The relationship between costs and quality in nonprofit nursing homes

L. Di Giorgio [*] M. Filippini [†] G. Masiero [‡]

August 2013
Preliminary draft: please do not quote.

Abstract

We investigate the relationship between costs and quality in nonprofit nursing homes, which represents a key issue in the present context of adoption of cost containment measures. We estimate a total cost function for nursing home services using data from 45 nursing homes in Southern Switzerland between 2006 and 2010. Quality is measured by means of clinical indicators regarding process and outcomes that are derived from the Minimum Data Set. Conversely from many previous studies, we use panel data and estimate fixed effects models and control for unobserved heterogeneity. This allows to capture nursing homes specific features that may explain differences in quality or costs levels. We find evidence that poor levels of quality regarding outcome, as measured by the prevalence of severe pain and weight loss, lead to higher costs. Quality endogeneity is addressed through IV and GMM approaches using measures of residents empowerment through families as instruments.

Keywords: nursing home costs, nonprofit, quality indicators, cost-quality tradeoff

JEL classification: I10, L3.

---

[*]Institute for Health Metrics and Evaluation (IHME), University of Washington, Seattle; Department of Economics, Università della Svizzera italiana (USI), Switzerland. Corresponding author. E-mail address: laura.di.giorgio@usi.ch. We thanks Andrew Street for helpful comments and invaluable advice during our stay with the Policy team at the Center for Health Economics at the University of York. Also, we thank the Swiss National Science Foundation for financial support to the project. Any mistake is clearly our fault.

[†]Department of Economics, Università della Svizzera Italiana (USI); Swiss Federal Institute of Technology (ETH), Zurich, Switzerland.

[‡]Department of Engineering, University of Bergamo, Italy; Department of Economics, Università della Svizzera Italiana (USI), Switzerland.

1 Introduction

Ensuring good quality of care to nursing home (NH) residents is a major concern in many health care systems. Boosting quality levels must take into account cost containment measures, which are required to manage increasing health expenditures and ageing population. This twin objective of the nursing home sector - high quality and affordable costs- calls for better understanding of the potential trade-off between costs and quality. Quality aspects need to be integrated in empirical evidence of NH costs. The literature on NH costs is extensive, but marginally addresses quality of care. Most of these studies do not include measures of quality. Some of them use imprecise or indirect measures, such as the number of deficiency citations, information about staffing or mortality rates. Others rely on modeling quality as a latent variable (Gertler and Waldman, 1992; Carey, 1997). Moreover, the majority of these studies use cross-sectional designs and do not account for unobserved heterogeneity that may affect both costs and quality. To our knowledge, only Wodchis et al. (2007) use panel data models.

Failure to account for quality in cost functions is responsible for omitted variable bias (Braeutigam and Pauly, 1986). This bias is even more pronounced when comparing individual efficiency levels, as these techniques are particularly sensitive to model misspecification (Newhouse, 1994; Cremieux and Ouellette, 2001).

Donabedian (1988) conceptualizes the measurement of quality in the NH sector in terms of three dimensions: Structure (S), Process (P) and Outcome (O). The SPO-framework is widely accepted in empirical analyses of quality. Failure to include information about these three dimensions of care are due to measurement deficiencies and limitations in data availability. Recently, the introduction of the Resident Assessment Instrument (RAI) in the U.S. and some European countries, started a comprehensive and multidimensional assessment of all NHs residents health status. These data, also called Minimum Data

Set (MDS), are used to develop a battery of clinical indicators of quality that meet the taxonomy of the SPO model (Zimmerman, 1995). These indicators are categorized in two groups: indicators of quality regarding the process and indicators of quality regarding quality regarding the outcome. As such, they offer a unique tool to measure and compare quality of NHs in different domains of care (Berg et al., 2012).

A positive relationship between costs and quality is generally expected when higher levels of quality can be provided only through more costly equipment or additional staff employment. However, adverse inpatient events may be costly to treat because they involve additional resource utilization for extra care. The relationship between costs and quality may therefore depend on the dimension considered. Better procedures are expected to increase costs, while prevention of development of adverse outcomes may actually reduce costs (Weech-Maldonado et al., 2006; Wodchis et al., 2007).

This paper investigates the relationship between quality and costs in nursing home care, taking into account refined quality measures. We improve the specification of the cost function for the production of NHs care services used in previous analysis (Di Giorgio, Filippini et al., 2012; 2013) by incorporating quality measures based on the taxonomy of the SPO-model. We also aim to disentangle the impact of different dimensions of quality on costs.

The remainder of the paper is organized as follows: section 2 describes how quality for NHs services can be measured and presents the SPO-framework more in detail. Section 3 reviews previous studies on the relationship between costs and quality. In section 4 and 5 we report our data and describe the rationale behind quality indicators included in the following empirical analysis. We also detail the empirical strategy. Section 6 and section 7 respectively provide discussion of the findings and concluding remarks.

2 Quality

2.1 Definition and measurement

No universal definition of quality exists in health research. The Institute of Medicine (IOM, 2001) states that "quality of care is the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge".[1] This definition has significantly influenced the literature on quality and is very much related to the paradigm of quality proposed by Donabedian (1988). His seminal article on the assessment of quality of care represents the foundation of modern quality assessment, providing a framework of reference with guidance validity. Donabedian proposed the so called SPO-framework. Structure is defined by the attributes of the setting in which care is provided, such as material resources (e.g. equipment), human resources (e.g. staffing levels) and organizational structure (e.g. payment system). Process refers to the activities of practitioners to give care, such as making a correct diagnosis and implementing the treatment accordingly. Outcome defines the change in health status of the patient. The success of this paradigm lies in its broad scope, which encompasses older concepts of quality. Table 1 shows how different measures of quality used in the literature fall within the dimensions of the SPO-framework. With the development of the quality indicators derived from the RAI, direct clinical measures of quality regarding process and outcome are available.

Only a few countries have adopted the RAI. Many use different systems to measure quality in the NH sector (Nakrem et al., 2009). Each measure of quality has advantages and disadvantages, which are discussed below. Some relatively old indicators (non-clinical) are still considered valid and are often combined in empirical studies with clinical quality indicators derived from RAI.

Previous studies attempt to capture NHs quality differences mainly using in-

[1] Other well recognized definitions are provided by the UK Department of Health (1997), the Council of Europe (1998), and the WHO (2000). For a detailed exposition of the most influential and known definitions of quality, see Legido-Quigley et al. (2008).

Structure	Process	Outcome	
		Objective	Subjective
Room size	Staffing information	Mortality rates	Resident satisfaction
Equipment	Mistakes rate	Hospitalization	Family satisfaction
Staffing levels	Deficiency citations	Quality indicators (RAI)	Deficiency citations
Nurse skill mix	Quality indicators (RAI)		

Table 1: Classification of quality indicators based on the SPO-framework (Donabedian, 1988).

dicators of structure or indirect signals. Probably the most recognized indicator with current validity is the number of deficiency citations (Castle and Ferguson, 2010). Deficiency citations have the advantage of representing different dimensions of reduced quality but suffer from detection bias due to high variability in the use of citations among states/countries. Another important indicator that is employed extensively in the literature is the use of resources, in particular, information on staffing. While earlier studies focus on staffing levels as determinant factors (McKay, 1989; Farsi et al., 2005; Farsi et al., 2008), recent studies recognize the need to extend this dimension to staff characteristics, such as staff turnover, worker stability and skill levels (Castle and Engberg, 2005; Castle and Engberg, 2007; Dormont and Martin, 2012; Spilsbury et al., 2011), as well as willingness of leadership (Rantz, 2004). A recent systematic review of Bostick et al. (2006) shows not only evidence of association between higher licensed staff and quality, but also a significant relationship between staff turnover and quality indicators such as pressure ulcers, weight loss and functional ability.

