Working paper No: 21/2011 May 2011 LSE Health

Irene Papanicolas, Alistair McGuire

Using a latent variable approach to measure the quality of English NHS hospitals

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Using a latent variable approach to measure the quality of English NHS hospitals

Irene Papanicolas, Alistair McGuire

First Published in May 2011

Working paper no. 21/2011

LSE Health
The London School of Economics and Political Science
Houghton Street
London WC2A 2AE

British Library Cataloguing in Publication Data A catalogue record for this publication is available from the British Library ISBN [978-0-85328-463-5]

Corresponding author
Irene Papanicolas
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Email: I.N.Papanicolas@lse.ac.uk

Acknowledgments The authours would like to acknowledge Professor Elias Mossialos for his comments and support on this piece of work.

Abstract

The provision of performance information is essential for ensuring and improving the performance of health care systems. However, the lack of reliable quality information is a key problem in evaluating and improving health care. This paper estimates the performance of English NHS Acute Trusts using Dr. Foster data from the years 1996 – 2008 to investigate health outcomes after elective treatment for Acute Myocardial Infarction (AMI) and Hip Replacement. A latent variable approach is used to calculate the hospital quality effect in determining short and long term mortality and readmission rates for each year. These measures are then used to compare current and past quality of care within and across NHS Acute Trusts. Our results support that this method is well suited to measuring provider quality of care. Using these quality measures we are able to investigate the performance of NHS Acute Trusts across this time period and identify hospitals where further scrutiny of low quality is required in the future.

Keywords: Measuring Quality; Latent Variables; Health.

Contents

1 Introduction

Often timely and relevant data on quality of care does not exist, because it is costly and difficult to collect. Yet, even when such data is available it is not straightforward to use. We are still far from having complete data sets informing us of all the factors that influence outcomes, or measures for the appropriate clinical processes of care required to obtain good outcomes. Indeed, while work on the evidence base of medicine is growing, and many conditions have been identified as amenable to health care (Holland, 1988; Nolte et al., 2004), there still remains considerable uncertainty about the clinical effectiveness of over half of current medical practice (Tovey, 2007; Maynard, 2008). Even with good data, multidimensionality is a problem - as quality of medical care has many dimensions: outcomes, processes and others - ideally all of which would be integrated into a quality evaluation (McClellan and Staiger, 1999).

Increasingly measures of health outcome are used to inform policy, whether it is through investigations into surgical performance (Spiegelhalter et al., 2002), to produce publicly available indicators of performance for hospitals (Healthcare Commission, 2004; New York State Department of Health, 2004), or to produce research to evaluate health care reforms (Farrar et al., 2009; Jarman et al., 1999). Use of outcomes to compare quality of care assumes that variation attributable to other factors can be properly accounted for, such that any residual variation in outcomes, such as observed mortality and morbidity, is indicative of variation in quality of care (Lilford et al., 2004).

While outcome measures are influenced by quality of care they are also a result of data collection techniques, data quality, patient case-mix, and chance. Definitions of outcomes can vary considerably across institutions influencing the comparability of data. Even with a simple outcome such as death, systematic differences can arise with definitions and the way they are applied across institutions, for example by classifying patients who come to the hospital to die under the palliative code Z51.5 or not (Hawkes, 2010a). Moreover, trying to identify groups, or cases, of patients whose outcome is being compared can also be difficult. Some cases such as Child Birth or Fractured Neck of Femur are very easy to identify, but other areas such as Stroke or Infertility are harder to classify, and data coding and collection may vary across institutions (Lilford et al., 2004). Both these issues pose challenges in using data for to compare providers.

The type of information will also differ markedly when obtained from different sources such as case records or administrative data. Routine administrative data often contains little information on co-morbidity and severity of disease which will have a large impact

on outcome. Clinical databases, run by various bodies, may record more detailed clinical information, but are also likely to vary considerably with regards to data quality (Aylin et al., 2007). In England, HES is often regarded as unreliable by clinicians because of problems in its early years, notably its inability to secondary diagnoses for most patients (McKee and James, 1997; McKee et al., 1999). However, since then data quality has improved considerably (Audit Commission, 2004). Indeed Aylin et al. (2007) note that, "if suitable predictive models could be developed using this routinely collected information source, they would be a valuable tool for generating measures of performance adjusted for case mix".

The most commonly discussed challenge of using outcome measures as a quality metric is accounting for differences in patient case-mix. Fair comparison of quality between providers needs to consider the differences in patient factors that will influence outcomes, such as patient severity or co-morbidity. Typically, some sort of risk adjustment is employed to address the attribution problems associated with outcome measures, and control for the other influencing factors. Yet, even when outcome measures are risk adjusted, as is the case for hospital standardised mortality ratios (HSMRs) for example, they still run the risk of not accounting for factors that cannot be identified and measured accurately. Mistaking such errors for differences in quality is known as "case-mix fallacy" (Lilford et al., 2004).

