Jacobs et al., 2006)

where $\sigma^2 = \sigma_u^2 + \sigma_\nu^2$, $\lambda = \frac{\sigma_u}{\sigma_\nu}$ and if $\lambda > 0$ there is inefficiency in the error term. $\varphi(.)$ is the probability density function and $\Phi(.)$ displays the cumulative distribution function of the standard normal distribution (Jacobs et al., 2006).

In order to calculate the individual cost efficiency (CE) scores for the Cobb-Douglas cost function, the following equation is used:

$$CE_i = exp(u_i)$$ (9)

(Jacobs et al., 2006)

In order to estimate consistent coefficients several assumptions are made. Firstly, variables are exogenous under $E(\nu|\boldsymbol{Q}, \boldsymbol{Y}) = 0$. The expected value of the idiosyncratic error term ν is zero considering the explanatory variables Q and Y (Wooldridge, 2013). It is highly debatable if this assumption will stand in this context, because it is likely that organizations try to influence their efficiency by changing their inputs (Jacobs et al., 2006). Secondly, the inefficiency term u is assumed to be distributed half-normal, as described above. Both of these assumptions are restrictive and have to be true in order to derive consistent coefficients and an adequately estimated efficiency term. Part of these restrictive assumptions can be dropped when applying a method in which the longitudinal aspect of the data is respected, a panel data method. (Jacobs et al., 2006).

4.1.3   Respecting the longitudinal aspect of the data

Panel data can be used to understand what distributional assumption on the inefficiency term fits the real data best (Jacobs et al., 2006). Furthermore, when applying a fixed effects method, you can allow the error term to be correlated with explana-

tory variables and still be able to estimate a consistent coefficient (Jacobs et al., 2006; Greene, 2018; Wooldridge, 2013). Also, inadequate correction for (unobserved) confounders could lead to omitted variable bias in the estimated coefficients and in the inefficiency term of a pooled analysis of the panel data. Therefore, the most important advantage of panel data methods is the ability to adjust for time invariant unobserved heterogeneity between units of observation, i.e. the unobserved differences between hospitals that are constant over time. There has been an evolution of use different panel data models concerning different aspects over time (Jacobs et al., 2006; Greene, 2004). Greene (2004) showed that the 'fixed effects' methods in SFA by Schmidt & Sickles (1984) and Cornwell et al. (1990) cannot differentiate between (unobserved) heterogeneity between units of analysis and inefficiency. In the models by Schmidt & Sickles (1984) and Cornwell et al. (1990) inefficiency is assumed to be time invariant. Therefore, the inefficiency term captures both unobserved heterogeneity and inefficiency. Greene (2004) concluded that rankings that were derived based on these 'fixed effects' methods were wrong for they were based on inefficiency scores that also captured effects of unobserved heterogeneity between units of analysis. The use of an improved fixed effects panel data model for a cost function in SFA that can distinguish between unobserved heterogeneity and inefficiency, is explored in this thesis. This method is called the 'true fixed effects model' (Greene, 2004, 2005a,b). The 'True fixed effects'-model is written as:

$$C_{it} = \alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + u_{it} + \nu_{it}, \tag{10}$$

(Greene, 2004, 2005a,b)

where $\boldsymbol{\beta}$ is a vector and it is assumed that:

- $corr(\boldsymbol{x}_{it}, u_{it}, \nu_{it}) = 0$;

- $\nu_{it} \sim N(0, \sigma_i^2)$;

- $u_{it} \sim |N(0, \sigma_i^2)|$ from which $u_{it} \geq 0$ follows;

Greene (2005a).

The inefficiency term is assumed to be time variant. ν_{it} is normally distributed, u_{it}

is half normally distributed and the explanatory variables, idiosyncratic noise and inefficiency term are uncorrelated.

In the true fixed effects model inefficiency is assumed to be time varying (u_{it}) and α_i captures the time invariant unobserved heterogeneity between units of analysis (Greene, 2004).

It is favoured to use the true fixed effects model if it is assumed that there is little to no variation of the unobserved heterogeneity within units of observation, and most unobserved heterogeneity is between units and constant over time (Greene, 2005a). This is a strong assumption on the variation of unobserved heterogeneity. The true fixed effects method that allows inefficiency to be time variant is favoured because there is a reasonable chance that the assumed to be time invariant inefficiency term of the alternative models will include the effects time invariant unobserved heterogeneity(Greene, 2005a).

An important observation that is necessary to be able to apply a fixed effects method, is that there should be change in levels in the variables over time within units. When there is lack of change within hospitals over time this could lead to biased results and would plead for not using a fixed effects method (Jacobs et al., 2006).

There are pitfalls in the 'true' fixed effects model by Greene et al.. Firstly there can be an incidental parameter problem when making use of non-linear models with a small number of periods and a large number of parameters (Greene, 2004). Secondly, there is a risk of creating an over specified model. If inefficiency does not vary over time, the inefficiency will be captured in the time constant α_i that captures the hospital specific effect. This leads to an underestimation of the inefficiency term. This is in contrast with other fixed effects panel models in SFA that are more likely to overestimate it (Greene, 2004). The fixed effects method is more likely to overestimate the inefficiency term because unobserved heterogeneity will captured in the inefficiency term since both are assumed to be constant over time (Greene, 2004). While the true fixed effects method is more likely underestimate the inefficiency term because time invariant inefficiency will be split from the time varying inefficiency term

(Greene, 2004). In the true fixed effects method time invariant inefficiency will be included in the time invariant constant or dummy variables capturing the time constant unobserved heterogeneity.

The risk over underestimation of the inefficiency term by splitting time invariant heterogeneity from inefficiency, is preferred over not recognizing the time invariant heterogeneity at all (Greene, 2004).

4.2   Environmental influences

There are several options to correct for environmental influences that affect operating costs of hospitals from the patient level. Environmental influences need to be corrected otherwise the estimated coefficients of the explanatory variables of interest might (partly) capture effects that are caused by differences in environmental influences like for example patient case-mix. Correcting for environmental influences or other observable confounding factors can be done in three ways (Jacobs et al., 2006).

It can be handled by grouping the units of analysis into groups that are known to have almost similar environments (Jacobs et al., 2006). Also, a model can be made which puts constraints into the model. This means that observable confounding factors are included in the model as explanatory variables (Jacobs et al., 2006).

At last, the data can be prepared before running the model. This preparation of the data consists of adjusting the data on the different observable confounding environmental influences, before using it in the efficiency analysis.

In this study the case-mix adjustments on the patient level are done before running the model. This method is often concluded to be the most sensible approach to deal with observable heterogeneity caused by environmental influences (Jacobs et al., 2006). Furthermore, the method was proposed in previous Nordic literature on cost and quality in the hospital setting (Kittelsen et al., 2015).

4.2.1   Case-mix adjustments at the patient level

Correction for environmental influences on quality indicators is necessary. Correction for environmental influences on output DRG-points is not necessary as it is already

Table 2: Case-mix adjustment variables.

(Kittelsen et al., 2015)		
	Variable name	**Variable Definition**
DRG groups	DRG	*Diagnose related group, Nordic DRG.*
Patient characteristics	Agegrp*	*Age dummies for groups: 0, 1-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90+*
	Male	*Sex: male =1, female=0*
Treatment Variables	TransInOwnhospital TransOutOwnHospital TransInOtherHospital TransOutOtherHospital	*Dummies for transfer into and/or out of the hospital department after stay*
	Charlson	*Charlson index*
	NumSecDiagnoses	*Number of secondary diagnoses*

partly inserted in the calculation of the weighted DRG points by the Nordic DRG system. The case-mix adjustments on the quality indicators in this research follow a method proposed by (Kittelsen et al., 2015). In their article they follow Ash et al. (2003) by calculating the observed-to-expected ratio of the quality indicators. The expected value is calculated in a case-mix adjustment model including different environmental influences that need to be adjusted for. The case-mix adjusting model corrects for differences in composition of DRG between hospitals, patient characteristics and treatment variables. Patient characteristics included are sex and age. Treatment variables included are dummies for transfer into and/or out of the hospital department after stay, Charlson index and the number of secondary diagnoses. These variables indicate the severity of disease. In Kittelsen et al. (2015) there are in total five case-mix adjustment models discussed. However, in their study model 2 including DRG, patient characteristics and treatment variables is tested to achieve the best possible case-mix adjustment of all models discussed. Therefore, in this thesis it is decided to use this model as case-mix adjustment model. The case-mix adjustment variables are shown in table 2.

The observed-to-expected ratio is a ratio of the sum of all observed health outcomes

in a certain hospital in all patients in all DRG groups and the sum of all expected health outcomes in a certain hospital in all patients in all DRG groups (Kittelsen et al., 2015). The observed-to-expected ratio is called performance indicator and can be written like this:

$$P_h = \frac{\sum_{k=1}^{K} \sum_{i=1}^{N_{hk}} \omega_{ihk}}{\sum_{k=1}^{K} \sum_{i=1}^{N_{hk}} \hat{\omega}_{ihk}}, \tag{11}$$

(Kittelsen et al., 2015)

where k is DRG $k \in (1, \ldots, K)$ and h is hospital $h \in (1, \ldots, H)$. N_{hk} is the number of patients in DRG k in hospital h. ω_{ihk} is the observed value of the quality indicator for patient i in hospital h in DRG group k. $\hat{\omega}_{ihk}$ is the expected value of the quality indicator for this same patient i in hospital h in DRG group k. The time dimension of the data is ignored in the case-mix adjustments.

$\hat{\omega}_{ihk}$ is predicted in two ways. For the part of the model that corrects for differences in composition of DRG between hospitals $\hat{\omega}_{ihk}$ is predicted by dividing the observed value of quality indicator within each DRG groups with the total patients within this DRG group, resulting in the average value of quality indicator within each DRG group. This can be written as:

$$\hat{\omega}_{ihk} = \frac{\sum_{g=1}^{H} \sum_{j=1}^{N_{gk}} \omega_{jgk}}{\sum_{g=1}^{H} N_{gk}} \tag{12}$$

(Kittelsen et al., 2015)

The expected value of the quality indicator for patient and treatment variables is predicted using a logit model since all quality indicators are binominal variables. As performed in Kittelsen et al. (2015) not all explanatory variables in the model have the same impact in all DRG groups (Kittelsen et al., 2015). Therefore, for each DRG k the expected value of predicted health outcome is estimated with the maximum likelihood estimation of a logit model where ω_{ihk} is the value of the quality indicator for patient i at hospital h in DRG group k. Within this logit model that is written as:

$$\hat{\omega}_{ihk} = \frac{e^{\beta_{0k} + \boldsymbol{\beta}_k \boldsymbol{z}_{ihk} + \varepsilon_{ihk}}}{1 - e^{\beta_{0k} + \boldsymbol{\beta}_k \boldsymbol{z}_{ihk} + \varepsilon_{ihk}}}, \tag{13}$$

(Kittelsen et al., 2015)

$\hat{\omega_{ihk}}$ is predicted by setting the error term $\varepsilon = 0$ and using the estimated β_{0k} and $\boldsymbol{\beta}_k$

that are coefficients specific to each DRG k. $\boldsymbol{z}_{ihk}$ is the vector for case-mix variables that are individual specific. ε_{ihk} is an error term. The distribution of the error term is assumed to be normal (Kittelsen et al., 2015). The expected values of the health outcomes have to be predicted first in order to be able to estimate the performance indicator for each hospital for each case-mix adjustment model. As all quality indicators are based on adverse health events or negative health outcomes, they are considered to display a lower quality of care when the values of these indicators are higher. So performance indicator P_h is interpreted to display a lower quality of care when its value is higher (Kittelsen et al., 2015). For example: If the observed mortality rate in a hospital is higher than the expected mortality rate in this same hospital, the P_h for that hospital will be higher. This relative higher observed mortality rate leads to a higher P_h for mortality rate in that hospital. In this case a higher P_h displays a lower quality of care

4.3   Stata

Stata version 16.0 was used to run the SFA. Furthermore, a package of codes from Kumbhakar et al. (2015) was used to be able to run the SFA panel data models.

4.4   Variables

Table 3 'Description of quality variables' describes the quality variables before case-mix adjustment. Table 4 'Description of variables' describes the dependent and independent variables that are included in the analysis.

The explanatory variables that are included are quantitative hospital output, defined by DRG-points. The hospital output is defined by different quantitative output variables based on DRG-points of four patient categories: elective patients, emergency patients, day patients and polyclinic (outpatient) patients. Elective and emergency patients stay at least one night in the hospital. In addition to quantitative output, qualitative hospital output is included in the model. Qualitative output included are case-mix adjusted mortality rate within 90 days after last admittance, case-mix ad-

Table 3: Description of quality variables

	Variable description	**Variable name**	**Variable definition**
Health outcomes as quality indicators	Readmission	readm_30_emgc	*Emergency readmission within 30 days after discharge.*
	Mortality	mort30last mort90last mort180last mort365last	*All-cause out of hospital mortality.* *Dummies for 30, 90, 180 and 365 days after admission.*
	Patient Safety Indicators	PSI12vt_pe PSI13sepsis PSI15ac_punc PSI18ob_trau Bed_sores	*Table 1 Patient safety indicators.*

justed readmission rate within 30 days and case-mix adjusted patient safety indicators (see section 3.1.1 'Quality indicators').

To overcome collinearity problems only one of the mortality indicators is included in the analysis. The decision on which mortality variable to include in the analysis is based on a trade-off between the size of overlap between the mortality variable and the emergency readmission within 30 days and the certainty that the quality of care of the last admittance was of influence on the occurrence of the adverse health event death. One could theorize that the longer the adverse health event happened after the last admittance, the lower the probability that the quality of care of the last admittance was of influence on the occurrence of the adverse health event. 90-day mortality is less likely to partly overlap information with within 30 days emergency readmission than 30-day mortality. Moreover, mortality within 90 days has a lower probability that the death was influenced by anything other than the health care provided during the last admittance than within 180 and 365 days mortality. So, it is decided to include the performance indicator of mortality within 90 days after last admittance as qualitative variables of the mortality indicators.

Table 4: Description of variables of use.