Similarly, the advantages and disadvantages of the quality indicators based on the SPO model are discussed in Castle and Ferguson (2010). Structural indicators have the advantage of being easy to measure and data are often available. The disadvantage is that the presence of structural attributes does not imply its best use. Castle and Ferguson (2010) maintain that structural quality indicators are necessary but not sufficient. Indicators of process are usually easy to interpret as they inform on the provision of a particular treatment. Even in this case, it cannot be determined whether or not the provided treatment is appropriate. Finally, outcome indicators are of natural interest, as they measure the change in patients' health status. The main problem with these indicators is that it is extremely difficult to isolate the effect of care and changes in health, as the latter may be influenced by many uncontrolled factors.

Recently, interpersonal aspects of care to NHs residents received increasing attention. Residents' satisfaction seems to be a valid indicator with great potential even though it is not without limitations. People's reluctance to reveal their opinions and the inability of severe residents to understand and answer questions are among the most important.

2.2 Concerns about quality indicators

The recent development of clinical quality indicators has improved measurement of quality, but with some limitations. Firstly, due to the absence of a universally accepted definition of quality, the selection of quality indicators to include in empirical analyses is, to some extent, arbitrary (Castle and Ferguson, 2010). This is an issue because of the usually low correlation among quality indicators. Indeed, facilities with excellent outcomes in some dimensions may perform poorly in others. The choice of indicators may therefore affect the perception of NH quality. Secondly, detection bias occurs if higher quality NHs are the more vigilant in looking for and detecting quality issues (Mor et al., 2003). Since staff of the NH, rather than an independent authority, assesses residents health status, risk of detection bias exists. Thirdly, variation in clinical quality indicators may be due not only to changes in quality, but also in risk or to error (Arling et al., 1997). To cope with this issue, different risk-adjustment techniques are used. While first studies of NH quality mainly used adjustment methods at the facility level (Nyman, 1988; Zinn et al., 1994; Zinn et al., 1993a), more recently risk-adjustment has been performed at the individual level. Different approaches include stratification, covariates model (Mukamel, 1997) and standardization (Zinn et al.,1993b). For some clinical indicators of quality that are considered particularly relevant in detecting the presence of problematic cases of quality shortcomings, no risk-adjustment is required. Among these are presence of daily physical restraints (Berg et al., 2002), dehydration and fecal impaction (Arling et al., 1997; Karon et al., 1999). The main issue of risk-adjustment techniques is that they may only partially capture the risk-factors of residents, resulting in biased estimates of quality coefficients may occur (Mukamel et al., 2008). To address this issue, instrumental variables techniques have been discussed (Angrist et al., 1993). Risk-adjustment is also of concern when risk-adjustment factors are themselves a function of quality. In these cases, quality scores could be over-adjusted, giving credit for poor quality (Mukamel et al., 2008). Finally, quality indicators are often criticized because they reflect a bio-medical perspective and neglect consumers' value of quality.[2]

3 Empirical evidence on the impact of quality on costs

Empirical models using non-clinical quality measures mainly focused on the impact of specific factors on costs, such as market structure, forms of organization, or reforms implemented in the NH sector. Quality measures are usually introduced as control factors. From these studies, some use staffing information (e.g. Crivelli et al., 2002; Farsi et al., 2005, 2008; Dormont and Martin, 2012; Konetzka et al., 2004) or deficiency rates (e.g. Harrington et al., 2001). Another strand of literature exploits determinants of quality variability. Factors considered include the impact of state regulations (Bowblis and Lucas, 2012), ownership form (Grabowski et al., 2013) and competition (e.g. Brekke et al., 2010; Castle et al., 2008; Forder and Allan, 2011; Grabowski, 2004; Starkey et al., 2005).

We focus our review on studies that try to disentangle the relationship between costs and quality using clinical indicators derived from RAI. Their main

[2] One possibility to include residents' voice it to use family and residents satisfaction scores (Sangl et al., 2007), however these data are not available.

Study	Quality indicators	Empirical strategy (Data)	Effects on costs
Mukaml and Spector, 2000	Incidence of functional decline	Weighted least squares models (Cross-sectional data)	Inverted U-shaped
	Incidence of bedsores		Inverted U-shaped
	Mortality		Inverted U-shaped
Laine et al., 2005a	Prevalence of pressure ulcers	Stochastic frontier models (Cross-sectional data)	Positive effects
	Prevalence of weekly use of depressants		No significant effects
	Prevalence of depression w/o treatment		No significant effects
Laine et al., 2005b	Prevalence of depression w/o treatment	Stochastic frontier models (cross-section)	No significant effects
	Prevalence of pressure ulcer		Positive effects
	Prevalence of use of depressants		Positive effects
	Prevalence of use of physical restraints		No significant effects
Weech-Maldonadoet al., 2006	Worsening of pressure ulcers	Two-stage least squares models (Cross-sectional data)	Inverted U-shaped
	Mood decline		U-shaped
Wodchis et al., 2007	Prevalence of physical restraints	Random and fixed effects models (Panel data)	Negative effects
	Prevalence of use of depressant		No significant effects
	Prevalence of urinary incontinence		Positive effects
	Incidence of urinary incontinence		U-shaped
	Prevalence of skin ulcers		No significant effects
	Incidence of skin ulcers		Negative effects
	Prevalence of pain		No significant effects

Table 2: Overview of selected studies investigating the relationship between costs and quality in NHs.

contribution is summarized in Table 2, with details on the choice of quality indicators, the empirical approach and the results obtained.

Mukamel and Spector (2000) is one of the first studies to investigate the relationship between costs and quality using the RAI-derived quality indicators. The authors estimate a variable cost function for NHs in New York State. Three indicators of outcome quality are included: activity of daily living (ADL), pressure ulcers and mortality. Regression-based risk adjustment is applied (Mukamel, 1997). Weighted ordinary least-squares is used to tackle the issue of different sample size in the calculation of the outcomes variables.[3] The authors report an inverted U-shaped relationship between costs and quality, although only few coefficients are statistically significant. The loss of statistical significance is attributed to high multicollinearity among higher-order terms of quality indicators. Due to the availability of only weak instruments, the endogeneity issue of quality is ignored.

An important contribution to the cost-quality relationship is provided by Laine et al. (2005a, 2005b) which implement Stochastic Frontier Models (SFM). In both studies, endogeneity of quality is not addressed. The first study (2005a) models a stochastic production frontier for the Finnish long-term care sector in 2001 where the dependent variable is specified as the case-mix weighted patient days and covariates include only input characteristics. Ward characteristics and quality are modeled following Battese and Coelli (1995), i.e. technical inefficiencies are specified as a function of quality indicators. Quality is measured linearly by three continuous indicators: the prevalence of high-risk pressure ulcers, the prevalence of weekly use of depressants and hypnotics, and the prevalence of depression with no treatment. The latter two indicators are not risk adjusted. The prevalence of pressure ulcers is the only quality indicator significantly associated with technical inefficiency. The suggested relationship is that higher prevalence of pressure leads to higher technical efficiency.