Finally, we should not forget the presence of random error, which is inherent to any measurement. Random error can take the form of type 1 (false negatives) or type 2 (false positives) error. The only way to limit the amount of both types of error is to ensure the sample size is large enough. Unfortunately this is not always possible when dealing with health care, as it depends on the prevalence of particular treatments and conditions. Many studies have noted the challenges associated with sample size and outcomes, which limit the number of procedures that can be assessed (Birkmeyer et al., 2002; Dimick et al., 2004).

In using outcome measures as quality metrics, the challenge lies in controlling for all these factors in order to extract the true quality signal from the measure. With the exception of chance, every other factor has components that can be measured and others that cannot (Lilford et al., 2004). Case-mix adjustment is often used to control for variance in outcome, indeed there is a wide literature on different statistical methods used to produce case-mix adjusted outcomes (Jarman et al., 2005; Iezzoni, 1994; Iezzoni et al., 1996; Iezzoni, 1997; Shahian et al., 2001, 2010). However, worryingly enough different case-mix adjustment methods can produce different results, identifying different providers

as good or bad performers depending on the adjustment technique used (Iezzoni et al., 1996; Shahian et al., 2010).

Even if a case-adjustment technique was uniformly applied to providers outcomes could still vary systematically across providers due to differences in the other areas. For example, outcomes could differ because of systematic factors such as systematic differences in data recording, or systematic differences to patient behaviours before and after treatment. For example, hospitals that treat less educated patients may have worse outcomes because they have lower adherence to medications after treatment. An alternative method to control for noise was put forward by McClellan and Staiger (1999) who used information on individual patient characteristics at the individual hospital level to adjust hospital quality measures at the hospital level and thus take account of any systematic biases which are embedded within these measures. The approach, in recognition that the measurement of hospital quality is difficult, begins with latent variable approach to address measurement issues, before going on to refine this measure with a vector autoregression framework.

A latent variable is a variable that can not be observed directly, such as hospital quality. Latent variable assessment takes the observable data and combines it to make assumptions about the unobservable, latent, phenomenon. The latent variable measured by this method will consist of the 'true' variance with both random and systematic errors. This type of technique has been used extensively in the areas of psychological and education testing (Hambleton and Cook, 1977), political science (Treier and Jackman, 2008), and increasingly in epidemiology (Muthén, 1989). In educational testing, information is collected about the subject's ability through their answers on various questions which are indicative of their underlying ability. Conditional on the latent trait, the multiple responses are assumed to be independent observations, and thus correlation amongst the responses is induced by variability in the latent ability amongst the subjects. Thus by modelling their responses the latent ability can be estimated (Landrum et al., 2000).

This model can be extended to health, where the latent trait being modelled is provider quality. By using hospital-specific intercepts derived from a patient level equation which maps quality of outcome (e.g. 30-day mortality) against patient characteristics we are able to create latent measures of quality, as they pick up the unmeasured systematic aspects which are retained after controlling for observable variation in patient outcomes. While these latent outcome measures will still be noisy, they will filter out much of the estimation error that is otherwise present due to systematic differences in patient mix across hospitals rather than differences in care. Additionally, as these estimates are normally calculated using large samples of patients it is possible to eliminate much of the noise inherent

in raw outcome measures, that make quality measures difficult to interpret. While a latent variable approach is essentially another case-mix adjustment technique it is an improvement to other methods as the latent measure provides a composite measure of quality for each provider than result in different outcomes.

These latent outcome estimates of hospital quality thus have a number of attractive features. They can incorporate information on quality measures in a systematic manner, are relatively easy to compute from available data and overcome the risk of over-estimation which is common when aggregate data are combined with individual observations. These hospital intercepts, which estimate the mean value of quality measure holding patient characteristics constant, are therefore less noisy and less likely to be inconsistent estimates than crudely observed aggregate measures of hospital quality. In this paper, we will explore the use of latent variable modelling to measure quality of health care providers. We find that this methodology can be useful for quality measurement in situations where random or systematic measurement error is a problem, where the phenomenon under study is not directly observable, and where many indicators are needed to describe the different aspects of the phenomenon. The next section will outline the empirical methods used in the paper, before we go on to describe the data used for the model. Finally, we will present the results of the analysis and conclude.

2 Empirical Model

This paper introduces the basic approach which analyses the determinants of hospital quality through a two-step process, with hospitals being the unit of analysis. In the first step, individual patient level data is used to create latent outcome measures at the hospital level using multiple individual outcome measures (mortality and readmission rates), adjusted for individual level patient characteristics. This process allows the amount of noise surrounding these outcome measures to be minimized, thus creating more robust quality measures at the hospital level. The latent outcome measures are then used in the second step to examine how different hospital and social characteristics influence quality of care.

2.1 Creating Latent Outcome Measures

The first step of the analysis uses the quality measures provided at the individual patient level over a given amount of years to estimate the relative difference in the mean value

of outcomes of each hospital holding patient characteristics constant. These hospital intercepts are estimated using the following equation:

$$Y_{iht}^{k} = \beta q_{1h}^{k} + \sum \phi X_{jht} + u_{iht}\,, \tag{1}$$

where Y^k represents the quality outcome measure (mortality and readmission rates at different time intervals), with i denoting the individual patient, h the hospital and t the year. $\sum \phi$ represents a group of individual control variables for patient characteristics (age, gender, socioeconomic deprivation, co-morbidities and type of admission). While β represents the fixed effects of all NHS hospitals in which treatment was provided to the sample of patients. The model is run with no constant term, and thus a variable for every hospital can be included. As there is no reference case, the β^k coefficients will represent the intercept value of the hospital's mortality/readmission regression.