	Variable description	Variable name	Variable Definition
Dependent variable	Operating costs	ln_kost	*Natural logarithmic transformed operating costs of hospitals in kNOK, already deflated*
Quantitative explanatory variables	DRG points daypatients	ln_v_dagb	*Natural logarithmic transformed DRG points for patients with day treatment without overnight stay*
	DRG points polyclinic patients (outpatient clinic)	ln_v_polk	*Natural logarithmic transformed DRG points for patients receiving health care treatment at the polyclinic*
	DRG points emergency patients	ln_v_emrg	*Natural logarithmic transformed DRG points for emergency patients*
	DRG points elective patients	ln_v_elective	*Natural logarithmic transformed DRG points for elective patients*
Qualitative explanatory variables	Readmission	perf_u2_readm30_emgc	*Performance indicator for acute readmission to hospital within 30 days after discharge*
	Mortality	perf_u2_mort30last perf_u2_mort90last perf_u2_mort180last perf_u2_mort365last	*Performance indicator for out of hospital mortality from any cause. Dummies for 30, 90, 180 and 365 days after discharge*
	Patient Safety Indicators	perf_u2_PSI12vt_pe perf_u2_PSI13sepsis perf_u2_psi15ac_punc perf_u2_psi18ob_trau perf_u2_bed_sores	*Performance indicator for patient safety indicators as defined by OECD. See Table 1 Patient safety indicators.*

4.5   Stochastic frontier models of use

The cost function is estimated through several different Stochastic Frontier models. These models will be discussed in this subsection. **Model A** is a pooled or cross-section estimation of a Cobb-Douglas cost function with only quantitative output. **Model B** is a pooled estimation of a Cobb-Douglas cost function with both quantitative and qualitative output. **Model F** is a pooled estimation of a flexible model B including quadratic terms for the qualitative variables. **Models C** are the panel data models. Model C_A and C_B are, respectively, a true fixed effects estimation of a Cobb-Douglas Cost function with only quantitative output and a true fixed effects estimation of a Cobb-Douglas cost function with both quantitative and qualitative output. Model C_F is the true fixed effects of the flexible model B including quadratic terms for the qualitative variables. All models C are estimated by a dummy estimation method of the true fixed effects model, following Kumbhakar et al. (2015).

A general note on the case-mix adjusted qualitative variables is that they are not transformed into their natural logarithmic forms. This would not make a difference as their levels are close to one and deviating around one.

An overview of the models can be found in table 5.

MODEL A - QUANTITATIVE OUTPUT

$$\ln Cost_{it} = \alpha_0 + \boldsymbol{\beta} \ln \boldsymbol{DRG}_{it} + (u_{it} + \nu_{it}) \tag{14}$$

MODEL B - QUANTITATIVE AND QUALITATIVE OUTPUT

$$\ln Cost_{it} = \alpha_0 + \boldsymbol{\beta} \ln \boldsymbol{DRG}_{it} + \boldsymbol{\gamma} \boldsymbol{PI}_{it} + (u_{it} + \nu_{it}) \tag{15}$$

MODEL F – FLEXIBLE MODEL

$$\ln Cost_{it} = \alpha_0 + \boldsymbol{\beta} \ln \boldsymbol{DRG}_{it} + \boldsymbol{\gamma} \boldsymbol{PI}_{it} + \boldsymbol{\delta}(\boldsymbol{PI}_{it})^2 + (u_{it} + \nu_{it}) \tag{16}$$

In model A, B and F the panel data set is analyzed as if it is a cross sectional data set. So, time dimension t is suppressed.

MODEL C - TRUE FIXED EFFECTS MODEL

Model C_A

$$\ln Cost_{it} = \boldsymbol{\beta} \ln \boldsymbol{DRG}_{it} + \alpha_i + (u_{it} + \nu_{it}) \qquad (17)$$

Model C_B

$$\ln Cost_{it} = \boldsymbol{\beta} \ln \boldsymbol{DRG}_{it} + \boldsymbol{\gamma PI}_{it} + \alpha_i + (u_{it} + \nu_{it}) \qquad (18)$$

Model C_F

$$\ln Cost_{it} = \boldsymbol{\beta} \ln \boldsymbol{DRG}_{it} + \gamma \boldsymbol{PI}_{it} + \boldsymbol{\delta}(\boldsymbol{PI}_{it})^2 + \alpha_i + (u_{it} + \nu_{it}) \qquad (19)$$

True fixed effects model C is estimated by inserting dummies for every unit. This can be written down as:

Model C_A

$$\ln Cost_{it} = \alpha_i * D_i + \boldsymbol{\beta} \ln \boldsymbol{DRG}_{it} + (u_{it} + \nu_{it}) \qquad (20)$$

and Model C_B

$$\ln Cost_{it} = \alpha_i * D_i + \boldsymbol{\beta} \ln \boldsymbol{DRG}_{it} + \boldsymbol{\gamma PI}_{it} + (u_{it} + \nu_{it}), \qquad (21)$$

Model C_F

$$\ln Cost_{it} = \alpha_i * D_i + \boldsymbol{\beta} \ln \boldsymbol{DRG}_{it} + \boldsymbol{\gamma PI}_{it} + \boldsymbol{\delta}(\boldsymbol{PI}_{it})^2 + (u_{it} + \nu_{it}) \qquad (22)$$

For all equations:

- $\ln Cost_{it}$ is the natural logarithm of the operating costs of hospital i in time t;

- hospital $i = 1, ...22n$;

- $\boldsymbol{DRG}_{it}$ is the vector of the quantitative variables defined by DRG-points per DRG-group (DRG) $d = 1, ..4m$ for each hospital i in time t;

- $\boldsymbol{PI}_{it}$ is the vector of the qualitative variables defined by performance indicators (PI) of the quality variables $p = 1, ...7q$ for each hospital i in time t;

- $\boldsymbol{\beta}$ is the vector of coefficients of quantitative variables d. $\boldsymbol{\gamma}$ is the vector of coefficients of qualitative variables p and $\boldsymbol{\delta}$ is the vector of coefficients of the quadratic terms of the qualitative variables p;

- D_i is the dummy variable for hospital i. $D_i = D_1, ...D_{22}$ are all dummy variables for the hospitals in the data;

- α_i is the time constant unobserved heterogeneity of each hospital i;

- and time $t = 1, ...7T$.

Table 5: Overview of stochastic frontier models of use.

Cross section models	Model A	$\ln Cost_{it} = \alpha_0 + \boldsymbol{\beta} \ln \boldsymbol{DRG}_{it} + (u_{it} + \nu_{it})$
(Panel data treated as cross sectional	Model B	$\ln Cost_{it} = \alpha_0 + \boldsymbol{\beta} \ln \boldsymbol{DRG}_{it} + \boldsymbol{\gamma} \boldsymbol{PI}_{it} + (u_{it} + (\nu_{it})$
data, t suppressed)	Model F	$\ln Cost_{it} = \alpha_0 + \boldsymbol{\beta} \ln \boldsymbol{DRG}_{it} + \boldsymbol{\gamma} \boldsymbol{PI}_{it} + \boldsymbol{\delta}(\boldsymbol{PI}_{it})^2 + (u_{it} + \nu_{it})$
Panel models	Model C_A	$\ln Cost_{it} = \boldsymbol{\beta} \ln \boldsymbol{DRG}_{it} + \alpha_i + (u_{it} + \nu_{it})$
	Model C_B	$\ln Cost_{it} = \boldsymbol{\beta} \ln \boldsymbol{DRG}_{it} + \boldsymbol{\gamma} \boldsymbol{PI}_{it} + \alpha_i + (u_{it} + \nu_{it})$
	Model C_F	$\ln Cost_{it} = \boldsymbol{\beta} \ln \boldsymbol{DRG}_{it} + \boldsymbol{\gamma} \boldsymbol{PI}_{it} + \boldsymbol{\delta}(\boldsymbol{PI}_{it})^2 + \alpha_i + (u_{it} + \nu_{it})$

5   Results

5.1   Descriptive statistics

The data contains 39,930,986 observed patient hospital admissions [7] that are collected from 22 different hospitals over a period of 7 years from 2008 to 2014 (Table 6). The unit of observation in the cost analysis is a hospital. The total number of observations over 7 years is 151. 22 unique hospital units observed over 7 years would lead to a balanced panel of 154 hospital observations. However, this is an unbalanced panel data set for there are 3 missing hospital observations [8].

Table 6: Overview of observations in data

Number of hospitals	22
Total number of patient observations	39,930,986
Number of years	7
Number of hospital observations	151

Table 7 summarizes the data on cost, quantitative and qualitative output variables. Costs are reported in kNOK[9]. Mean operating costs per unit of observation (hospitals) are ca. three billion NOK or 300 million euro [10]. The range of operating costs goes from a hospital with a minimum value of operating costs of 589 million NOK to a hospital with a maximum value of operating costs of 14.3 billion NOK. This is a difference of over 13 billion NOK between the lowest and highest observed value of operating costs in the data (Table 7). The mean number of DRG-points of emergency, elective, day and polyclinic patients are respectively 26,501, 16,668, 3,681 and 9,108 DRG-points.

The scatter plot of operating costs and either one of the DRG-points variable (the number of DRG-points of emergency, elective, day and polyclinic patients) shows heteroscedasticity (Appendix D). In order to deal with the heteroscedasticity, the variables

---

[7]This is the number of patient observations from the patient level data. The number of patient observations from the hospital level cost data is 39,700,334. The origin of the difference in this counting is unknown. A possible explanation is difference or bias in registration.

[8]There are two observations unmatched in the year 2008 and there is 1 observation unmatched in the year 2010. These observations had to be deleted from the data set manually due to uncertainty in fit of the match in the merging process of the patient level and hospital level data

[9]NOK is Norwegian Krone and 1 kNOK = 1000 NOK

[10]1 euro ≈ 10 NOK

are transformed to the natural logarithm of operating costs and the natural logarithm of number of DRG-points of emergency, elective, day and polyclinic patients. Appendix D includes the scatter plots of before and after transformation of the variables to their natural logarithms.

Table 7 also summarizes the data on qualitative output variables and the case-mix adjusted quality variables, called performance indicators (PI). The mean number of emergency readmissions within 30 days, mean number of deaths within 30, 90, 180 and 365 days are respectively 15,072, 972, 1,263, 1,405 and 1,569. The mean number of deaths is increasing over the four variables, where 365-day mortality has the highest mean. This is expected for the way they are calculated is overlapping. The number of death within 365 days includes also the number of death within 180, within 90 and within 30 days.

The mean number of patients with adverse health events[11] for pulmonary embolism/ deep vein thrombosis (OECD PSI-12) and sepsis (OECD PSI-13) are respectively 198 and 175. This is relatively higher than the mean number of patients with adverse health events accidental cut, puncture or haemorrhage during medical care (OECD PSI-15), obstetric trauma (OECD PSI-18) and bed sores that have a mean number of patients with adverse health events of respectively 57, 42 and 88. In general, the means of the number of cases of these patient safety indicators are a low compared to the mean of the total number of patients (264,444). Furthermore, accidental cut, puncture or haemorrhage during medical care (OECD PSI-15) and obstetric trauma (OECD PSI-18) have a minimum level of zero. This reveals that there are hospitals in the dataset where the number of patients reported with this type of adverse health event is zero. Possibly this could be explained by the fact that these adverse health events did not occur in all hospitals or that they were not reported in all hospitals.

The mean levels of the predicted number of cases of the quality indicators and the performance indicators[12] are also reported in table 7. In general, the mean of

---

[11]These adverse health events are reported based on the patient safety indicators (see table 1)

[12]The performance indicators are the ratio of the observed-to-expected number of cases, see equation 9 in chapter 4 'Method'. The performance indicators (PI) are to be interpreted as that a hospital with a higher level of PI for a certain quality indicator has a lower quality for

Table 7: Summary of qualitative output variables.

Variable	Mean	Std. Dev.	Min	Max
Number of hospital observations N= 151				
Quantitative variables				
Costs	3036796	2624917	589178	14251069
ln_kost	14·64	0·75	13·29	16·47
v_emrg_p	26501·06	17052·05	6241·67	83062·45
v_elective_p	16667·60	18283·60	2592·77	98833·00
v_dagb_p	3681·11	2645·82	454·14	12617·76
v_polk_p	9108·05	7799·05	921·16	41604·05
Number of patients	264443·60	203104·40	30738	1049609
Observed number of cases health outcome quality indicators				
readm30_emgc	15072·31	11674·81	2408	62015
mort30last	971·56	535·08	220	2231
mort90last	1262·76	688·77	329	2886
mort180last	1404·71	770·27	366	3504
mort365last	1569·15	898·27	385	4598
psi12vt_pe	197·65	155·67	23	981
psi13sepsis	174·70	142·22	27	743
psi15ac_punc	57·36	55·70	0	290
psi18ob_trau	42·42	37·56	0	162
bed_sores	87·58	71·79	14	371
Predicted number of cases health outcome quality indicators				
pred2_readm30_emgc	14892·54	11584·75	2279·66	62115·51
pred2_mort30last	1040·93	609·03	216·90	3109·37
pred2_mort90last	1346·01	791·75	278·87	4025·46
pred2_mort180last	1493·55	879·52	313·65	4458·67
pred2_mort365last	1661·43	980·51	355·24	4945·84
pred2_psi12vt_pe	208·21	148·65	40·25	837·81
pred2_psi13sepsis	178·85	147·56	23·98	847·73
pred2_psi15ac_punc	62·68	50·75	4·97	260·55
pred2_psi18ob_trau	43·17	39·68	0·32	207·99
pred2_bed_sores	99·28	63·88	18·40	343·54
Performance indicators - case-mix adjusted quality indicators				
perf_u2_readm30_emgc	1·015	0·102	0·698	1·320
perf_u2_mort30last	0·955	0·125	0·612	1·328
perf_u2_mort90last	0·966	0·128	0·601	1·301
perf_u2_mort180last	0·969	0·137	0·586	1·402
perf_u2_mort365last	0·970	0·178	0·542	1·614
perf_u2_psi12vt_pe	0·941	0·209	0·556	1·757
perf_u2_psi13sepsis	1·018	0·270	0·342	1·911
perf_u2_psi15ac_punc	0·805	0·325	0·000	1·883
perf_u2_psi18ob_trau	0·921	0·545	0·000	2·435
perf_u2_bed_sores	0·920	0·465	0·275	2·696

the performance indicators for the quality indicators are reported to be around one or slightly lower than one [13]. If the mean performance indicator(PI) over all hospital observations is lower than one, this implies that at the mean the observed number of cases are lower than the expected number of cases at the mean level. A mean PI that is not very close to one could be explained by a certain group of good or bad performing hospitals included in the data, or by insufficient case-mix adjustment.