Laine et al. (2005b) provide a similar cross-sectional analysis which shifts

[3] The authors used the inverse of the squared root of the sample size as weights.

the focus from productive efficiency to cost efficiency. The analysis is performed using data at the ward level obtained aggregating individual-level data. The authors include quality indicators regarding process, the prevalence of depression without treatment and prevalence of pressure ulcers adjusted for risk, in the deterministic part of the cost frontier. Indicators of output quality, i.e. the prevalence of use of depressants and hypnotics and the prevalence of use restraints, are modeled following Battese and Coelli (1995). The mean values of the indicators over a three-years period is taken without risk adjustment. The underlying idea is to allow indicators of process quality to affect the production process itself, while outcome is restricted to have an impact on the level of inefficiency. The results show that a worse outcome in terms of higher prevalence of pressure is associated with higher costs, while poor process quality measured by the weekly use of depressants and hypnotics is associated with higher inefficiency. However, the impact of these quality indicators is relative low.

Weech-Maldonado et al. (2006) investigate the impact of quality on costs in U.S. NHs. Using cross sectional data from around 750 facilities, they test the inverted U-shaped theory by adding squared and cubic terms of quality. Quality is measured by changes in physical and psychological outcome indicators, i.e. worsening pressure ulcers and mood decline. Indicators are adjusted for risk using the covariates model (Mor et al., 1998). A weighted 2-stage least squares regression is estimated to address endogeneity of quality indicators. Socio-demographic characteristics at the county-level as well as the presence of alternative service providers are used as instruments for quality scores. However, the validity of these instruments is not tested. The results show an inverted U-shaped relationship between costs and pressure ulcers. The opposite pattern arises for mood decline, showing that different indicators of quality may lead to different types of relationships.

Additional evidence based on data from Ontario, Canada, is provided by Wodchis et al. (2007). The authors estimate individual-effects models where total costs are regressed on output, labor price, some exogenous variables and quality indicators adjusted for risk using resident-level covariates model, with the only exception of prevalent physical restraint use. Heteroskedasticity, autocorrelation and endogeneity issues are discussed. However, due to the lack of a valid instrument, endogeneity is ignored. The analysis shows a negative relationship between costs and use of daily physical restraints, as well as worsening incontinence. Antipsychotic use, the prevalence of ulcers and the prevalence of severe pain are not statistically significant.

Most of the studies presented above find correlation between some quality indicators and costs. However, the association is weak and the approaches used are hardly comparable. As suggested in the Introduction, the majority of these studies use cross-sectional designs and do not account for unobserved heterogeneity that may affect both costs and quality. To our knowledge, only Wodchis et al. (2007) use panel data and an estimation approach that considers unobserved heterogeneity. Unobserved heterogeneity may represent a serious problem in analyses of costs-quality relationship due to the difficulty in measuring quality. If the risk-adjustment technique used in cross-sectional studies does not perfectly capture the facility-specific features, then the results may be biased. Also, only few studies address the potential endogeneity of quality, and virtually no test is provided on the validity of the instruments.

In the following section we propose an empirical approach to investigate the relationship between costs and quality using data from Swiss NHs. As compared to previous studies, we are able to control for unobserved heterogeneity by exploiting a panel data set. Also, we try to address the potential issue of quality endogeneity. Our specification of the cost model is improved by including four clinical measures of quality regarding NH process and outcome.

4 Model specification and data

4.1 Choice of quality indicators

To select the most appropriate clinical quality indicators for our cost analysis, we consider three strands of literature. First, we consult the medical recommendations on the pertinence of the indicators to reveal quality issues in NHs. Second, we consider studies on the technical requirements that quality indicators need to satisfy to be included in empirical analyses. And finally, we look at previous studies investigating the relationship between costs and quality using quality indicators analysed by Zimmerman (1995) (see section 3).

From the medical literature we consult the numerous lists of recommended indicators to use in benchmarking analyses of NHs (Berg et al., 2002; Morris et al., 2003; Rantz et al., 2004).

From the medical-statistical literature, we derive three main criteria that should be satisfied for the empirical analysis (Berg et al., 2002; Laine et al., 2005b): a relatively large variation in the quality scores, the absence of multicollinearity between the indicators and other variables, and a relatively large number of observations from which the quality indicators are calculated. The issue of the denominator is motivated by statistical properties since some quality indicators capture the onset of rare events. In these cases, the relevant question is whether the observed frequency of the event can be considered as a "true score", or it is driven by random shocks. Indeed, standard errors of rare events are large leading to problems in the comparison of quality among facilities. The minimum number of observations for benchmarking is 20 (Berg et al., 2002).

Finally, we consider previous economic studies analysing the relationship between costs and quality with particular focus on the selection of quality indicators (see section 3).

Based on these criteria, we select 4 quality indicators from the 22 available in our dataset. The two indicators of process are the presence of antipsychotic use for low-risk residents (Q_{Antips}) and the presence of daily physical restraints use (Q_{Restr}). The two indicators of outcome include the prevalence of weight loss (Q_{Weight}) and the prevalence of severe pain (Q_{Pain}). Finally, we control for time-invariant quality aspects regarding the structure of NHs through the econometric specification of the model (see section 5).

An alternative approach to the use of single quality indicators is to combine the 22 quality indicators into a composite measure, as suggested for instance by the U.S. Institute of Medicine (2006). Though, combining different measures of quality implies choosing (arbitrary) weights for remarkably heterogeneous quality events. Moreover, different and small numbers of eligible residents for some quality indicators across facilities may represent a serious problem with the composite measure of quality. Also, factors affecting costs cannot be clearly identified. Finally, Bayesian hierarchical models to adjust quality rates for uncertainty associated with the number of events are not completely satisfactory. We emphasize Donabedian's approach and are interested in separate indicators for process, outcome and structure. Therefore, we prefer to focus on single quality measures which are more reliable and meaningful.

4.2 Detailing the cost function

The model specification used in this paper draws from the model specification used in a companion paper focusing on cost efficiency of different institutional forms (Di Giorgio, Filippini, Masiero, 2012). As previously discussed, in order to identify the impact of quality on costs, the cost model includes four quality indicators derived from the RAI. Total costs are a function of output (Y), the prices for labor, capital and material (P_l, P_k, P_m), the case-mix of residents (Q_1), the institutional form of the NH (IF), the nursing staff ratio (Q_2), four indicators of quality (Q_{Pain}, Q_{Weight}, Q_{Antips}, Q_{Restr}), and a time trend (τ)

which captures technological progress:[4]

$$C = f(Y,\, P_l,\, P_k,\, P_m,\, IF,\, Q_2,\, Q_{Pain},\, Q_{Weight},\, Q_{Antips},\, Q_{Restr},\, Q_1,\, \tau). \quad (1)$$

The price of labor is calculated as the weighted average wage of different professional categories employed in the NH (doctors, nurses, administrative and technical staff). We choose to include only one price of labor to avoid multicollinearity problems that typically arise with labor prices for different categories. The price of capital is calculated as the sum of mortgage costs, amortization and costs related to capital purchases divided by the capital stock, which is approximated by the number of beds. The price for material and meals is computed by taking the remaining costs and dividing them by the number of meals provided each year. This item mainly includes costs for food, energy and administrative costs.

The main difference among nonprofit nursing homes lays in their *institutional form*, which underlines property rights or legal constraints affecting different institutions. Hence *public-law* nursing homes are public administrative units without a separate judicial status from the local public administration and are diectly integrated into it. Conversely, *private-law* NHs usually take the form of a foundation. Generally, foundations are created by natural persons, private legal entities or local governments. Since the decision-making process may vary across institutional forms, we include a dummy for the institutional form (IF) equal to one when the NH is a *public-law* organization, and 0 otherwise.

Q_2 is the nursing staff ratio, that is the ratio between the number of nurses employed in NHi and the number of nurses that should be employed according to the guidelines of the regulator (optimal amount of staff).[5] Because nursing care is a labor-intensive service, staffing levels have been recognized as a good

[4] In a non-competitive environment such as the Swiss one, there is no reason to assume that NHs minimize costs. In this case, the estimated costs function is a "behavioral cost function" (Evans, 1971) and can still be used to make a comparison among firms.

[5] As compared to other quality indicators related to staff levels, our indicator is conceptually different. The nursing staff ratio is the deviation from the optimal number of nurses that should be employed according to guidelines rather than the number of staff nurses employed.

indicator for quality (Bostick et al., 2006).

In addition to the nursing staff ratio, we include four additional indicators of quality derived from the MDS that measure the prevalence of adverse events, i.e. the prevalence of antipsychotic use for low-risk residents (Q_{Antips}), daily physical restraints use (Q_{Restr}), weight loss (Q_{Weight}) and severe pain (Q_{Pain}).

Q_{Antips} is risk-adjusted based on the stratification approach, Q_{Restr} is a *sentinel indicator* and as such no risk-adjustment is required (Berg et al., 2002). Due to lack of data at the resident level, we further control for case-mix differences using an index at the facility-level (Q_1).This index measures average patients' assistance need by means of normal daily activities such as eating, personal care or physiological activities and is calculated centrally on a yearly basis by the regulator. Patients are classified in one out of five categories according to their severity level. A value between 0 and 4 is assigned where higher values indicate more severe cases.[6] We expect this case-mix indicator to be correlated with patients' risk factors that are not observable. Moreover, any unobserved facility-specific risk factors feature is captured by the individual effects. We acknowledge that the classification system used may be less precise as compared to adjustments based on clinical information at individual level. However, as previously discussed, even complex systems of risk adjustments present important shortcomings.

For the estimation of the cost model in (1), we use a log-log functional form. This implies that the cost elasticities are not allowed to vary with output. When choosing the functional form, parsimony in the number of coefficients to be estimated is traded off against flexibility. A translog functional form would require interacting all quality indicators with the production factors. The number of parameters to be estimated would expland to $(n + 1)(n + 2)/2$, leading to an important loss of degrees of freedom given our sample size.[7]

[6]Note that this is not the RUG's classification system of residents. As compared to the RUGs system, our case-mix measure is not derived from the MDS. The main advantage is that case-mix differences are less likely to reflect quality levels.

[7]In a preliminary analysis, we also tried to estimate: 1) a full-translog cost model and 2)

Input prices and total costs are divided by the material price in order to satisfy the homogeneity condition in input prices.[8] The log-log form of eq. (1) is:

$$\begin{aligned}\ln\left(\frac{C}{P_m}\right) &= \delta_0 + \delta_Y \ln Y + \delta_{Q_1} \ln Q_1 + \delta_{P_l} \ln \frac{P_l}{P_m} + \delta_{P_k} \ln \frac{P_k}{P_m} + \delta_{IF} IF \\ &\quad + \delta_{Q_2} Q_2 + \delta_{Q_{Antips}} Q_{Antips} + \delta_{Q_{Restr}} Q_{Restr} + \delta_{Q_{Weight}} Q_{Weight} \\ &\quad + \delta_{Q_{Pain}} Q_{Pain} + \delta_t \tau + \varepsilon, \qquad (2)\end{aligned}$$

where ε is the error term which may contain individual effects δ_i. The individual subscript i and the time subscript t are omitted for simplicity.

The estimation of the cost function in (2) is based on the assumption that output, input prices and quality are exogenous variables. In the case of the NHs included in the sample, output is likely to be exogenous because NHs have to accept all residents in a given catchment area and residents do not have free choice of facility. Also, the excess demand framework due to subsidized prices leads to occupation rates of about 100%. For the same reasons, case-mix is also likely to be exogenous. Moreover, reimbursement systems are linked to NH-specific case-mix, further reducing incentives to attract less costly customers. Input prices can be considered exogenous because NHs have to follow the guidelines imposed by the regulator.

As with respect to quality, it is important to distinguish between the indicator nursing staff ratio and the four clinical quality indicators derived from RAI. The nursing staff ratio is strongly regulated by the canton and a NH is not allowed to deviate too much from the optimal staff size imposed by the canton. Therefore, we can exclude the presence of endogeneity.[9] The four clinical indicators of quality are instead not regulated and may therefore be endogenous.

an hybrid translog cost model. In the hybrid translog cost function the quality indicators were included only in linear form. The results of the full translog were not satisfactory, probably due to multicollinearity problems and the loss of degrees of freedoms. The results of the hybrid cost function were very similar to those obtained with the log-log functional form.

[8] The cost function is linear homogenous of degree 1 in input prices when a 10% increase in all input prices leads to a 10% increase in total cost.

[9] The Durbin-Wu-Hausman test performed using the lagged Q_2 as instrumental variable does not reject exogeneity at the 99% level.

As we will see later in the empirical analysis, we try to address the potential endogeneity issue by using instrumental variables.

4.3 Data and descriptive statistics

To conduct the empirical analysis, we merge two datasets on costs and quality of long stay (chronic) patients NHs from a region in Switzerland (Ticino). The first dataset includes yearly resources use at the organization level extracted from the annual reports of NHs. It includes 45 NHs over a 10-years period, from 2001 to 2010. The second dataset contains information derived from the MDS on 22 quality indicators at the organization level for the period 2006-2010, excluding the year 2008. These indicators measure the presence of adverse events in a facility.[10] Due to missing values in the data set, no quality scores were available for three NHs for the years 2006 and 2007. We also exclude observations in which the denominator of the quality score is smaller than 20. This leads to a loss of other 14 observations. Complete data pertaining to 45 NHs observed over a 4-years period, 2006, 2007, 2009 and 2010 were used. The total number of observations is 163.

In Table 3 we provide descriptive statistics for the main costs and quality variables. Median values are not shown because of the similarity with mean values. The data show that on average a resident day costs 255 Swiss francs (SFr.). The difference between minimum and maximum costs is of almost SFr. 100. This may be due also to differences in output, as the number of resident days ranges between 30000 to more than 64000 days. Average case-mix of residents is 3.15, with important differences among NHs (2.38-3.83). The average price of labor and material is approximately SFr. 84000 and SFr. 9.60 respectively, and NHs are very homogenous in these respects. The price of capital shows higher variation, from SFr. 1500 to almost SFr. 17000. These differences are due to renovation or enlargement investments. At the approximation point, the shares of capital-, material- and labor costs are 7%, 12% and 81%, respectively.