Sepsis (OECD PSI-13) and emergency readmission within 30 days after discharge have a level of the PI that is higher than one at the mean. The maximum reported level of the PI for bed sores and the PI for PSI18 obstetric trauma level in the data is respectively 2.696 and 2.435. This implies that in the hospitals where these maximum values are reported, the absolute level of observed number of adverse health events were more than two times the absolute level of the expected number adverse health events of adverse health events. Possibly, this could be explained by poor quality of health care in these hospitals or by an incomplete case-mix adjustment for the occurrence of these adverse health events.

All variables included in the analysis change over time and the change between units is relatively higher than the change within units (See Appendix E). The Pearson correlation between the quantitative output variables is high (See Appendix C). Nevertheless, they will all be included in order for the quantitative output to be complete. The collinearity problem leads to less precise estimations of the coefficient of the quantitative variables and therefore interpretation should be done with care. As expected there are high Pearson correlations between the mortality variables (See Appendix B). The mortality quality variables collect overlapping information and therefore one of the mortality variables is included in the analysis to overcome this problem of multicollinearity.

---

this quality indicator.

[13]It is the mean of the PI values over all hospital observations (N=151). The mean PI over all patient observations would be equal to 1.

5.2   Cross section models A, B and F

In this section the results of the estimation of Model A, B and F are discussed. These models are estimated by using the data as if it is a cross-sectional dataset and therefore the longitudinal aspect of the data is not taken into account (See table 5). A 5%-significance level is set. A coefficient is considered to be not significant if the 95%-confidence interval contains zero and the p-value is higher than 0.05. The first three columns of table 8 report the estimated results based on the cross-section models.

For all three models, the Wald test on the fit of the cost function shows a p-value smaller than 0.0001. The log likelihood of the model is increasing from model A to F from 123.765 to 180.425.

The coefficients of the natural logarithm of the DRG-points for day patients, emergency patients and elective patients are positive and significant at the 1%-significance level (p<0.01). The coefficient of the natural logarithm of the DRG-points for polyclinic patients is positive and significant at the 1%-significance level model A and B[14].

In model B the qualitative variables show significant, positive coefficients for the patient safety output variables: pulmonary embolism/deep vein thrombosis (OECD PSI-12), sepsis (OECD PSI-13), accidental cut, puncture or haemorrhage during medical care (OECD PSI-15) and obstetric trauma (OECD PSI-18)[15]. The qualitative variable for the patient safety output variable bed sores is negative and significantly different from zero. The coefficients of the PIs of the patient safety indicators in model F are positive and significant for pulmonary embolism/deep vein thrombosis (OECD PSI-12) and obstetric trauma (OECD PSI-18). In addition, the coefficient of the PI of 90-day mortality is negative and significant in model F.

---

[14]In model F the coefficient of natural logarithm of the DRG-points for polyclinic patients is positive and significant if instead of the PI of mortality within 90 days, the PI mortality of within 180 or 365 days is used as quality variable

[15]In model B the coefficient of the PI of mortality within 30 days is negative and significant if used as quality variable instead of the PI of mortality within 90 days

Table 8: Stochastic frontier cost function model A, B, F, C_A, C_B and C_F

	Cross-section models						Panel data models					
	Model A		Model B		Model F		Model C_A		Model C_B		Model C_F	
dependent variable: ln_kost	Coefficient	95% CI	Coefficient	95% CI	Coefficient	95% CI	Coefficient	95% CI	Coefficient	95% CI	Coefficient	95% CI
ln_v_dagb	0.249***	-0.156, 0.341	0.189***	0.098,0.280	0.223***	0.138,0.308	0.411***	0.354,0.468	0.387***	0.330, 0.444	0.388***	0.332,0.445
ln_v_polk	0.163***	0.060,0.266	0.183***	0.083,0.282	0.087*	-0.005,0.179	0.092***	0.025,0.158	0.112***	0.032, 0.191	0.128***	0.049,0.208
ln_v_emrg	0.217***	0.139,0.295	0.227***	0.141,0.312	0.252***	0.177,0.327	0.326***	0.246,0.407	0.342***	0.254, 0.431	0.360***	0.271,0.448
ln_v_elective	0.313***	0.252,0.375	0.322***	0.260,0.384	0.328***	0.266,0.390	0.168***	0.121,0.216	0.189***	0.138, 0.240	0.196***	0.144,0.248
perf_u2_readm30_emgc			-0.107	-0.290,0.076	0.719	-0.734,2.172			0.078	-0.027, 0.184	0.896**	0.047,1.744
perf_u2_mort90last			-0.152*	-0.331,0.028	-3.478***	-4.593,-2.363			0.110*	-0.016,0.236	-0.051	-1.067,0.965
perf_u2_psi12vt_pe			0.109***	0.035,0.184	0.466**	0.063,0.868			0.037*	-0.006,0.079	0.458***	0.237,0.680
perf_u2_psi13sepsis			0.172***	0.114,0.230	-0.220	-0.490,0.050			0.036**	0.003,0.068	0.029	-0.106,0.164
perf_u2_psi15ac-punc			0.102***	0.048,0.155	0.055	-0.123,0.233			0.024	-0.008,0.056	-0.062	-0.161,0.037
perf_u2_psi18ob_trau			0.064***	0.030,0.098	0.179***	0.076,0.282			-0.001	-0.025,0.022	0.014	-0.068,0.096
perf_u2_bed_sores			-0.037**	-0.069,-0.005	0.044	-0.073,0.160			0.003	-0.020,0.027	0.000	-0.072,0.071
$(perf\_u2\_readm30\_emgc)^2$					-0.416	-1.126,0.293					-0.380*	-0.787,0.026
$(perf\_u2\_mort90last)^2$					1.665***	1.112,2.219					0.090	-0.421,0.601
$(perf\_u2\_psi12vt\_pe)^2$					-0.155	-0.351,0.040					-0.211***	-0.317,-0.105
$(perf\_u2\_psi13sepsis)^2$					0.154**	0.032,0.276					0.001	-0.060,0.062
$(perf\_u2\_psi15ac\_punc)^2$					0.020	-0.080,0.120					0.050*	-0.007,0.107
$(perf\_u2\_psi18ob\_trau)^2$					-0.051**	-0.097,-0.004					-0.008	-0.039,0.024
$(perf\_u2\_bed\_sores)^2$					-0.025	-0.069,0.018					0.000	-0.025,0.026
... hospital dummies							yes		yes		yes	
constant	6.039***	5.702, 6.376	6.144***	2.815,9.472	7.601***	5.780,9.421	5.998***	5.216,6.780	5.345***	4.438, 6.252	4.337***	3.049,5.626
Inefficiency												
$\ln(\sigma_u^2)$	-3.961***	-4.800, -3.123	-15.218	-16712.299,16681.863	-17.173	-19981.281,19946.935	-14.922	-217.090,187.245	-16.034	-1140.079,1108.011	-14.401	-442.317,413.516
$\ln(\sigma_v^2)$	-5.337***	-6.310, -4.363	-4.892***	-5.192,-4.591	-5.228***	-5.458,-4.997	-6.713***	-6.940,-6.487	-6.830***	-7.059,-6.600	-6.957***	-7.200,-6.713
σ_u	0.138		0.000		0.000		0.001		0.000		0.001	
σ_v	0.069		0.087		0.073		0.035		0.033		0.031	
σ^2	0.024		0.008		0.005		0.001		0.001		0.000	
λ	1.989		0.006		0.003		0.017		0.009		0.020	
LR test statistic H_0: $\sigma_u = 0$	1.802		0.000		0.000		0.000		0.000		0.000	
Prob>=chibar2	0.090		1.000		1.000		1.000		1.000		1.000	
CE	0.89704		0.99960		0.99985		0.99954		0.99974		0.99940	
Fit of cost function												
Loglikelihood	123.765		155.055		180.425		292.568		301.376		310.941	
Wald Chi-squared	6284.983		11212.042		15750.023		70074.802		78765.147		89423.871	
Prob>Chi2	0.000		0.000		0.000		0.000		0.000		0.000	
Number of observations	151		151		151		151		151		151	
Degrees of freedom	147		140		133		126		119		112	
Convergence (1=Yes)	1		1		1		1		1		1	

* p<.1, ** p<.05, *** p<.01

Flexible form   The coefficients of the quadratic terms of the PIs of within 90 day mortality is positive and significant (p<0.01). The coefficient of the quadratic terms of the performance indicators of sepsis (OECD PSI-13) is positive and significant and the coefficient for the quadratic term of the performance indicator of obstetric trauma (OECD PSI-18) is negative and significant (p<0.05)[16]. The coefficient of the performance indicator of sepsis is not significant which makes it very difficult to interpret what the relationship between sepsis and costs would look like. Yet, the flexible function of cost to 90-day mortality and cost to obstetric trauma (OECD PSI-18) can be derived from these coefficients (Table 8). The coefficient of the PI of 90-day mortality is -3.478 and the coefficient of the quadratic term is 1.665 (Table 8). The coefficient of the PI of obstetric trauma is 0.179 and the coefficient of the quadratic term is -0.051 (Table 8).

The relationship between costs and the PI of 90-day mortality could be written as $\ln Costs = constant - 3.478 * (PI_{mort90last}) + 1.665 * (PI_{mortlast})^2$ and is visualized in red in figure 2-(a). This function is convex to the origin, and the minimum point is at a performance indicator for mortality within 90 days of ca. 1.044[17]. In figure 2-(a) this optimal point is marked together with the point on the curve where the mean level of PI as reported in table 7 is set, in this case a PI of 0.966. In figure 2-(b) the relationship of cost and quality is displayed in blue. The blue form is the relationship between costs and quality based on the inverse of PI of 90-day mortality.

The relationship between costs and the PI of obstetric trauma could be written as $\ln Costs = constant + 0.179 * (PI_{obstetrictrauma}) - 0.051 * (PI_{obstetrictrauma})^2$ and is visualized in red in figure 3-(a) and the relationship of costs and quality based on the inverse of PI of obstetric trauma is displayed in blue in figure 3-(b)[18]. This function

---

[16]The coefficient for the quadratic term of the PI of pulmonary embolism/deep venous thrombosis is negative and significant if the PI of mortality within 365 days is used, instead of PI of mortality within 90 days as quality variable

[17]This is derived by solving the optimization problem, setting the first order condition (FOC) equal to zero.$-3.478 + 2 * 1.665 * (PI_{mort90last}) = 0$, leading to $PI_{mort90last} = \frac{3.478}{(2*1.665)} = 1.044$

[18]The minimum level of the PI of obstetric trauma is zero as in some hospital observations there are zero observed cases of obstetric trauma (Table 7). Taking the inverse of zero is impossible, and could indicate a quality level approaching infinity. This is incorrect. Therefore, in order to compute figure 3 (a) and (b) adequately, the observations with zero observed cases of obstetric trauma are omitted (N=130).

(a)   (b)
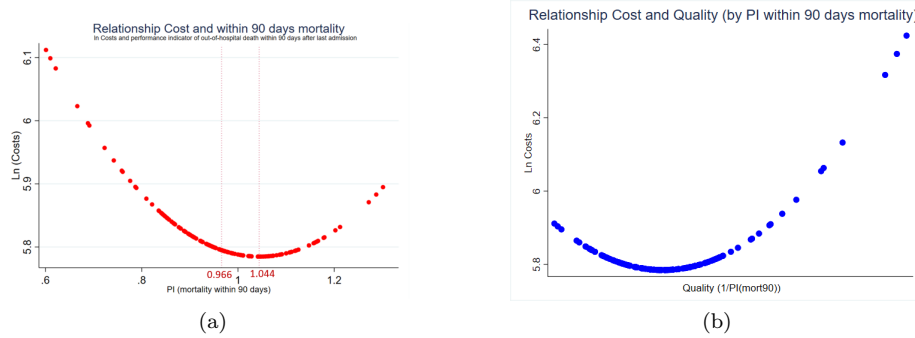
Figure 2: Relationship Cost and Quality by PI of 90-day mortality

(a)   (b)
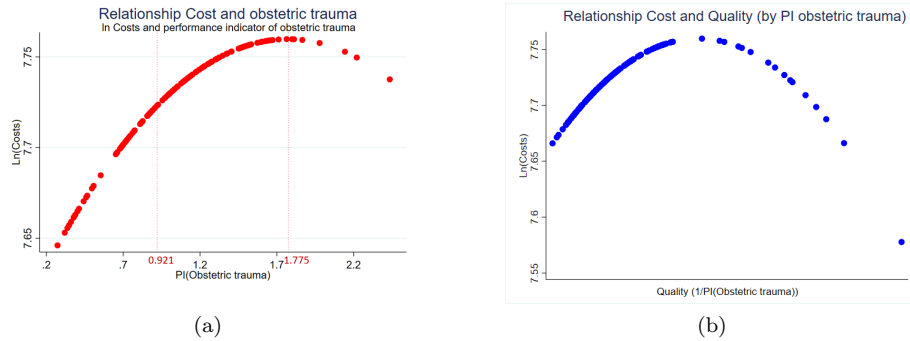
Figure 3: Relationship Cost and Quality by PI of obstetric trauma

is concave to the origin, and the maximum point is at a PI for obstetric trauma of ca. 1.755 [19].

5.3   True fixed effects model C

In this section the results of the panel estimation of Model C_A, C_B and C_F are discussed. By estimating these models with the true fixed effects method, the longitudinal aspect of the data is respected (See section 4.1.3. 'Respecting the longitudinal aspect of the data' and table 5). This panel data method assumes that the levels of the outcome variable (operating costs) and the levels of the explanatory variables (quan-

---

[19]This is derived by solving the optimization problem, setting the first order condition (FOC) equal to zero. $0.179 - 2 * 0.051 * (PI_{obstetrictrauma}) = 0$, leading to $PI_{obstetrictrauma} = \frac{0.179}{(2*0.051)} = 1.755$

titative and qualitative output) change over time within and between the units that are observed multiple times over time. Appendix E confirms that this assumption is met by showing that each variable of use is having a coefficient for within and between variation that is not zero.