[10] Unfortunately, data at the resident-level were not available.

Variables	Mean	Std. Dev.	Min.	0.25	0.75	Max.
Average cost (SFr./Y)	255.73	21.48	213.01	242.03	268.83	359.64
Annual resident days (Y)	25434	10231	8955	19041	30128	64275
Average dependency index (Q_1)	3.15	0.30	2.38	2.95	3.38	3.83
Average labor price in SFr. per employee per year (P_l)	83680	4068	69415	81784	85776	97512
Average capital price in SFr. per bed (P_k)	6011	2320	1510	4552	7354	16914
Average material price in SFr. per meal (P_m)	9.60	1.49	6.85	8.73	10.12	16.11
Nursing staff ratio (Q_2)	0.93	0.07	0.74	0.88	0.97	1.12
Prevalence of antipsychotic use (Q_{Antips})	0.32	0.12	0.08	0.24	0.4	0.88
Prevalence of physical restraints use (Q_{Restr})	0.20	0.10	0.00	0.13	0.26	0.50
Prevalence of weight loss (Q_{Weight})	0.07	0.05	0.00	0.04	0.09	0.25
Prevalence of severe pain (Q_{Pain})	0.21	0.12	0.00	0.11	0.27	0.61

Notes: All monetary values are in 2005 Swiss francs (SFr.), adjusted by the national Consumer Price Index.

Table 3: Descriptive statistics of the main costs, inputs and quality variables.

Regarding the indicators of quality, the data show that, as expected, the nursing staff ratio is very close to 1, and little variation is present (0.74-1.12). On average, 32% of low-risk patients use antipsychotics, but in some NHs this value reaches 88% suggesting a serious problem in the NH sector. The average prevalence of daily physical restraints use is 20%, and ranges between 0 and 50%. The average prevalence of residents who lost weight unexpectedly is 7%, and this percentage ranges between 5 and 25%. Finally, on average, the prevalence of residents suffering from severe pain is 21%, but reaches more than 60% in some cases.

An interesting question is whether NHs that perform well in one quality dimension perform also well in the other quality domains. To answer this question, we compute the correlation among indicators (including the staff ratio) and Kendall's rank correlation coefficient (Kendall, 1955). The latter measures the similarity of the ordering of the NHs when ranked based on the scores of the quality indicators. Both measures indicate a very low correlation among quality indicators ($< 25\%$).

5 Methodology

The focus of this paper is to analyze the impact of quality of process and quality of outcome on costs. We use a "classical" regression approach for panel data rather than stochastic frontier models. From the econometric point of view, the classical estimators to use with panel data are Ordinary Least Square (OLS), fixed-effects (FE) and random-effects (RE). The Breusch-Pagan test (1980) suggests the use of individual effects models ($\chi^2(1)$=32.18, P-value=0.000) as compared to the pooled model. Individual effects are used to capture quality regarding time-invariant structural aspects of NHs. The FE model treats the individual effects as fixed parameters. These are allowed to be partially correlated with regressors, thus accommodating a limited form of endogeneity (Cameron and Trivedi, 2010). This feature is particularly appealing in studies of costs-quality relationship due to the potential endogeneity of the indicators. Instead, the RE model assumes that the unobservable individual effects are random variables distributed independently of the regressors, that is: $\delta_i \sim (\delta, \sigma_{\delta^2})$ and $v_{it} \sim (0, \sigma_{v^2})$, and the coefficients are estimated with the Generalized Least Square (GLS) method. The Hausman test rejects the null hypothesis of no systematic difference in coefficients between the RE and the FE at the 5% level ($\chi^2(10)$=19.70, P-value=0.032). Given that the percentage of within variation of the variables of interest with respect to the overall variation is satisfactory, the FE estimates should be fairly precise (Cameron and Trivedi, 2005). Therefore, for the present empirical analysis the FE model represents the preferred estimator. The OLS and the RE estimates are presented for comparative purposes.

Standard errors are corrected using the cluster robust estimator based on

Stock and Watson (2006) in all models. Stock and Watson (2006) show that the cluster-robust estimator is preferred in FE models if serial correlation is expected, and it is reasonable to rely on asymptotic theory. In our sample, each cluster contains a sufficient number of observations so that clustered standard errors would be preferred (Kezdi, 2004).[11]

Further, in order to take into account the potential endogeneity of quality, we also evaluate an instrumental variables approach. We consider both the Two-Stage Least Squares (2SLS) approach and the efficient Generalized Method of Moments (GMM) approach combined with the FE model. The GMM approach has the advantage of consistency in the case of arbitrary heteroskedasticity and shows higher flexibility than 2SLS, in particular to test the validity of the instruments. Both approaches come at the price of poor finite sample performance, in particular in the case of weak correlation between the instruments and the endogenous variable.[12] In this analysis, we prefer the GMM approach as it allows errors clustering for panel data and provides a battery of tests to check the validity of the instruments.[13]

A valid instrument must satisfy two requirements: the instrument z must be correlated with the endogenous variable x, $Cov(x, u) \neq 0$, and uncorrelated with the error term u, $Cov(z, u) = 0$. In the case of a single endogenous variable, the first condition is tested with a simple regression of z on x. A statistically significant coefficient provides evidence of the correlation between instrument and endogenous variable. In the case of multiple endogenous regressors, the *Shea partial* R^2 (1997) measure should be used, as it takes into account the

[11] Kezdi (2004) states that a sample of 50 clusters is close enough to infinity for accurate inference if the number of observations for cluster is not too small. A cluster is considered small if it contains less than five observations per cluster (Rogers, 1994). In the present case, the significance of the coefficients remain unchanged when standard errors are clustered as compared to not clustered.

[12] In particular, the efficient GMM approach may suffer from poor finite sample properties as the optimal weighting matrix of the efficient GMM estimator is a function of fourth moments, which require large sample size (Hayashi, 2000).

[13] A possible alternative to clustered standard errors for 2SLS estimates is bootstrapped standard errors. However, in the present case standard errors become so large, that the entire statistically significance gets lost.

intercorrelation among the instruments.[14] However, this does not exclude the possibility of weak instruments, which lead to a very high asymptotic bias.

The second condition can be tested when there are more instruments for an endogenous variable. In this case, the C-statistic, also called "difference-in-Sargan" statistic, can be used to test the orthogonality condition of a subset of instruments (Hayashi, 2000).

As shown in previous studies (Mukamel and Spector, 2000; Wodchis et al., 2007), it is not easy to find good instruments for quality. In this study, we rely on three hypotheses. First, the number of relatives visits exert pressure on the staff and management of the NH to keep an adequate level of quality. Second, the quality offered by the NH depends on the average quality offered by surrounding NHs. Third, the share of adults and elderly people living in the area of the NH exercise an indirect pressure on the quality offered by a NH.