One dummy is omitted from the analysis to overcome multicollinearity problems. The omitted dummy is the dummy for Oslo University Hospital (OUS). The three columns on the right side of table 8 display the estimations based on the panel data models.

The Wald test on the fit of cost function shows a p-value smaller than 0.0001. The log likelihood of the model is increasing from model C_A to C_F from 292.568 to 310.941. The degrees of freedom decrease from 126 to 112 from model C_A to model C_F.

In all three panel data models, the coefficients of the natural logarithm of the DRG-points for day patients, polyclinic patients, emergency patients and elective patients are positive and significant at the 1%-significance level ($p<0.01$).
Significance of coefficients of qualitative variables differ between model C_B and C_F. The coefficient of the performance indicators of sepsis (OECD PSI-13) is positive and significantly different from zero at the 5%-significance level ($p<0.05$) in model C_B. This is the only significant coefficient of in total the seven quality variables in model C_B. The coefficients of the PI of emergency readmission within 30 days after discharge and the PI of pulmonary embolism/ deep vein thrombosis (OECD PSI-12) are positive and significantly different from zero in model C_F.

Flexible form   The coefficient of the quadratic term of the performance indicator of pulmonary embolism/ deep vein thrombosis (OECD PSI-12) is negative and significantly different from zero at the 1%-significance level in model C_F. The coefficients of the PI of pulmonary embolism/deep vein thrombosis and the quadratic term are 0.458 and -0.211 respectively.

The relationship between costs and the PI of pulmonary embolism/ deep vein thrombosis could be written as $\ln Costs = constant + 0.458 * (PI_{PE/DVT}) - 0.211 *$
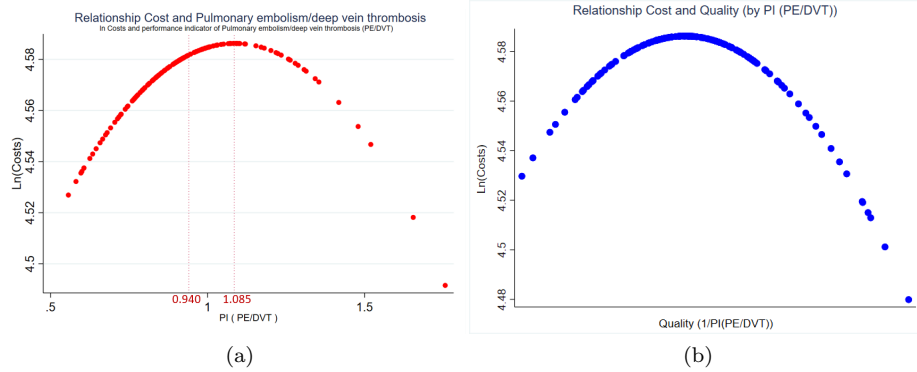
Figure 4: Relationship Cost and Quality by PI of pulmonary embolism/ deep vein thrombosis

$(PI_{PE/DVT})^2$ and is visualized in red figure 4-(a). This function is concave to the origin, and the maximum point is at a performance indicator for pulmonary embolism/ deep vein thrombosis of ca. 1.085[20]. The relationship of costs and quality based on the inverse of PI of pulmonary embolism/deep venous thrombosis is displayed in blue in figure 4-(b).

Dummy variables   Table 9 shows the coefficients of the dummy variables of model C_A, C_B and C_F that were left out from table 8. The omitted dummy is Oslo University Hospital (OUS). The signs of the significant coefficients of the dummies are negative with respect to this omitted dummy. A significant negative coefficient of a dummy means that the hospital belonging to this dummy has, ceteris paribus, relatively lower operating costs than the hospital belonging to the reference, omitted dummy. The coefficients of the dummy variables indicate the impact of the unobserved differences between the hospitals on the dependent variable, operating costs.

The interpretation of the significant coefficient -0.369 for the dummy for Vestre-viken based on model C_B would be as following: When considering the operating costs for the hospital in Vestreviken, the dummy for Vestreviken will be one and the other

---

[20]This is derived by solving the optimization problem, setting the first order condition (FOC) equal to zero. $0.458 - 2*0.211*(PI_{PE/DVT}) = 0$, leading to $PI_{PE/DVT} = \frac{0.458}{(2*0.211)} = 1.085$

dummies will be zero. Ceteris paribus, the hospital specific effect of Vestre Viken hospital is correlated with a 36.9% lower operating costs compared to the operating costs of Oslo University hospital. There are more significant coefficients of dummy variables in model C_A than in model C_B and there are more significant coefficients of dummy variables in model C_B than in model C_F. This indicates the size of unobserved heterogeneity is bigger in models A/C_A including only quantitative output variables than in models B/C_B or F/C_F including qualitative output variables as well.

Table 9: Estimated coefficients of dummy variables from true fixed effects models.

	Model C_A		Model C_B		Model C_F	
	Coefficient	*95% CI*	*Coefficient*	*95% CI*	*Coefficient*	*95% CI*
Dummy Vestreviken	−0·355***	-0.421,-0.290	−0·369***	-0.450,-0.288	−0·313***	-0.411,-0.216
Dummy Akerhus	−0·185***	-0.261,-0.109	−0·194***	-0.290,-0.098	−0·150***	-0.261,-0.040
Dummy Innlandet	−0·417***	-0.480,-0.355	−0·417***	-0.499,-0.335	−0·378***	-0.472,-0.283
Dummy Østfold	−0·334***	-0.428,-0.241	−0·329***	-0.442,-0.215	−0·281***	-0.404,-0.157
Dummy Vestfold	−0·420***	-0.510,-0.330	−0·388***	-0.482,-0.294	−0·326***	-0.434,-0.218
Dummy Telemark	−0·355***	-0.457,-0.252	−0·325***	-0.435,-0.214	−0·238***	-0.372,-0.105
Dummy Sørlandet	−0·407***	-0.482,-0.333	−0·392***	-0.494,-0.290	−0·334***	-0.447,-0.222
Dummy Stavanger	−0·360***	-0.438,-0.282	−0·327***	-0.418,-0.237	−0·267***	-0.375,-0.158
Dummy Fonna	−0·370***	-0.483,-0.257	−0·337***	-0.463,-0.211	−0·248***	-0.394,-0.103
Dummy Bergen	−0·173***	-0.225,-0.120	−0·163***	-0.220,-0.106	−0·124***	-0.192,-0.057
Dummy Førde	−0·210***	-0.350,-0.070	−0·151*	-0.316,0.013	−0·033	-0.216,0.151
Dummy Nord-Trøndelag	−0·294***	-0.418,-0.170	−0·227***	-0.362,-0.091	−0·106	-0.263,0.051
Dummy St. Olavs sykehus Trondheim	−0·194***	-0.254,-0.134	−0·181***	-0.251,-0.110	−0·124***	-0.213,-0.035
Dummy Møre og Romsdal	−0·279***	-0.389,-0.169	−0·212***	-0.338,-0.087	−0·120	-0.266,0.026
Dummy Finnmark	−0·036	-0.213,0.142	0·016	-0.184,0.215	0·152	-0.071,0.375
Dummy UNN	−0·029	-0.112,0.054	−0·013	-0.110,0.083	0·058	-0.056,0.172
Dummy Nordland	−0·117*	-0.235,0.000	−0·076	-0.204,0.053	0·010	-0.138,0.158
Dummy Helgeland	−0·209**	-0.380,-0.039	−0·136	-0.327,0.054	−0·018	-0.228,0.192
Dummy Diakonsykehus	−0·180**	-0.350,-0.011	−0·118	-0.297,0.060	0·010	-0.186,0.205
Dummy Lovisenberg	−0·182**	-0.350,-0.014	−0·121	-0.300,0.058	0·024	-0.169,0.218
Dummy Haraldsplass	−0·121	-0.297,0.055	−0·066	-0.254,0.123	0·073	-0.134,0.279

Omitted dummy variable: OUS

5.4   Cost efficiency scores

'CE' in table 8 displays the average overall cost efficiency score for every model. Again, a 5%-significance level is set. Overall, average cost efficiency based on model A is 89.7%.

$\ln (\sigma_u)^2$, the natural logarithm of the variance of the inefficiency term u, is -3.961 (95% CI -4.800- -3.123) with a p-value smaller than 0.01. This indicates that there is inefficiency. σ_u and σ_ν display the standard deviation of the inefficiency term u and idiosyncratic term ν respectively. σ^2 is the total variance of the error term. σ_u is 0.138 and σ_ν is 0.069, leading to a λ of 1.990 (95% CI 1.901-2.078). This implies that the proportion of inefficiency in the error term ϵ is bigger than the proportion of random noise.

The Loglikelihoodratio-test (LR-test) tests the null hypothesis (H_0) that σ_u, the standard deviation of inefficiency term u, is zero. The p-value of this test is 0.090. Therefore, the H_0 σ_u=0 cannot be rejected in model A at the 5%-significance level.

Overall, average cost efficiency in model B is 99.96% and average cost efficiency in model F is 99.99%. In model B, F, C_A C_B and C_F, the inefficiency term is not significantly different from zero. $\ln (\sigma_u)^2$, the natural logarithm of the variance of the inefficiency term u, is not significantly different from zero. In addition, λ is < 1 meaning that the proportion of σ_u is smaller than σ_v. The LR-test for inefficiency in model B and F on H_0 σ_u=0 fails to reject the null-hypothesis at the 5%-significance level. Thus, there is no significant inefficiency found in these models.

Tabel 10 shows the overall cost efficiency score per model and the average cost efficiency scores for every hospital over 7 years at four decimals. In model A average cost efficiency scores seem to differ between hospitals. Ranking of hospitals based on the average cost efficiency scores of model A is shown in table 11. Ranking based on the average cost efficiency scores of the other, more complex models is not performed because in these models no significant inefficiency term was found (see table 8). In addition, the cost efficiency scores in the more complex models are fairly close to one and there is no difference between hospitals at the four decimal level (see table 10).

Table 10: Average cost efficiency scores 2008-2014 per hospital

Hospital code	Model A	Model B	Model F	Model C_A	Model C_B	Model C_F
1	0.7896	0.9996	0.9999	0.9995	0.9997	0.9994
2	0.9255	0.9996	0.9999	0.9995	0.9997	0.9994
5	0.8271	0.9996	0.9999	0.9995	0.9997	0.9994
11	0.9487	0.9996	0.9999	0.9995	0.9997	0.9994
12	0.8954	0.9996	0.9999	0.9995	0.9997	0.9994
23	0.9673	0.9996	0.9999	0.9995	0.9997	0.9994
24	0.9314	0.9996	0.9999	0.9995	0.9997	0.9994
25	0.9525	0.9996	0.9999	0.9995	0.9997	0.9994
26	0.9190	0.9996	0.9999	0.9995	0.9997	0.9994
27	0.9423	0.9996	0.9999	0.9995	0.9997	0.9994
28	0.8916	0.9996	0.9999	0.9995	0.9997	0.9994
29	0.9004	0.9996	0.9999	0.9995	0.9997	0.9994
30	0.9304	0.9996	0.9999	0.9995	0.9997	0.9994
34	0.8875	0.9996	0.9999	0.9995	0.9997	0.9994
35	0.9408	0.9996	0.9999	0.9995	0.9997	0.9994
36	0.7933	0.9996	0.9999	0.9995	0.9997	0.9994
37	0.7923	0.9996	0.9999	0.9995	0.9997	0.9994
39	0.8463	0.9996	0.9999	0.9995	0.9997	0.9994
40	0.8925	0.9996	0.9999	0.9995	0.9997	0.9994
51	0.9341	0.9996	0.9999	0.9995	0.9997	0.9994
52	0.9336	0.9996	0.9999	0.9995	0.9997	0.9994
54	0.8881	0.9996	0.9999	0.9995	0.9997	0.9994
Overall	0.8970	0.9996	0.9999	0.9995	0.9997	0.9994

Table 11: Ranking hospitals on cost efficiency

Based on average cost efficiency scores 2008-2014 per hospital model A		
Hospital code	Model A	rank A
1	0.7896	22
2	0.9255	10
5	0.8271	19
11	0.9487	3
12	0.8954	13
23	0.9673	1
24	0.9314	8
25	0.9525	2
26	0.9190	11
27	0.9423	4
28	0.8916	15
29	0.9004	12
30	0.9304	9
34	0.8875	17
35	0.9408	5
36	0.7933	20
37	0.7923	21
39	0.8463	18
40	0.8925	14
51	0.9341	6
52	0.9336	7
54	0.8881	16
Overall	0.8970	

Change over time   In figure 5 the change of the cost efficiency scores per hospital over time 2008-2014 are shown for model A, B, C_A and C_B. Every color displays an unique hospital unit. The y-axis displays the cost efficiency score. As discussed before, only model A has a significant cost inefficiency term. The range of cost efficiency levels over time in model A on the y-axis of figure 5-(a) is between 0.750 to 1.000, while the range of cost efficiency levels for the other models are somewhere between 0.9995 and 0.9997. The cost efficiency scores of model B, C_A and C_B are closely approximating one. The cost efficiency scores of these models are calculated based on inefficiency terms that are not significantly different from zero. Therefore, it is not possible to rank the hospitals as there is done for the results of model A in table 11. Nonetheless, in figure 5-(b) the change of cost efficiency levels over time of each hospital based on model B are visualized. Each line represents one of the hospitals. The lines representing hospitals at the bottom do not cross with the lines representing hospitals at top with the highest efficiency score. So, there seems to be some sort of ranking between hospitals over time. There is a smaller margin between the lines representing the hospitals based on panel data model C_A and C_B. In addition, lines are crossing each other. Therefore, there is no clear visual difference between any of the hospital cost efficiency lines in figure 5-(d).
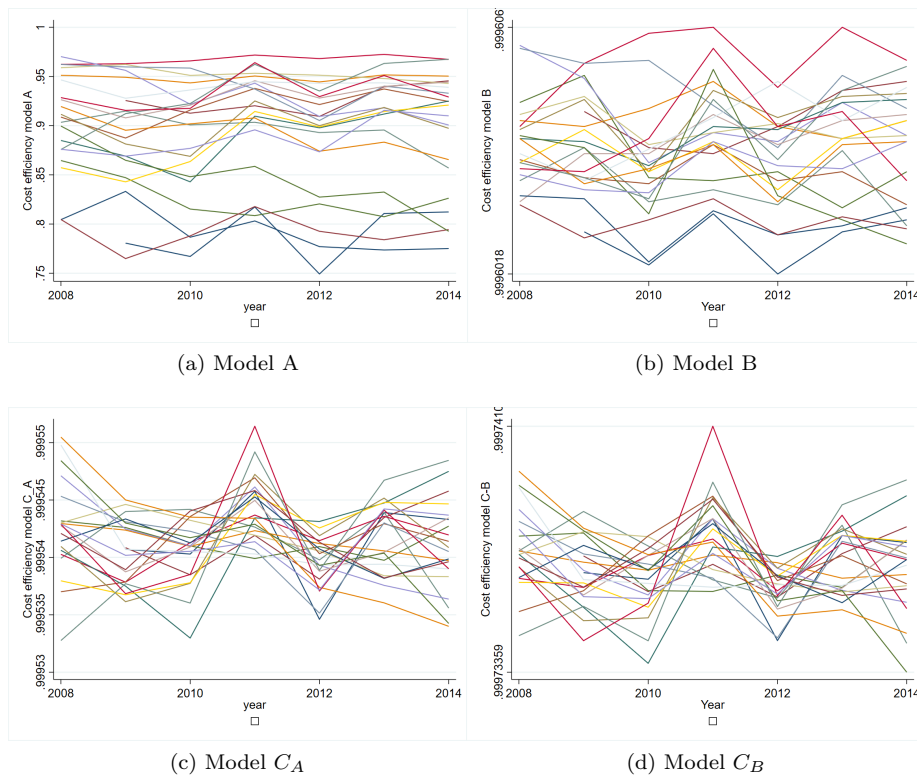
(a) Model A

(b) Model B

(c) Model C_A

(d) Model C_B

Figure 5: Change in the cost efficiency score per hospital based on model A, B and C_A and C_B over 2008-2014

6 Discussion

The research objective to find the relationship between cost and quality in the Norwegian hospital setting is tested by estimating several cross-section and panel data stochastic frontier cost functions. The cost function includes both quantitative and qualitative explanatory variables, and quadratic terms to allow for flexibility. The estimated models have a good fit (p-value Wald Chi-squared < 0.001) which is expected of cost functions since they generally have a good fit due to the influence that hospital size has on total operating costs. Therefore, it can be concluded that the independent variables in the models explain a significant part of the value of the dependent variable, which is operating costs. The results provide insight in the existence of a cost/quality trade-off and U-shaped relationship between cost and quality in the Norwegian hospital setting between 2008 and 2014. Before discussing the most important findings on the cost/quality trade-off and flexible relationship between cost and quality, another important finding is discussed.

6.1 Unobserved heterogeneity

The estimated set of coefficients of model A compared to C_A, model B to C_B and model F to C_F, are not similar in size, direction and/or significance. Furthermore, (part of) the coefficients of the hospital specific dummies in the panel models are significantly different from zero. One of the reasons of the inconsistency of results between the cross-section and the panel models could be the presence of time constant unobserved heterogeneity. In that case, the estimated coefficients in cross-section model A, B and F are affected by time constant unobserved heterogeneity, which indicates that the estimated results from cross-section models A, B and F are subject to omitted variable bias. Hence, the correlation that might be indicated by the estimated coefficients of cross section models A, B and F should not be interpreted causally. Model A, including only quantitative output variables, seems to have the highest level of omitted variable bias as most coefficients of the dummy variables are significant in model C_A compared

to C_B and C_F. This is plausible and can possibly be explained because qualitative output variables are not included in model A. An interesting insight based on the coefficients of the dummy variables from table 9, is that the significant coefficients of dummy variables are negative. This indicates that, all else being equal, hospitals represented by these dummy variables have a significantly lower overall operating cost than the hospital belonging to the reference dummy (OUS) due to a hospital specific effect. Thus, there is time constant unobserved difference between hospitals in the data affecting operating costs in the cross-section model. There it has not been corrected for these time constant unobserved differences captured in the coefficients of the dummy variables of models C. The value of the coefficients of the dummy variables indicate how much impact the unobserved effect or difference has on the dependent variable.

There are multiple tenable causes of finding unobserved heterogeneity between hospitals that indicates omitted variable bias in the cross-section models. Firstly, it could be caused by inadequate case-mix adjustments on the health outcome quality indicators. Unfortunately, there is no statistical test that can test whether case-mix adjustments are complete. One explanation for inadequate case-mix adjustments could be a difference between hospitals in complexity of cases within each DRG-type. Typically, more complex cases within each DRG-type are admitted or treated in a university hospital. This hospital with more severe or complex cases within the DRG-type will come out worse than they might be due to within DRG variation of severity of disease.

Secondly, institutional culture influences the performance and quality of care (Mannion & Davies, 2018; Jacobs et al., 2012). Institutional culture is difficult to define. Mannion and Davies (2018) define organizational culture as "the shared ways of thinking, feeling, and behaving in healthcare organisations". There is no correction for institutional culture in the models. Thus, all existing differences in institutional cultures between hospitals that are constant over time, are captured in the unobserved heterogeneity.

Thirdly, the model specification that the inefficiency term is time varying could lead to the capture of time invariant inefficiency in time constant α_i. In this case the value of α_i could be wrongly interpreted as unobserved heterogeneity, while it actually

is inefficiency. The time persistent differences between hospitals might be expressions of inefficiency. In this case unobserved heterogeneity can be overestimated, while the inefficiency term is underestimated. The underestimation of inefficiency in the true fixed effects model is further discussed in 6.3. 'Cost efficiency'.

Of course, it is important to note that also in the panel data models omitted variable bias could occur in presence of time varying unobserved heterogeneity.

6.2   Cost and Quality

In this section, the overall evidence on the cost/quality relationship is discussed. First, a general note is made on the inconsistency between the coefficients over the models. Then, evidence on both a negative and a positive relationship between cost and quality is discussed. Afterwards, the existence of flexible relationships is discussed. Finally, the coefficients of the quantitative variables are discussed.

Overall, there is inconsistency in sign, level and significance of the coefficients of the qualitative variables between the cross-section and panel models (A, B and F vs. C_A, C_B and C_F) and between the non-flexible and flexible models (B and C_B vs. F and C_F). The coefficients of the variables in the panel models, estimated through the true fixed effects method, are solely driven by the variation of the variables within hospitals (*after this: within variation*) over time, while the variation between hospitals is captured in the fixed effects coefficient α_i. In the cross-section models the estimated coefficients are driven by between and within variation. So, between variation is not captured in the coefficients of the quantitative and qualitative variables that are estimated by the panel models. This difference in capturing within or both within and between variation between the panel and cross-section models, could possibly be another explanation of the inconsistency in the sign, level and significance of the estimated coefficients of the variables.

The most consistent result on a cost/quality correlation is the positive and significant coefficient for the correlation between costs and the patient safety indicator pulmonary embolism/deep venous thrombosis. This result is found in model B, C_B and C_F, so both in cross-section and panel models. As a lower value of PI of pul-

monary embolism/deep venous thrombosis is interpreted as a higher quality of care, cost and quality based on the PI of pulmonary embolism/deep venous thrombosis are negatively correlated. This contradicts the existence of a cost/quality trade-off. It indicates that a higher quality of care, which is a lower number of cases of pulmonary embolism/deep venous thrombosis, is correlated with lower operating costs. In addition, the coefficients of the qualitative output variable for sepsis in model B and C_B and the coefficients of the qualitative output variable for obstetric trauma in model B and F indicate a negative association between quality and costs. The negative correlation between cost and quality based on these quality variables are found in more than one of the models. Still, the discussed estimated coefficients of these variables are not consistent over all the models (B, F, C_B and C_F).

If causal inference could have been drawn from this model, the coefficients of these qualitative output variables could have been practically interpreted as a decrease of the performance indicator by a value of one to be correlated with a decrease of operating costs by 100% times the value of the coefficient of the variable of interest. In order for the ratio of observed to expected number of cases of sepsis to decrease with one, the observed cases of sepsis need to decrease with the absolute value of the expected cases. It would be more practical and realistic to interpret the coefficient as a decrease of PI of sepsis of 0.1 to be correlated with a decrease of operating costs by 10% times the estimated coefficients of the variable of interest. E.g. interpreting the coefficients, a decrease of the PI of sepsis by 0.1, which is an increase in quality, is correlated with a decrease of operating costs of 0.36% (based on model C_B).

For pulmonary embolism/deep venous thrombosis a drop of the PI by 0.1 is correlated with a decrease of operating costs of 4.58% (based on model C_F). A decrease of PI is an increase in quality as for a decrease in PI the observed number of cases decreases relatively to the expected number of cases. The mean observed number of patients with pulmonary embolism/deep venous thrombosis (198) is relatively low compared to the mean number of patients (264,444) (see table 7). Thus, 4.58% seems to be a relatively big impact on operating costs for this relatively small group of patients with pulmonary embolism/deep venous thrombosis.

The coefficient of the PI of sepsis is not significant in models F or C_F. Nor are the coefficients of the PI of bed sores and acute puncture. Another argument for this inconsistency would be that there is a lack of degrees of freedom in models F and C_F, as this number decreased by 7 after adding the quadratic terms. This loss of power for the model could make it difficult to estimate significant coefficients for relatively rare adverse health events like sepsis, bed sores, and acute puncture, which have a low number of cases compared to the number of patients. The coefficients of the PI of emergency readmission within 30 days are not significantly different from zero in cross-section models or in model C_B. But, in model C_F the estimated coefficient for this variable is positive and significant, indicating a positive relationship between operating cost and the PI of emergency readmission within 30 days. This hints at a negative relationship between cost and quality based on the PI of emergency readmission within 30 days. One argument for the difference between the models, is that the estimated coefficients in cross section model B and F are subject to omitted variable bias caused by time constant unobserved heterogeneity. This does not explain, however, why the coefficient of the PI of emergency readmission within 30 days is not significant in model C_B, but it is in model C_F.

Other inconsistent, but interesting results on the existence of a cost/quality trade-off are briefly discussed here. Firstly, the estimated coefficient of PI for within 90-day mortality is negative and significant at the 1%-significance level in cross-section model F. This indicates a possible cost/quality trade-off between cost and quality based on within 90-day mortality. Further on in the discussion there is a more extensive elaboration about the flexible relationship between cost and quality based on within 90-day mortality. Secondly, in model B a negative, significant coefficient of the PI of bed sores is estimated. This could indicate a possible cost/quality trade-off. In neither the flexible model F nor panel models the estimated coefficient for PI of bed sores are significant.

Estimates from model C_F on the PI of pulmonary embolism/deep venous thrombosis indicate an inverted U-shaped relationship between quality and costs with the turning point at a PI of 1.085 (figure 4). This means that hospitals with a PI lower

than 1.085 and thus higher quality of care, will operate in an area where there is a negative relationship between cost and quality, while hospitals with a PI higher than 1.085 and thus a lower quality of care, will operate in an area where there is a positive relationship between cost and quality. The coefficient of PI of pulmonary embolism/deep venous thrombosis on costs is positive, and the mean PI of pulmonary embolism/deep venous thrombosis is with 0.940 lower than the PI of 1.085 at the turning point. This indicates that the mean of the hospitals operate in an area of marginal returns where there is a negative relationship between quality and costs.

A similar inverted U-shaped relationship between quality and costs with a turning point at a PI of 1.755 is found based on the estimated coefficient of the PI of obstetric trauma and the quadratic term of this in model B. The coefficient of the PI of obstetric trauma on costs is positive with a mean PI of 0.921 being lower than the PI of 1.755 at the turning point. Again, this indicates that the mean of the hospitals operate in an area of marginal returns where there is a negative relationship between quality and costs.

To explain the finding of this inverted U-shaped relationship between quality and costs based on these two patient safety indicators (obstetric trauma and pulmonary embolism/deep venous thrombosis), I reflect on Fleming's findings that were discussed in the theoretical framework (Fleming, 1991). It could be theorized based on Fleming's model that hospitals currently operate in an area of marginal return, like area C in Fleming's model (figure 1-(a)), and that the inverted U-shaped relationships from figure 3 and 4 are due to hospitals operating at a moderate level of quality. This same concept is seen areas B and C in Fleming's model, where quality is not poor, but not at its best either. It is unclear if there could be a second turning point, like area D from Fleming's cubic model. In order to research this, a third-degree function needs to be estimated to allow for a cubic relationship. This could be an interesting concept to look into in future research.

The relationship between cost and quality based on 90-day mortality is U-shaped. As discussed before the coefficient PI for 90-day mortality is not significant in model B, but in model F, the coefficients for both the PI and the quadratic term for 90-day

mortality are significant at the 1%-significance level. The convex functional form has a turning point at a PI of 90-day mortality of 1.044. The negative sign of the coefficient of the PI of 90-day mortality indicates a negative association between cost and 90-day mortality. This could be interpreted as a positive association between cost and quality, and is a cost/quality trade-off. The mean level of the PI of 90-day mortality is 0.966. At this level of PI of 90-day mortality, there is a negative derivative in figure 2-(a). Since PI and quality are inversely correlated, this indicates again positive marginal costs for the next unit of quality. Therefore, a cost/quality trade-off is found for quality based on the PI of 90-day mortality.

However, it also indicates that the lower the quality, so the higher the PI for 90-day mortality, the cost/quality trade-off turns at some point into an area of marginal returns. In the area of marginal returns costs can be reduced by increasing quality. The turn-point of the cost/quality trade-off for 90-day mortality and costs is at a performance indicator of 1.044. If the PI becomes higher than that, the quality becomes worse, and the curve is in the area of marginal returns, whereby increasing quality, costs could be reduced.

The quality is interpreted to worsen when the PI increases. The PI increases if the number of observed cases increases relatively to the expected number of cases. Unfortunately, it is difficult to interpret what practical implications for each individual hospital the change in cost/quality trade-off has over the value of PI over the graph and what implication the turning point has. The level of PI depends on ratio between the observed number of cases and the expected number of cases of death within 90 days. The number of cases in this ratio are hospital specific. So, the PI ratio and the number of cases that the cost/quality trade-off is based on, varies per hospital. Practical interpretation of the cost/quality trade-off could be made based on the observed and predicted number of cases of each hospital. However, this is not advisable because there is uncertainty in the model specification and causal inference cannot be drawn from the models of use.