We identify two variables related to the first hypothesis: the weighted average distance (traveling time) between the residents' location and the NH facility, and the weighted population density of the area served by each NH. The first measure captures the travelling time necessary for family members to reach the NH. When a NH serves residents from more municipalities, travelling times are weighted by the relative importance of the municipality in terms of population.[15] The same approach is used to calculate weighted population density. Population density is calculated as the ratio between the number of inhabitants and surface in hectare. The same weight is applied when more municipalities are served by the same NH. These variables are expected to capture residents empowerment through family members (voice). Higher population density and shorter travelling time are expected to increase the likelihood of being visited,

[14] The F diagnostic for weak instrument for the joint significance of the instruments in first-stage regression does not recognize situations in which some instruments are good while others are weak.

[15] Weights correspond to the relative percentage of people living in a municipality w.r.t. the whole catchment area of the NH. This approach works also in the case a NH serve only the population of one municipality, as in this case we measure the distance between the center of the village and the NH.

as empirically shown by Dillmann et al. (2002).

For the second hypothesis, we build a variable to capture pressure from the presence of other NHs located in geographical proximity. For each year and NH, pressure is measured as the average score of each quality indicator of all NHs located in the vicinity. Vicinity is defined by the eight districts in which the region considered in this analysis is further divided.[16] The underlying motivation is that managers of NHs located close by affect each other's. Travel time or transport costs are increasingly used in the literature to investigate the impact of competition on quality (Brekke et al., 2010; Forder and Allan, 2011). Although in the present case competition is not direct, managers may still compete for other reasons such as reputation.

Finally, we include the percentage of young, adult and elderly population in the catchment area of each NH. Population structure is expected to capture the extent to which the population is interested and involved in issues relating to quality of NH services.

We also consider lagged values of quality indicators as a natural instrument. Lagged values are an attractive instrument due to the high correlation with the endogenous variable. Nevertheless, caution is necessary in the presence of serial correlation in the data, as this may invalidate the instruments (Angrist and Kruger, 2001). To test for autocorrelation in panel data set, we use the test developed by Wooldridge (Drukker, 2003; Wooldridge, 2002).

In Table 4 we provide some descriptive statistics of the instruments discussed:[17]

The average distance in terms of travelling time is about 4 minutes, with longest travelling time being almost 11 minutes. The average population density in each area served by the NH is of 10 inhabitants per hectare, but shows high

[16] The region considered in the analysis is further divided in 8 districts: Mendrisio, Lugano, Vallemaggia, Locarno, Bellinzona, Riviera, Blenio and Leventina. Given only few NHs are located in northern districts, Vallemaggia, Leventina and Blenio are pooled together.

[17] The descriptive statistics of the instruments for the second hypothesis are not shown as they are less informative.

Instruments	Mean	Std.Dev.	Min	Max
Average distance from residents to NH	4.30	2.70	0.2	10.40
Population density in area served by each NH	10.62	14.30	0.13	80
Percentage young people	0.28	0.03	0.21	0.33
Percentage adults	0.26	0.02	0.21	0.34
Percentage elderly	0.20	0.04	0.12	0.30

Table 4: Descriptive statistics of instruments.

variability reaching a peak of 80. The percentage of young-, adults- and elderly individuals is on average 28, 26 and 20, respectively.

6 Results

The estimation results are presented in Table 5. Standard errors are provided in parentheses. The statistics for R^2 and the number of observations (N) are provided at the end of the table.

The coefficients are very similar among the different panel models, with the exception of the output coefficient which is lower in the FE model. The OLS model does not consider the unobserved heterogeneity. The similarity of the RE and the FE estimates suggests a low correlation between individual effects and covariates. The results of FE combined with GMM (FE-GMM), which take into account the potential endogeneity of the quality indicators, are also very similar to the RE and FE estimates. Note that in this model we lose one year of observations due to the inclusion of lagged values for the quality indicators (N=113). The main difference w.r.t. the FE estimates is that the quality coefficients lose their significance. However, as we discuss later in more detail, the quality of our instruments is low. Therefore, the results obtained with FE-GMM could be biased.

In the following we interpret the results obtained with the FE model, as we believe that the bias induced by potential endogeneity of the quality indicators is less severe than the bias induced by weak instruments in the FE-GMM.

The output coefficient (δ_Y) is positive and smaller than 1, suggesting that an increase of 10% in output increases total costs by 7.5%. The coefficient of case-mix (δ_{Q_1}) shows that more severe patients are more costly to treat. The share of labor costs (δ_{P_l}) is estimated at around 90%, while the estimated share of capital is 6% (δ_{P_k}). These values are very close to the actual share costs, 82% and 7% respectively. The form of organization (δ_{IF}) does not seem to affect total costs.

Coefficients	OLS	RE	FE	FE-GMM
δ_Y	0.875***	0.853***	0.751***	0.859***
	(0.017)	(0.019)	(0.046)	(0.110)
δ_{Q_1}	0.277***	0.254***	0.219***	0.299***
	(0.081)	(0.060)	(0.079)	(0.091)
δ_{P_l}	0.874***	0.910***	0.916***	0.923***
	(0.040)	(0.027)	(0.026)	(0.018)
δ_{P_k}	0.062***	0.059***	0.059***	0.067***
	(0.015)	(0.013)	(0.015)	(0.018)
δ_T	0.002	0.004*	0.005***	0.008***
	(0.002)	(0.002)	(0.002)	(0.003)
δ_{IF}	-0.009	-0.007	-	-
	(0.014)	(0.015)	-	-
δ_{Q_2}	0.485***	0.513***	0.480***	0.489***
	(0.089)	(0.069)	(0.071)	(0.074)
$\delta_{Q_{Pain}}$	0.076	0.061**	0.056**	0.055
	(0.046)	(0.028)	(0.027)	(0.067)
$\delta_{Q_{Weight}}$	-0.061	0.098**	0.102**	0.270
	(0.087)	(0.044)	(0.043)	(0.234)
$\delta_{Q_{Antips}}$	0.033	0.034	0.026	0.105
	(0.054)	(0.025)	(0.025)	(0.138)
$\delta_{Q_{Restr}}$	-0.119**	-0.071*	-0.064	0.087
	(0.048)	(0.040)	(0.042)	(0.125)
δ_0	-2.355***	-4.657***	-3.621***	-
	(0.439)	(0.328)	(0.541)	-
R^2	0.984	0.983	0.981	0.988
N	163	163	163	113

Notes: Significance levels: * = 10%, ** = 5%, *** = 1%.

Table 5: Estimated coefficients of OLS, RE, FE, and FE-GMM models.

Consider now the main variables of interest: the quality indicators. The nursing staff ratio (δ_{Q_2}) is highly statistically significant. As expected, the higher the relative number of staff working in a NH, the higher the costs. The coefficient is stable among all the models. Regarding the other four indicators of quality, the magnitude and sign are pretty constant too, but the significance levels slightly change. Both individual-effects models show a positive and significant association between costs and the prevalence of weight loss ($\delta_{Q_{Weight}}$) as well as the prevalence of severe pain ($\delta_{Q_{Pain}}$). The use of daily physical restraints ($\delta_{Q_{Restr}}$) is instead associated with lower costs, but only weakly statistically significant. No association is found between the prevalence of antipsychotic use ($\delta_{Q_{Antips}}$) and costs.

The time trend (δ_T) is statistically significant, but its coefficient is very small. As discussed in a companion paper (Di Giorgio, Filippini et al., 2013), total costs have remained pretty constant since the year 2006 as a consequence of the introduction of global budgets.

We now present the statistics provided by the GMM approach to test the validity of our instruments, when all instruments discussed in section (5) are used. The *Shea Partial* R^2 statistics show that the percentage of variability in the quality indicators explained by the instruments is 10%, 17%, 7% and 12% for Q_{Pain}, Q_{Weight}, Q_{Antips} and Q_{Restr}, respectively. The endogeneity test does not reject exogeneity at the 99% ($\chi^2(4)$=3.081, P-value=0.544). Since our instruments are likely to be weak we cannot be confident that the exogeneity of quality indicators is well assessed. Nevertheless, we believe that in the case of endogeneity, the bias is likely to be very limited due to the institutional setting of the nursing home sector and the strong regulation system. In fact, nursing homes activities are regulated by the local government in a relatively effective way. Therefore, we can plausibly assume that nursing homes do not vary their levels of quality according to the cost of services.

7 Conclusions

To ensure good quality of long term care while keeping costs under control, a better understanding of the relationship between costs and quality is needed. In the NH sector, quality improvements represent a main concern since the ageing of the population is putting the system under financial pressure.

In this paper, we investigated the relationship between costs and quality according to the SPO-framework developed by Donabedian (1988). We used the recently published data on quality indicators for Swiss NHs derived from the RAI. In addition to the nursing staff ratio, we considered two additional indicators of process quality, i.e. the use of antipsychotics for low-risk residents and the presence of daily physical restraints, and two indicators of the outcome quality, i.e. the prevalence of weight loss and the prevalence of severe pain. As compared to previous studies, we estimated an individual effects model based on a panel data that allowed to control for unobserved heterogeneity. We estimated a log-log total costs function and included quality indicators of process and outcomes as covariates. The empirical analysis showed some evidence of a positive relationship between clinical indicators of quality regarding outcomes, the prevalence of severe pain and the prevalence of weight loss, and total costs. We also found some evidence that higher prevalence of daily physical restraint use is associated to lower costs, even though the relationship is weakly significant. The use of antipsychotics is positively related to costs, although not significantly. Finally, staffing levels are strongly correlated with costs.

From a policy point of view, a correlation between costs and quality may suggest that quality aspects should be incorporated in funding schemes designed for nursing home care. Accounting for this correlation may allow the regulator to combine quality and costs objectives and provide appropriate incentives through improved financing schemes for NHs.

References

[1] Angrist, JD, Imbens, GW and Rubin, DB. (1993). Identification of causal effects using instrumental variables. NBER Technical Paper No 136. National Bureauc of Economic Research: Cambridge, MA.

[2] Angrist, JD and Krueger, AB. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. Journal of Economic Perspectives, 15(4): 69–85.

[3] Arling, G, Karon, SL, Sainfort, F, Zimmerman, DR and Ross, R. (1997). Risk Adjustment of Nursing Home Quality Indicators. The Gerontologist, 37(6): 757-766.

[4] Battese, GE and Coelli, TJ. (1995). A Model for Technical Inefficiency Effects in a Stochastic Frontier Production for Panel Data. Empirical Economics, 20: 325-332.

[5] Baum, CF, Schaffer, ME and Stillman, S. (2003). Instrumental Variables and GMM: Estimation and Testing. Working Paper No 545, Boston College Department of Economics. The Stata Journal, 3(1): 1-31.

[6] Berg, K, Mor, V, Morris, J, Murphy, KM, Moore, T and Harris, Y. (2002). Identification and Evaluation of Existing Nursing Homes Quality Indicators. Health care financing review, 23(4): 19-36.

[7] Bhargava, A, Franzini, L and Narendranathan, W. (1982). Serial Correlation and the Fixed Effects Model. The Review of Economic Studies, 49(4): 533-549.

[8] Birnbaum H, Bishop C, Lee A and Jensen G. (1981). Why do nursing home costs vary? The determinants of nursing home costs. Medical Care, 19: 1095-1107.

[9] Bostick JE, Rantz MJ, Flesner MK and Riggs CJ. (2006). Systematic Review of Studies of Staffing and Quality in Nursing Homes. Journal of the American Medical Directors Association, 7(6): 366-376.

[10] Bowblis, JR and Lucas, JA. (2012). The impact of state regulations on nursing home care practices. Journal of Regulatory Economics, 42: 52-72.

[11] Braeutigam, RR and Pauly, MV. (1986). Cost Function Estimation and Quality Bias: The Regulated Automobile Insurance Industry. RAND Journal of Economics, 17: 606-617.

[12] Brekke, KR, Siciliani, L and Straume, OR. (2010). Price and quality in spatial competition. Regional Science and Urban Economics, 40: 471-480.

[13] Breusch, T and Pagan, A. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. The Review of Economic Studies, 47(1): 239-253.

[14] Cameron AC and Trivedi PK. (2005). Microeconometrics. Methods and applications. Cambridge University Press: New York.

[15] Cameron AC and Trivedi PK. (2010). Microeconometrics Using Stata. Revised Edition. Stata Press: College Station, Texas.

[16] Carey, K. (1997). A Panel Data Design for Estimation of Hospital Cost Functions. The Review of Economics and Statistics, 79(3): 443-453.

[17] Castle, NG, and Engberg, J. (2005). Staff Turnover and Quality of Care in Nursing Homes. Medical Care, 43(6): 616-626.

[18] Castle, NG and Engberg, J. (2007). The Influence of Staffing Characteristics on Quality of Care in Nursing Homes. Health Research and Educational Trust, 42(5): 1822-1847.

[19] Castle, NG, Liu, D and Engberg, J. (2008). The Association of Nursing Home Compare Quality Measures with Market Competition and Occupancy Rates. Journal for Healthcare Quality, 30(2): 4-14.

[20] Castle, NG, and Ferguson, JC. (2010). What Is Nursing Home Quality and How Is It Measured? The Gerontologist, 50(4): 426-442.

[21] Cremieux, PY, and Ouellette, P. (2001). Omitted Variable Bias and Hospital Cost. Journal of Health Economics, 20: 271–282.

[22] Crivelli, L, Filippini, M and Lunati, D. (2002). Regulation, ownership and efficiency in the Swiss nursing home industry. International journal of health care finance and economics, 2(2): 79-97.

[23] Di Giorgio L, Filippini, M and Masiero, G. (2012). The impact of the institutional form on the cost efficiency of nursing homes. Working Paper N. 12-03, University of Lugano.

[24] Di Giorgio, L, Filippini, M and Masiero, G. (2013). Implications of global budget payment system on nursing home costs. CEPRA Working Paper N. 1/2013, University of Lugano.

[25] Dillmann, JL, Yeatts, DE and Cready, CM. (2013). Geographic distance, contact and family perceptions of quality nursing home care. Journal of Clinical Nursing, 22(11-12): 1779-1782.

[26] Donabedian, A. (1988). The Quality of Care. How Can It Be Assessed? JAMA, 260(12), 1743-1748.

[27] Dormont, B and Martin, C. (2012). Quality of service and cost-efficiency of French nursing homes. 9th European Conference on Health Economics (ECHE), Zurich, July 18-21, 2012.

[28] Drukker, DM. (2003). Testing for serial correlation in linear panel-data models. Stata Journal, 3: 168–177.

[29] Evans, R. (1971). "Behavioural" Cost functions for hospitals. The Canadian Journal of Economics, 4: 198-215.