The U-shaped relationship between cost and quality based on the PI of 90-day mortality, and the inverted U-shaped relationship between cost and quality based on

the PI of obstetric trauma, are based on estimations by cross-section model F. Model F is the only model where the coefficients of the qualitative variables of 90-day mortality and obstetric trauma are significantly different from zero. As stated earlier, it is important to be aware that the estimated results of the cross-section model F could be subject to omitted variable bias.

A general note about figures 2, 3 and 4 is that they visualize the relationship between cost and the quality variables (relatively 90-day mortality, obstetric trauma and pulmonary embolism/deep venous thrombosis) where the value of the intercept is included in the calculations and the level of all other variables is set to zero. It is unrealistic that the levels of the other variables are zero. Therefore, the levels of operating costs cannot be interpreted from these figures. However, the change and shape of the relationship is not affected and can be interpreted. To get an better insight on the real levels of operating costs on the y-axis, the mean level of each of the other variables could be inserted in the calculations of the relationship between cost and the quality variables (relatively 90-day mortality, obstetric trauma and pulmonary embolism/deep venous thrombosis).

Finally, the coefficients of the quantitative output differ over the 6 models. As discussed earlier, comparing cross-section models to panel models, it can be noted that the coefficients of models A, B, and F are different from these same models in panel data analysis (C_A, C_B and C_F). Comparing the models including only quantitative to the ones including both qualitative and quantitative variables, the levels of the coefficients of the quantitative variables are also changing over the different models A to B and F or C_A to C_B and C_F. The fixed effects are not captured in the coefficients in the panel data models, where coefficients are solely driven by the within variation. Therefore, the coefficients for the quantitative output variables of the panel model could capture aspects of the long-run cost functions of hospitals compared to the coefficients of the cross-section model capturing aspects of the short-run cost functions. Also, there is a collinearity problem between the quantitative output variables which leads to a less precise in the estimated coefficients (See appendix C for Pearson correlation coefficients between quantitative variables and operating costs).

The value share of a quantitative variable over the output should be aligning with the estimated coefficient for this variable because DRG weights are directly used to value the output produced, and are thus directly linked to costs. In this case, the value shares in DRG-points are 3,681 for day patients, 9,108 for polyclinic patients, 26,501 for emergency patients, and 16,668 for elective patients (Table 7). This indicates that the coefficient for emergency patients should be highest and the coefficient for day patients should be lowest as these are the variables with, respectively, the highest and lowest value share of output. As shown in table 8, the coefficients for the quantitative variables are not aligning with the value shares of output. In model A, B, and F the coefficient for DRG-points of polyclinic patients is lowest and the coefficient for DRG-points of elective patients is highest. Also, in model C_A, C_B and C_F the coefficient for DRG-points of day patients is highest, while it is expected to be lowest given the value shares of output in DRG-points (Table 8).

If causal inference could have been drawn from this model, a careful interpretation of the estimated results from model C_B could for example be that a 1 percent increase of the DRG-points of day patients, polyclinic patients, emergency patients and elective patients is associated with, respectively, a 0.387, 0.112, 0.342, 0.189 percent increase of total operating costs in kNOK (table 8).

6.3   Cost efficiency

The tests on the significance of the inefficiency term show that inefficiency is found in model A, but not in the other models. The insignificant inefficiency terms of models B, F, C_B and C_F indicate that the estimation of an average cost function would have been sufficient and estimate the same coefficients as the stochastic frontier models.

An interesting finding is that even though there was no significant inefficiency found in model B, in figure 5 there seems to be some influence or effect that is keeping the lines displaying the upper hospital efficiency scores up and the lines displaying the lower hospital efficiency scores down. Based on this figure, it can be concluded that it appears that there is something going on that effects a differentiation between the hospitals. As this figure is based on model B, this could be the unobserved heterogeneity present

in the cross-section models, which was discussed earlier.

There are several explanations for finding an insignificant inefficiency term for the more complex models. It is important to note that it is reasonable to assume that in real life there is inefficiency to some extent, and that results of insignificant inefficiency terms are driven by the model specifications. Firstly, the cost function could be over specified. Overfitting could lead to bias in the estimates of the inefficiency term. The independent variables explain a great part of the costs while only an insignificantly small part of the costs are explained by inefficiency. In addition, quality aspects could be captured by the significant inefficiency term of model A, while in model B qualitative variables are added as explanatory variables making the quality aspects disappear from the inefficiency term. Yet, this does not explain why the inefficiency term appears insignificant in model C_A compared to a significant inefficiency term in model A. This can be explained by the specifications of the true fixed effects model. The true fixed effects model by Greene captures time constant inefficiency in the fixed effects term or individual specific constant term (α_i) which leads to an underestimation of inefficiency. In addition, time constant unobserved heterogeneity could be captured in the inefficiency term in the cross section models while in the true fixed effects model only time varying unobserved heterogeneity can influence or be captured in the inefficiency term in the panel data models. Both could lead to over- and underestimation of the inefficiency term, depending on the direction of the effect the unobserved heterogeneity has. Lastly, assumptions on the distribution of the error term and inefficiency are restrictive and could cause bias on the estimation of inefficiency if the assumptions are not met. If in real life the distribution of the inefficiency term is not half normal as assumed, the estimates of inefficiency by the models are invalid and biased.

6.4   Comparison to other findings

The finding of a cost/quality trade-off, or positive relationship between cost and quality for 90-day mortality is in contrast with the findings by Kittelsen et al. (2015). They find that better productivity is associated with lower mortality rates, so better quality. Based on this, they interpret a 'tendency' for hospitals to have higher costs if

mortality is higher, which indicates a negative relationship between quality and costs. The result from Bakke (2019) is in line with the finding of a cost/quality trade-off for 90-day mortality in this study. However, this result could be subject to unobserved heterogeneity, just like the significant result that is based on model B in this study. Therefore, results are not comparable and conclusions cannot be drawn on the existence of a cost/quality trade-off based on 90-day mortality in the Norwegian hospital setting between 2008 and 2014.

In addition, the findings of a negative relationship between cost and quality based on PSI are not in line with the results from Kittelsen et al. (2015). They did not find any correlation between quality based on PSI and productivity, while this thesis finds negative correlations between cost and quality based on PSI pulmonary embolism/deep venous thrombosis, sepsis and obstetric trauma. However, results on PSI in Kittelsen et al. (2015) might be biased by the fact that they use data from three different Nordic countries which could make quality comparison based on PSI difficult because of differences in coding.

Carey & Stefos (2011) find that costs can increase excessively when quality measured by adverse health events becomes poor. The result indicating a negative relationship between cost and quality based on PSI, is in line with the findings of this study.

However, the finding of a negative relationship and inverted U-shaped relationship between cost and quality is in contrast with Gutacker et al. (2013). In this thesis the finding of a negative relationship and inverted U-shaped relationship between cost and quality is based on patient safety indicators. Gutacker et al. (2013) find a U-shaped relationship between quality and costs on the patient level. Quality in their study is not based on (adverse) health outcomes, but on reported quality of life based on EQ-5D and Oxford scores (Gutacker et al., 2013). They find some evidence for a U-shaped relationship between quality based on positive quality scores and patient level costs.

The finding of Hvenegaard et al. (2011) of a U-shaped relationship between department level cost and quality based on mortality, is in line with the finding of this thesis of a possible U-shaped relationship between operating cost and quality based

on 90 day-mortality. Nonetheless, the coefficient of 90-day mortality is inconsistent over the models B, F, C_B and C_F.

Hvenegaard et al. (2011) find no significant relationship between cost and quality based on the adverse health event wound sores. The evidence in this thesis on the correlation between cost and quality based on bed sores is weak. A significant, positive correlation is found in cross-section model B indicating a possible cost/quality trade-off, but there is inconsistency of the estimated coefficients over the models and it is therefore plausible that this result is biased.

This thesis finds evidence that quality of care based on patient safety indicators is negatively correlated with costs. This result has important implications for both application in practice and in research. Cost minimization strategies could be done by improving quality, and future research could be focused on quality of care that is defined by more than just mortality or readmission as an outcome indicator.

6.5   Limitations

Model assumptions   None of the model specifications allow to draw a conclusion on causal inference. As discussed in the method section, the true fixed effects will underestimate the inefficiency term. The fixed effects, or hospital specific time constant effects, are assumed not to hold inefficiency. Still, time constant inefficiency might be captured in it. In addition, a strong assumption on an equal shift of the cost functions has to stand in order to estimate quadratic terms through the true fixed effects method. Furthermore, distributional assumptions on the error and inefficiency terms are restrictive. The model is misspecified if the distributional assumptions are not met.

Moreover, the cost efficiency term is not significant. This might indicate that an average cost function estimation would lead to similar results and estimating the cost function through stochastic frontier analysis is not necessary.

Lastly, there could be more than one reason for different estimations in panel models and cross section models. As discussed before, one of the explanations is the existence of unobserved heterogeneity. However, if the above mentioned problems

arise, the differences in estimations between panel and cross section models could also be explained by estimating misspecified models or not meeting model assumptions.

In short, it is plausible that the results are driven by model specification and the belonging assumptions. This could be yet another explanation for the inconsistency in results between the different models.

Confounding and unobserved heterogeneity   Confounding could happen on two levels. Firstly, it could happen on the patient level, influencing the likelihood of getting an adverse health event and thus influencing the quality indicator. Case-mix adjustments described in the method section are done in order to be able to compare the effect of quality indicators on costs between hospitals. If there is an inadequate case-mix adjustment differences between hospitals that are not adequately adjusted, this could be confounding the estimates. However, there is no method to test if case-mix adjustments are performed adequately. A variable that is not corrected for is the impact of organizational or institutional culture on quality of care or performance of a hospital. There is some literature on the correlation between culture and performance of hospitals (Mannion & Davies, 2018; Jacobs et al., 2012). Differences in culture between hospitals could possibly explain differences in performance between hospitals. It is debatable whether it is necessary to correct for this difference. The cultural differences could explain why there are differences in efficiency scores between hospitals. However, this does not necessarily imply that culture should have been corrected for. Secondly, adequate adjustment for variables affecting costs in a hospital that are different per hospital need to be done. Factors affecting costs in a hospital could for example be the quality of community care (Dusheiko et al., 2011). Furthermore, the geographic location of a hospital is found to be an important explanatory variable for cost efficiency in Nordic hospitals (Medin et al., 2011). Also, costs are possibly influenced by scale properties (Kittelsen et al., 2018).

Endogeneity   Due to endogeneity problems, no conclusion can be drawn on causal inference in this study. First of all, behaviour of hospitals is not taken into account. However, if hospitals decide on the process of production of health care in order to

minimize costs, this causes endogeneity problems. In addition, cost and quality indicators based on adverse health outcomes are endogenous. For instance, poor physical condition could lead to early death, so low treating costs, while at the same time severe patients are more costly as resources will be used to save them (Kittelsen et al., 2015). In order to be able to draw conclusions on causal inference, an adequate correction for endogeneity needs to be done.

There is no accepted method to estimate efficiency in SFA in presence of endogeneity (Mutter et al., 2012). To what extent inefficiency estimated by SFA is affected by endogeneity in the cost function is researched by Mutter et al. (2012). Their study predicts that including endogenous variables leads to a more precise estimation of inefficiency in case that the endogeneity is expected to be due to a correlation with random noise (ν) (Mutter et al., 2012).

Future research could focus on finding instrumental variables that affect quality indicators but not costs. Factors that affect quality of care but not necessarily costs could be for example, shift length (Griffiths et al., 2014; Barger et al., 2006; Stimpfel & Aiken, 2013). According to results from Griffiths et al. (2014) day shifts are not longer than 8 hours and night shifts are not longer than 10 hours in Norway, making differences in shift length small. Another factor that could influence quality of care but not necessarily costs could be medical leadership. A Norwegian study finds significant effect of task-oriented leadership on quality of care in nursing homes (Havig et al., 2011).

Still, it remains unsure if the one or multiple instrumental variables can correct the endogeneity problem in this case completely as there is no accepted approach to correct for endogeneity while estimating efficiency in SFA yet.

Definition quality of care   Most evidence is based on quality outcome indicators. Mortality and readmission rates are widely used as quality outcome variables. However, quality of hospital care is too complex to be based on a single quality indicator (Ferrier & Trivitt, 2013). In this thesis, the scope of quality is broadened by including patient safety indicators next to mortality and readmission rates. Unfortunately, this

study allows to draw conclusions on the single effects of these quality variables on costs only. There is no clear pattern of possible interaction between quality variables found. Ferrier & Trivitt (2013) explore how different quality indicators can be combined into one aggregate quality indicator that combines all relevant quality indicators. As it appears in their study, correcting for a quality outcome indicator is more important than correcting for a quality process indicator. The relationship between a combined index of quality existing of both outcome and process quality indicators and costs might be an interesting topic for future research.

Data   The quality of the collected data can be considered good. The data is directly derived from the official national registries and patient registries of hospitals within the same country. Therefore, the way of registering data will probably be quite homogenous. However, the size of the data might be insufficient for the methods of use, in particular the number of hospitals in the data. In the ideal situation the study is repeated with a dataset that is of the same quality but includes a bigger pool of hospitals and a longer observation period. This is practically difficult as we are bound to the available data based on the number of hospitals in the country.

6.6   Policy recommendation and implications for further research

Future research   As discussed before, Fleming's article on the cubic relationship and different derivatives in areas of different quality levels, could provide an explanation for the finding of both a U-shaped and an inverted U-shaped relationship between costs and quality depending on what quality indicator is used. Therefore, it would be very interesting to exploit the research on the cost/quality relationship fully by performing a third degree function in order to allow for a cubic relationship. In addition, the relationship between a combined index of quality proposed by Ferrier & Trivitt (2013) existing of both outcome and process quality indicators and costs might be an interesting topic for future research.

In addition, a follow-up qualitative study could provide insight on what factors play

a role in this heterogeneity and how you could measure and correct for them. This could possibly help to understand institutional behaviour and causes of unobserved heterogeneity which could help to find a way to reward hospitals for good quality of care in the payment schemes.

Also, it might be interesting to set up research utilizing both DEA and SFA methods on similar data subsets to understand what the impact of the model of use is on the results that are found in the hospital setting in Norway.

Finally, there was no time dummy included in the models in this study. In future research a time dummy could be included to capture technological progression over the years or other possible time trends.

Policy recommendation Case-mix correction for observable or unobservable differences between hospitals is insufficient to be able make a precise estimation of the cost efficiency. In the Norwegian finance scheme of the hospitals, hospitals are partly reimbursed by ABF, and partly by a base payment. In addition, they are implementing the inclusion of performance measures in the payments. This study shows that time constant unobserved heterogeneity between hospitals could be present and therefore should be taken into account during the implementation of financial rewards for performance measures.

Additionally, quantitatively defined cost efficiency should not be the only instrument used to benchmark hospital efficiency. Quality explains part of the costs that would have been addressed to inefficiency in a model with only quantitative output.

An interesting finding in this study for the health authority and leaders of hospital organizations is that the patient safety indicators as quality indicators show a positive relationship between the adverse health events and costs, which can be translated into a negative relationship of the quality of care with costs. Based on this result it can be concluded that cost minimization behaviour does not necessarily has to be focused on minimizing cost, because it could be done by focusing on improving quality and thereby indirectly minimizing costs.

Be aware that rewarding for quality of care in pay-for-performance finance schemes

could lead to cream skimming and create an incentive to focus to improve what is observable and measurable, over what goes unobserved and unmeasured (Ferrier & Trivitt, 2013). It is debatable if you should be willing to measure quality of care at the macro level, as it could lead to misinterpretation and wrongly incentivize health care organizations. Evidence from this thesis on the relationship between cost and quality in the Norwegian hospital setting is weak as results are inconsistent and inconclusive. As the evidence is weak, any incentive to improve quality that might be indicated by this evidence, shall not be too strong.

6.7 Conclusion

This study investigates the relationship between cost and quality in the Norwegian hospital setting. Multiple qualitative output variables show a correlation with operating costs over the different cross-sectional and panel data models. A negative correlation was found between operating costs and quality based on 30-day readmission (in model C_F), the patient safety indicator for sepsis (in model B and C_B), the patient safety indicator for pulmonary embolism and deep venous thrombosis (in model B, F and C_F), and the patient safety indicator for obstetric trauma (in model B and F). This indicates that costs can be decreased by decreasing the 30-day emergency readmission rates and by decreasing the number of cases of the patient safety indicators obstetric trauma, pulmonary embolism/deep venous thrombosis, and sepsis. Furthermore, quality based on patient safety indicator for pulmonary embolism and deep venous thrombosis and obstetric trauma seem to have an inverted U-shaped relationship with operating costs. One possible cost/quality trade-off was found for the qualitative output variable for 90-day mortality, as well as a U-shaped relationship between cost and quality based on 90-day mortality. The results on the relationships between costs and quality based on 90-day mortality and quality based on obstetric trauma are estimated in a cross-section model F. Time constant unobserved heterogeneity between hospitals is present. There is no significant cost inefficiency found in the more complex cost models including qualitative output variables.

Despite hinting evidence on both positive and negative relationships between cost and quality in different models, there still is uncertainty on the relationship between cost and quality due to the inconsistency in estimated results on the cost/quality relationship over the different models. There are clues that the relationship between cost and quality changes over the level of quality in a non-constant way. This inconsistency in results on the cost/quality trade-off is in line with the ambiguous and conflicting evidence that is yet reported in the current literature. Model specification might be an important factor of influence in this field of research.

Practical implications for health care authorities are that unobserved heterogeneity should be considered when ranking and rewarding hospitals based on their activity and

quality. This study also indicates that health care organizations could possibly aim for cost minimization by improving quality based on patient safety indicators.

References

Anthun, K. S., Goude, F., Häkkinen, U., Kittelsen, S. A. C., Kruse, M., Medin, E., ... Rättö, H. (2012). EuroHOPE nordic hospital comparison: Material, methods and indicators. *EuroHOPE Discussion Papers No 10.*.

Anthun, K. S., Kittelsen, S. A. C., & 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. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0168851017300246` doi: https://doi.org/10.1016/j.healthpol.2017.01.006

Ash, A., Shwartz, M., Peköz, E., Hanchate, A., & Iezzoni, L. (2003). Comparing outcomes across providers. *Risk Adjustment for Measuring Health Care Outcomes*, 342-345.

Bakke, E. (2019). Quality and costs at the hospital level in Norway. *Unpublished master's thesis University of Oslo*.

Barger, L., Ayas, N., Cade, B., Cronin, J., Rosner, B., Speizer, F., & Czeisler, C. (2006). Impact of extended-duration shifts on medical errors, adverse events, and attentional failures. *PLoS medicine*, *3*, 487.

Biørn, E., Hagen, T. P., Iversen, T., & Magnussen, J. (2003). The effect of activity-based financing on hospital efficiency: A panel data analysis of dea efficiency scores 1992–2000. *Health Care Management Science*, *6*, 271-283.

Carey, K. (2003). Hospital cost efficiency and system membership. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, *40*(1), 25-38. Retrieved from `https://doi.org/10.5034/inquiryjrnl_40.1.25` (PMID: 12836906) doi: 10.5034/inquiryjrnl_40.1.25

Carey, K., & Stefos, T. (2011). Measuring the cost of hospital adverse patient safety events. *Health Economics*, *20*(12), 1417-1430. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.1680` doi: 10.1002/hec .1680

Clement, J., Valdmanis, V., Bazzoli, G., Zhao, M., & Chukmaitov, A. (2008). Is more better? an analysis of hospital outcomes and efficiency with a dea model of output congestion. *Health care management science*, *11*, 67-77. doi: 10.1007/s10729-007-9025-8

Coelli, T. J., Rao, D. P., O'Donnell, C. J., & Battese, G. E. (2005). *Review of production economics* (2nd edition ed.). New York: Springer.

Cornwell, C., Schmidt, P., & Sickles, R. C. (1990). Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics*, *46*(1), 185–200.

Deily, M. E., & McKay, N. L. (2006). Cost inefficiency and mortality rates in florida hospitals. *Health Economics*, *15*(4), 419-431. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.1078` doi: 10.1002/hec .1078

Donabedian, A. (1978). The quality of medical care. *Science*, *200*(4344), 856–864. Retrieved from `https://science.sciencemag.org/content/200/4344/856` doi: 10.1126/science.417400

Donabedian, A. (2005). Evaluating quality of medical care. *The Milbank quarterly*, *83*, 691-729. doi: 10.1111/j.1468-0009.2005.00397.x

Drösler, S., Romano, P., & Wei, L. (2009). Health care quality indicators project: Patient safety indicators report 2009. *OECD, Directorate for Employment,*

Labour and Social Affairs, OECD Health Working Papers. doi: 10.1787/ 220112312723

Dusheiko, M., Gravelle, H., Martin, S., Rice, N., & Smith, P. C. (2011). Does better disease management in primary care reduce hospital costs? evidence from english primary care. *Journal of Health Economics*, *30*(5), 919 - 932. Retrieved from `http://www.sciencedirect.com/science/article/ pii/S0167629611001093` doi: https://doi.org/10.1016/j.jhealeco.2011.08 .001

Ferrier, G. D., & Trivitt, J. (2013). Incorporating quality into the measurement of hospital efficiency: a double DEA approach. *Journal of Productivity Analysis*, *40*(3), 337-355. Retrieved from `https://ideas.repec.org/a/ kap/jproda/v40y2013i3p337-355.html` doi: 10.1007/s11123-012-0305-z

Ferrier, G. D., & Valdmanis, V. (1996). Rural hospital performance and its correlates. , *7*, 63-80. Retrieved from `https://doi.org/10.1007/BF00158477` doi: 10.1007/BF00158477

Fleming, S. T. (1991). The relationship between quality and cost: Pure and simple? *Inquiry*, *28*(1), 29–38. Retrieved from `http://www.jstor.org/ stable/29772179`

Greene, W. (2004). Distinguishing between heterogeneity and inefficiency: Stochastic frontier analysis of the world health organization's panel data on national health care systems. *Health economics*, *13*, 959-80. doi: 10.1002/ hec.938

Greene, W. (2005a). Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis*, *23*, 7-32. doi: 10.1007/s11123-004-8545-1

Greene, W. (2005b). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics*, *126*, 269-303. doi: 10.1016/j.jeconom.2004.05.003

Greene, W. (2018). *Econometric analysis* (8th ed., International ed. ed.). Boston: Pearson.

Griffiths, P., Dall'Ora, C., Simon, M., Ball, J., Lindqvist, R., Rafferty, A., ... Aiken, L. (2014). Nurses' shift length and overtime working in 12 european countries: The association with perceived quality of care and patient safety. *Medical Care*. doi: 10.1097/MLR.0000000000000233

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, *80*(2), 223–255. Retrieved from `http://www.jstor.org/stable/1830580`

Grossman, M. (2000). Chapter 7 - the human capital model. In *Handbook of health economics* (Vol. 1, p. 347 - 408). Elsevier. Retrieved from `http://www.sciencedirect.com/science/article/pii/S1574006400801663` doi: https://doi.org/10.1016/S1574-0064(00)80166-3

Gutacker, N., Bojke, C., Daidone, S., Devlin, N. J., 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. P., & Kaarboe, O. M. (2006). The Norwegian hospital reform of 2002: Central government takes over ownership of public hospitals. *Health Policy*, *76*(3), 320-333. Retrieved from `https://ideas.repec.org/a/eee/hepoli/v76y2006i3p320-333.html`

Havig, A., Skogstad, A., Kjekshus, L., & Romøren, T. (2011). Leadership, staffing and quality of care in nursing homes. *BMC health services research*, *11*, 327. doi: 10.1186/1472-6963-11-327

Helsedirektoratet. (2018). Rammeverk for nasjonalt kvalitetsindikatorsystem for helse- og omsorgstjenesten. Retrieved from `https://www.helsedirektoratet.no/rapporter`

Helsedirektoratet. (2019a). Innsatsstyrt finansiering (ISF) regelverk. Retrieved from `https://www.helsedirektoratet.no/tema/finansiering/innsatsstyrt-finansiering-og-drg-systemet/innsatsstyrt-finansiering-isf`

Helsedirektoratet. (2019b). Om kvalitet og kvalitetsindikatorer. Retrieved from `https://www.helsedirektoratet.no/statistikk/kvalitetsindikatorer/om-kvalitet-og-kvalitetsindikatorer`

Helsedirektoratet. (2020). Helsedirektoratets strategi. Retrieved from `https://www.helsedirektoratet.no/rapporter/helsedirektoratets-strategi`

Hussey, P., Wertheimer, S., & Mehrotra, A. (2013). The association between health care quality and cost. *Annals of internal medicine*, *158*, 27-34. doi: 10.7326/0003-4819-158-1-201301010-00006

Hvenegaard, A., Arendt, J., 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? *The European journal of health economics : HEPAC : health economics in prevention and care*, *12*, 541-51. doi: 10.1007/s10198-010-0268-9

Jacobs, R., Mannion, R., Davies, H., Harrison, S., Konteh, F., & Walshe, K. (2012). The relationship between organizational culture and performance

in acute hospitals. *Social science & medicine (1982)*, *76*. doi: 10.1016/ j.socscimed.2012.10.014

Jacobs, R., Smith, P. C., & Street, A. (2006). *Measuring efficiency in health care: Analytic techniques and health policy.* Cambridge University Press. doi: 10.1017/CBO9780511617492

Kittelsen, S. A., Anthun, K. S., Goude, F., Huitfeldt, I. M. S., Häkkinen, U., Kruse, M., . . . Rättö, H. (2015). Costs and quality at the hospital level in the nordic countries. *Health Economics*, *24* (S2), 140-163. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.3260` doi: 10 .1002/hec.3260

Kittelsen, S. A., Anthun, K. S., Häkkinen, U., Kruse, M., & Rehnberg, c. (2018). Scale and quality in nordic hospitals. *Nordic Journal of Health Economics*, *6*, 29-44.

Kumbhakar, S., Horncastle, A., & Wang, H.-J. (2015). *A practitioner's guide to stochastic frontier analysis using stata.* doi: 10.1017/CBO9781139342070

Linna, M., Häkkinen, U., & Magnussen, J. (2006). Comparing hospital cost efficiency between norway and finland. *Health Policy*, *77* (3), 268 - 278. Retrieved from `http://www.sciencedirect.com/science/article/ pii/S0168851005001892` doi: https://doi.org/10.1016/j.healthpol.2005.07 .019

Linna, M., Häkkinen, U., Peltola, M., Magnussen, J., Anthun, K., Kittelsen, S., . . . Rehnberg, C. (2010). Measuring cost efficiency in the nordic hospitals–a cross-sectional comparison of public hospitals in 2002. *Health Care Manag Sci*, *13*, 346-57.

Mannion, R., & Davies, H. (2018). Understanding organisational culture for healthcare quality improvement. *BMJ*, *363*. Retrieved from `https://www.bmj.com/content/363/bmj.k4907` doi: 10.1136/bmj.k4907

McKay, N. L., & Deily, M. E. (2008). Cost inefficiency and hospital health outcomes. *Health Economics*, *17*(7), 833-848. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.1299` doi: 10.1002/hec.1299

Medin, E., Anthun, K. S., Häkkinen, U., Kittelsen, S. A. C., Linna, M., Magnussen, J., ... Rehnberg, C. (2011). Cost efficiency of university hospitals in the nordic countries: a cross-country analysis. *The European Journal of Health Economics*, *12*, 509-519. Retrieved from `https://doi.org/10.1007/s10198-010-0263-1` doi: 10.1007/s10198-010-0263-1

Morey, R., Fine, D., Loree, S., Retzlaff-Roberts, D., & Tsubakitani, S. (1992). The trade-off between hospital cost and quality of care: an exploratory empirical analysis. *Med care*, *30*, 677–698.

Mutter, R., Greene, W., Spector, W., Rosko, M., & Mukamel, D. (2012). Investigating the impact of endogeneity on inefficiency estimates in the application of stochastic frontier analysis to nursing homes. *Journal of Productivity Analysis*, *39*. doi: 10.1007/s11123-012-0277-z

Mutter, R., Valdmanis, V., & Rosko, M. (2010). High versus lower quality hospitals: A comparison of environmental characteristics and technical efficiency. *Health Services and Outcomes Research Methodology*, *10*, 134-153. doi: 10.1007/s10742-010-0066-y

Nayar, P., & Ozcan, Y. (2008). Data envelopment analysis comparison of hospital efficiency and quality. *J Med Syst*, *32*, 193-199.