[30] Farsi, M, Filippini, M and Kuenzle, M. (2005). Unobserved Heterogeneity in Stochastic Cost Frontier Models: An Application to Swiss Nursing Homes. Applied Economics, 37: 2127-2141.

[31] Farsi, M, Filippini, M and Lunati, D. (2008). Economies of scale and efficiency measurement in Switzerland's nursing homes. Swiss Journal of Economics and Statistics, 144: 359-378.

[32] Forder, J and Allan, S. (2011). Competition in the English nursing homes market. PSSRU Discussion Paper 2820, University of Kent.

[33] Gertler, PJ, and Waldman, DM. (1992). Quality-Adjusted Cost Functions and Policy Evaluation in the Nursing Home Industry. Journal of Political Economy, 100: 1232–1256.

[34] Grabowski, D. (2004). A Longitudinal Study of Medicaid Payment, Private-Pay Price and Nursing Home Quality. International Journal of Health Care Finance and Economics, 4(1): 5-26.

[35] Grabowski, DC, Fengb, Z, Hirthc, R, Rahmanb, M and Mor, V. (2013). Effect of nursing home ownership on the quality of post-acute care: An instrumental variables approach. Journal of Health Economics, 32: 12-21.

[36] Greene, W. (2012). Econometric Analysis. Prentice Hall: NJ.

[37] Harrington, C, Woolhandler, S, Mullan, J, Carrillo, H and Himmelstein, DU. (2001). Does Investor Ownership of Nursing Homes Compromise the Quality of Care? American Journal of Public Health, 91(9): 1452–1455.

[38] Hayashi, F. (2000). Econometrics. Princeton University Press: NJ.

[39] IOM. (1990). Medicare: A strategy for quality assurance (Vol.1). National Academy Press: Washington, DC.

[40] Institute of Medicine (2006). Performance Measurement: Accelerating Improvement. The National Academies Press: Washington, D.C.

[41] Karon, SL, Sainfort, F, Zimmerman, DR. (1999). Stability of Nursing Home Quality Indicators over Time. Medical Care, 37(6): 570-579.

[42] Kendall, MG. (1955). Rank Correlation Methods. Hafner Publishing Co: New York.

[43] Kezdi, G. (2004). Robust Standard Errors Estimation in Fixed-Effects Panel Models. Hungarian Statistical Review Special, 9: 96-116.

[44] Laine, J, Linna, M, Häkkinen, U and Noro, A. (2005a). Measuring the productive efficiency and clinical quality of institutional long-termcare for the elderly. Health Economics, 14: 245-256.

[45] Laine, J, Linna, M, Noro, A and Häkkinen, U. (2005b). The Cost Efficiency and Clinical Quality of Institutional Long-Term Care for the Elderly. Health Care Management Science, 8: 149-156.

[46] Legido-Quigley, H, McKee, M, Nolte, E and Glinos, IA. (2008). Assuring the quality of health care in the European Union. European Observatory on Health Systems and Policies: Observatory Studies Series No. 12.

[47] McKay, NL. (1989). Quality choice in Mediaid markets: The case of nursing homes. Medical Care, 35: 367-385.

[48] Mukamel, DB. (1997). Risk-adjusted outcome measures and quality of care in nursing homes. Medical Care, 35: 367-385.

[49] Mukamel, DB, Glance, LG, Li, Y, Weimer, DL, Spector, WD, Zinn, JS and Mosqueda, L. (2008). Does risk adjustment of the CMS quality measures for nursing homes matter? Medical Care, 46(5): 532-541.

[50] Mukamel, DB and Spector, WD. (2000). Nursing Home Costs and Risk-Adjusted Outcome Measures of Quality. Medical Care, 38(1): 78-89.

[51] Mor, V, Berg, K, Angelelli, J, Gifford, D, Morris, J and Moore, T. (2003). The Quality of Quality Measurement in U.S. Nursing Homes. The Gerontologist, 43(II): 37-46.

[52] Mor, V, Morris, J, Lipsitz, L and Fogel, B. (1998). Benchmarking quality in nursing homes: the Q-Metrics System. Canadian Journal of Quality in Health Care,14: 12-17.

[53] Nakrem, S, Vinsnes, AG, Harkless, GE, Paulsen, B and Seim, A. (2009). Nursing sensitive quality indicators for nursing home care: International review of literature, policy and practice. International Journal of Nursing Studies, 46: 848-857.

[54] Newhouse, JP. (1994). Frontier Estimation: How Useful a Tool for Health Economics? Journal of Health Economics, 13: 317-322.

[55] Nyman, JA. (1988). Improving the quality of nursing home outcomes. Medical Care, 26: 1158-1171.

[56] Rantz, MJ, Hicks, L, Grando, V, Petroski, GF, Madsen, RW et al. (2004). Nursing Home Quality, Cost, Staffing, and Staff Mix. The Gerontologist, 44(1), 24-38.

[57] Rogers, WH. (1994). Regression standard errors in clustered samples. Stata Technical Bullettin, 13: 19-23.

[58] Sangl, J., Bernard, S., Buchanan, J., Keller, S., Mitchell, N., Castle, N. G., et al. (2007). The development of a CAHPS instrument for nursing home residents. Journal of Aging and Social Policy, 19: 63-82.

[59] Shea, J. 1997. Instrument relevance in multivariate linear models: A simple measure. Review of Economics and Statistics, 79(2): 348-352.

[60] Spilsbury, K, Hewitt C and Stirk, L. (2011). The relationship between nurse staffing and quality of care in nursing homes: A systematic review. International Journal of Nursing Studies, 48: 732-750.

[61] Starkey, KB, Weech-Maldonado, R and Mor, V. (2005). Journal of Health Care Finance, 32(1): 67-81.

[62] Stock, JH and Watson, MW. (2006). Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression. NBER Technical Working Paper 323.

[63] Weech-Maldonado, R, Shea, D and Mor, V. (2006). The Relationship Between Quality of Care and Costs in Nursing Homes American Journal of Medical Quality, 21(1): 40-48.

[64] Wodchis, WP, Teare, GF and Anderson, GM. (2007). Cost and Quality Evidence From Ontario Long Term Care Hospitals. Medical Care, 45: 981–988.

[65] Wooldridge, JM. (2002). Econometric Analysis of Cross Section and Panel Data. MIT Press: Cambridge, MA.

[66] Wooldridge, JM. (2003). Introductory Econometrics. A modern approach. (2nd Ed.). South-Western College Pub: Cincinnati, Ohio.

[67] Zimmerman, D. (1995). Development and testing of nursing home quality indicators. Health Care Financing Review, 16(4): 107-127.

[68] Zinn, JS, Aaronson, WE.and Rosko, MD. (1993a). The use of standardized indicators as quality improvement tools: An application in Pennsylvania nursing homes. American Journal of Medical Quality, 8: 72-78.

[69] Zinn, JS, Aaronson, WE and Rosko, MD. (1993b). Variations in the outcomes of care provided in Pennsylvania nursing homes: facility and environmental correlates. Medical Care, 31: 475-487.

[70] Zinn, JS, Aaronson, WE and Rosko, MD. (1994). Strategic groups, performance, and strategic response in the nursing home industry. Health Service Research, 29: 187-205.