OECD. (2019). *Health at a glance 2019*. Retrieved from `https://www.oecd-ilibrary.org/content/publication/4dd50c09-en` doi: https://doi.org/https://doi.org/10.1787/4dd50c09-en

Schmidt, P., & Sickles, R. C. (1984). Production frontiers and panel data. *Journal of Business & Economic Statistics*, *2*(4), 367–374. Retrieved from `http://www.tandfonline.com/doi/abs/10.1080/07350015.1984.10509410`

Simar, L., & Wilson, P. W. (2000). A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics*, *27*. doi: 10.1080/02664760050081951

Statens legemiddelverket. (2016). The Norwegian health care system and pharmaceutical system. Retrieved from `https://legemiddelverket.no/english/about-us/the-norwegian-health-care-system-and-pharmaceutical-system`

Stimpfel, A., & Aiken, L. (2013). Hospital staff nurses' shift length associated with safety and quality of care. *Journal of nursing care quality*, *28*, 122-9. doi: 10.1097/NCQ.0b013e3182725f09

Valdmanis, V. G., Rosko, M. D., & Mutter, R. L. (2008). Hospital quality, efficiency, and input slack differentials. *Health Services Research*, *43*(5p2), 1830-1848. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-6773.2008.00893.x` doi: 10.1111/j.1475-6773.2008.00893.x

Wooldridge, J. M. (2013). *Introductory econometrics : a modern approach* (5th ed., international ed. ed.). S.l.: South-Western, Cengage Learning.

World Health Organization. (2004). *ICD-10 : international statistical classification of diseases and related health problems* (tenth revision, 2nd ed.). Retrieved from `https://apps.who.int/iris/handle/10665/42980`

World Health Organization. (2018). World health organization global health expenditure database: Health expenditure profile Norway. Retrieved from `http://apps.who.int/nha/database/country_profile/Index/en`

Zweifel, P., Kifmann, M., & Breyer, F. (2009). *Health economics* (Second edition ed.). doi: 10.1007/978-3-540-68540-1

Appendices

A Hospitals in data

Table 12: Hospitals in data

Variable code	Name hospital
OUS	Oslo University Hospital
VESTREVIKEN	Vestre Viken
AHUS	Akerhus University Hospital
INNLANDET	Innlandet Hospital Trust
ØSTFOLD	Østfold Hospital Trust
VESTFOLD	Vestfold Hospital Trust
TELEMARK	Telemark Hospital Trust
SØRLANDET	Hospital of Southern Norway
STAVANGER	Stavanger Hospital Trust
FONNA	Fonna Hospital Trust
BERGEN	Bergen Hospital Trust
FØRDE	Førde Hospital Trust
NORDTRØNDELAG	Nord-Trøndelag Hospital Trust
STOLAVS	St. Olavs University Hospital
MØREROMSDAL	Møre og Romsdal Hospital Trust
FINNMARK	Finnmark Hospital Trust
UNN	University Hospital of North Norway
NORDLAND	Nordland Hospital Trust
HELGELAND	Helgeland Hospital Trust
DIAKON	Diakonhjemmet Hospital
LOVISENBERG	Lovisenberg Diaconal Hospital
HARALDSPLASS	Haraldsplass Deaconess Hospital

B Correlation performance indicators of mortality quality variables

Table 13: Correlation Performance indicators of mortality quality variable

	perf_u2_mort30last	perf_u2_mort90last	perf_u2_mort180last	perf_u2_mort365last
perf_u2_mort30last	1.0000			
perf_u2_mort90last	0.9512	1.0000		
perf_u2_mort180last	0.8474	0.9524	1.0000	
perf_u2_mort365last	0.6019	0.7526	0.9133	1.0000

C Pearson correlation coefficients

Table 14: Pearson correlation coefficients

	ln_kost	ln_v_emrg	ln_v_elective	ln_v_dagb	ln_v_polk	perf_u2_readm30_emgc	perf_u2_mort90last	perf_u2_psi12vt_pe	perf_u2_psi13sepsis	perf_u2_psi15ac_punc	perf_u2_psi18ob_trau	perf_u2_bed_sores
ln_kost	1.0000											
ln_v_emrg	0.9543	1.0000										
ln_v_elective	0.9583	0.8942	1.0000									
ln_v_dagb	0.9644	0.9441	0.9017	1.0000								
ln_v_polk	0.9732	0.9454	0.9264	0.9703	1.0000							
perf_u2_readm30_emgc	-0.0416	0.1207	-0.1019	0.0085	-0.0329	1.0000						
perf_u2_mort90last	-0.3693	-0.2559	-0.4977	-0.2473	-0.3715	0.1153	1.0000					
perf_u2_psi12vt_pe	-0.0073	0.0050	0.0226	-0.0903	-0.0512	-0.0719	0.0180	1.0000				
perf_u2_psi13sepsis	-0.1410	-0.2065	-0.1647	-0.1876	-0.2230	-0.2688	0.2173	0.0982	1.0000			
perf_u2_psi15ac_punc	0.4833	0.4734	0.3913	0.4804	0.4510	-0.0267	-0.0341	-0.2044	-0.1471	1.0000		
perf_u2_psi18ob_trau	0.2404	0.2456	0.0753	0.3156	0.2689	0.1426	0.2312	-0.3103	-0.0858	0.3934	1.0000	
perf_u2_bed_sores	-0.1648	-0.1788	-0.1550	-0.1411	-0.1615	-0.2479	0.1251	0.0412	0.2360	-0.0417	0.0066	1.0000

D   Scatter plots of cost and quantitative variables

(a) Cost and DRG-points emergency patients (b) Natural logarithm of Cost and DRG-points emergency patients
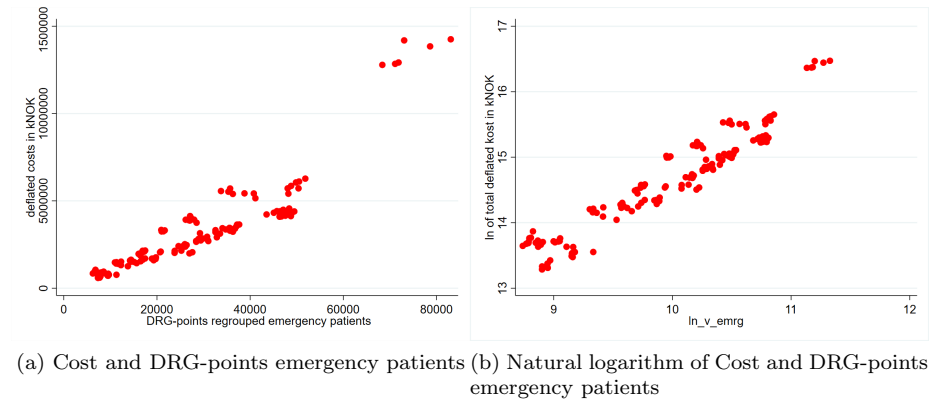
Figure 6: Scatter plots variable transformation Cost and DRG-points for emergency patients

(a) Cost and DRG-points elective patients   (b) Natural logarithm of Cost and DRG-points elective patients
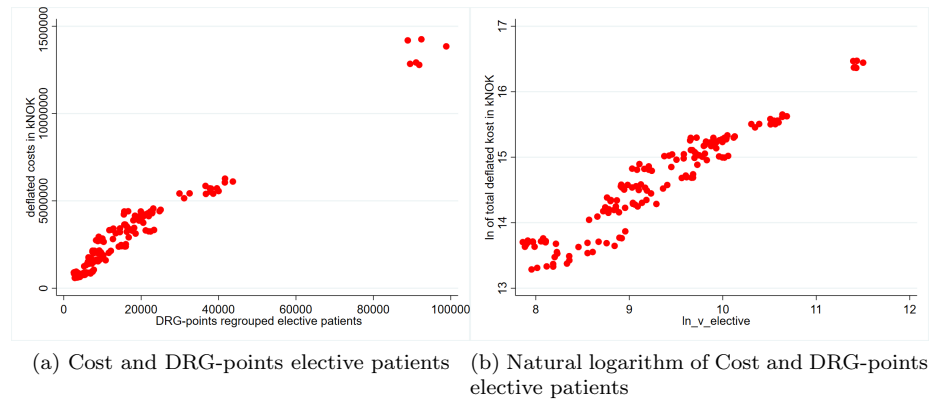
Figure 7: Scatter plots variable transformation Cost and DRG-points for elective patients

(a) Cost and DRG-points day patients

(b) Natural logarithm of Cost and DRG-points day patients
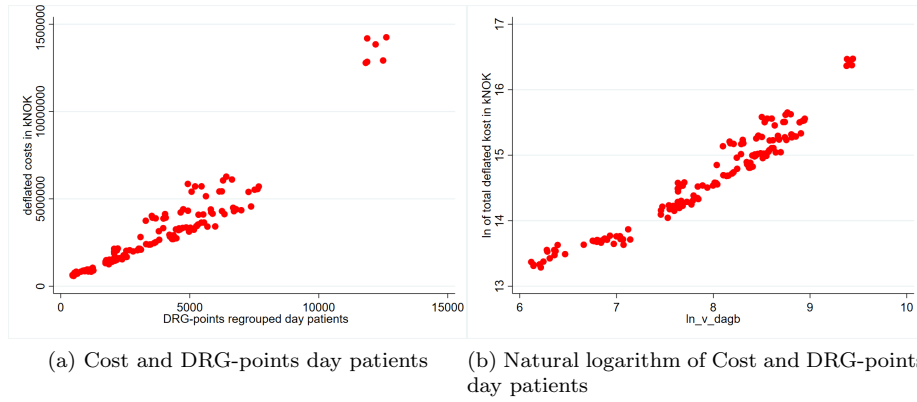
Figure 8: Scatter plots variable transformation Cost and DRG-points for day patients

(a) Cost and DRG-points polyclinic patients

(b) Natural logarithm of Cost and DRG-points polyclinic patients
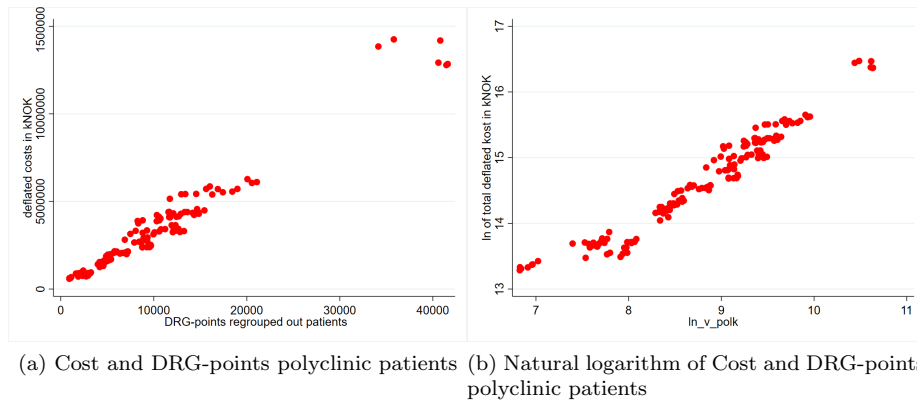
Figure 9: Scatter plots variable transformation Cost and DRG-points for polyclinic patients

E   Between and within variation

Table 15: Between and within variation of the included dependent and independent variables.

Variable		Mean	SD	Min	Max	Observations
ln_kost	overall	14·641	0·754	13·286	16.472	N=151
	between		0·769	13·347	16.415	n=22
	within		0·109	13·992	14.968	T-bar = 6.864
ln_v_emrg	overall	9·960	0·701	8·739	11.327	N=151
	between		0·712	8·861	11.214	n=22
	within		0·098	9·576	10.380	T-bar = 6.864
ln_v_elective	overall	9·332	0·854	7·860	11.501	N=151
	between		0·862	7·965	11.430	n=22
	within		0·176	8·359	9.846	T-bar = 6.864
ln_v_dagp	overall	7·917	0·835	6·118	9.443	N=151
	between		0·842	6·196	9.405	n=22
	within		0·147	7·255	8.463	T-bar = 6.864
ln_v_polk	overall	8·798	0·836	6·826	10.636	N=151
	between		0·846	6·907	10.570	n=22
	within		0·150	8·048	9.200	T-bar = 6.864
perf_u2_readm30_emgc	overall	1·015	0·102	0·698	1.320	N=151
	between		0·078	0·871	1.196	n=22
	within		0·068	0·776	1.235	T-bar = 6.864
perf_u2_mort30last	overall	0·955	0·125	0·612	1.328	N=151
	between		0·117	0·650	1.217	n=22
	within		0·052	0·820	1.097	T-bar = 6.864
perf_u2_mort90last	overall	0·966	0·128	0·601	1.301	N=151
	between		0·122	0·646	1.222	n=22
	within		0·050	0·849	1.128	T-bar = 6.864
perf_u2_mort180last	overall	0·969	0·137	0·586	1.402	N=151
	between		0·122	0·650	1.219	n=22
	within		0·070	0·844	1.227	T-bar = 6.864
perf_u2_mort365last	overall	0·970	0·178	0·542	1.614	N=151
	between		0·120	0·660	1.217	n=22
	within		0·135	0·793	1.411	T-bar = 6.864
perf_u2_psi12vt_pe	overall	0·940	0·209	0·556	1.757	N=151
	between		0·170	0·620	1.424	n=22
	within		0·127	0·342	1.290	T-bar = 6.864
perf_u2_psi13sepsis	overall	1·018	0·270	0·342	1.911	N=151
	between		0·204	0·639	1.536	n=22
	within		0·180	0·579	1.555	T-bar = 6.864
perf_u2_psi15ac_punc	overall	0·805	0·325	0·000	1.883	N=151
	between		0·276	0·068	1.494	n=22
	within		0·178	0·157	1.382	T-bar = 6.864
perf_u2_psi18ob_trau	overall	0·921	0·545	0·000	2.435	N=151
	between		0·477	0·000	1.767	n=22
	within		0·274	0·188	1.764	T-bar = 6.864
perf_u2_bed_sores	overall	0·920	0·465	0·275	2.696	N=151
	between		0·408	0·355	2.094	n=22
	within		0·236	−0·080	1.740	T-bar = 6.864