This patient level regression is run separately for each year t and quality outcome measure k. By saving the β^k coefficients for each regression, we obtain a vector of hospital intercepts for each quality outcome measure k, and year t. These estimates of quality are appealing as they provide relative rankings of hospital performance contributing to the quality outcomes, controlling for other observable influences. These estimates thus allow some of the noise to be removed from the outcome measures, and thus enable a more appropriate comparison of hospital trusts to one another.

2.2 Panel Data Estimation with Lagged Variables

Just as the performance of hospitals in obtaining a desired outcome for their patients will be influenced by the characteristics of patients, it will also depend on other factors such as the characteristics of the hospital itself. The second step of this analysis uses the latent quality measures calculated in the manner explained above and examines how they are associated with key hospital characteristics. This will inform us about what factors our quality metrics are associated with, but can also help draw conclusions about how well the latent indicator performs.

Hospital performance is likely to be influenced by hospital characteristics such as the type of hospital. Different types of hospitals (teaching, acute, specialist, foundation) may be associated with different quality care because of different underlying incentives or management models. For example, in addition to delivering medical care to patients, a teaching hospital will also provide clinical education and training to future health professionals, and

also invests in research and technology. These functions may result in different objectives and managerial style which can also contribute to quality differences. Specialist trusts and foundation trusts were introduced to the NHS in 2004. Specialist trusts are dedicated to providing elective care, while a foundation trust is a high performing hospital that has received more managerial and financial freedoms. These differences are likely to have an effect on hospital performance.

Other factors such as the number of patients treated (caseload) may also contribute to overall performance. The relationship between cases and outcomes is not clear. Increased caseload may result in lower quality due to overcrowding, or it can result in higher caseload as doctors become more experienced. Moreover, higher quality may lead to more cases as demand increases, or it can be the result of selecting fewer cases. Similarly average deprivation and co-morbidity is likely to be correlated with the latent outcome measure if there is a confounding relationship between them, that is if they are both correlated with another variable that influences quality. For example if more deprived patients are less likely to adhere to treatment after being discharged and thus are more likely to have bad outcomes.

However, it is unlikely that hospital performance is instantaneously influenced by a change in any of these variables, because of institutional, technological or even psychological reasons (Gujarati, 2003). In an institution structural forces dominate, at least in the short term. For example contractual obligations may prevent hospitals from switching sources of labour, on ancillary services immediately. Thus in certain regards institutions are 'locked in' to current conditions, at least until the medium term. Technological reasons may relate to the adoption of medical innovations. There will be a time gap between the introduction of a new technology and its adoption in routine health service provision. Finally, the psychological factors refer to the inertia of the status quo, managers and employers take time to adjust to change, thus even if some other factor such as changes in prices occur it may take a transition period for this change to translate into behaviour. Moreover, the performance of a hospital is likely to depend on its past performance. Thus it is important to consider how lags of hospitals own performance, as well as of the other explanatory variables influence the latent variable.

In order to examine the relationship between the exogenous factors discussed above and hospital performance, as measured by the latent variable, the following equations are estimated:

$$D30_{ht} = \alpha + \beta_1 D30_{h(t-n)} + \beta_2 \sum X_{ht} + \beta_3 \sum X_{h(t-1)} + \beta_4 H_{ht} + \epsilon_{ht} \quad (2)$$

$$D365_{ht} = \alpha + \beta_1 D365_{h(t-n)} + \beta_2 \sum X_{ht} + \beta_3 \sum X_{h(t-1)} + \beta_4 H_t + \epsilon_{ht} \quad (3)$$

$$R28_{ht} = \alpha + \beta_1 D30_{ht} + \beta_2 R28_{h(t-n)} + \beta_3 \sum X_{ht} + \beta_4 \sum X_{h(t-1)} + \beta_5 H_t + \epsilon_{ht} \quad (4)$$

$$R365_{ht} = \alpha + \beta_1 D365_{ht} + \beta_2 R365_{h(t-n)} + \beta_3 \sum X_{ht} + \beta_4 \sum X_{h(t-1)} + \beta_5 H_t + \epsilon_{ht} \quad (5)$$

where $D30_{ht}$ and $D365_{ht}$ denote the 30-day and 365-day hospital mortality intercepts gained from the first stage analysis, representing the latent mortality for each hospital h at year t, and where $R28_{ht}$ and $R365_{ht}$ denote the 30-day and 365-day term hospital readmission intercepts, representing latent readmissions for each hospital h at year t. The lag variables $D30_{h(t-n)}$ and $D365_{h(t-n)}$ take into account the latent mortality measures of n years prior to year t, while variables X_1 and X_2 control for hospital type and treatment characteristics of the hospital such as average length of stay of patients, number of cases admitted and average waiting time). H_t represents yearly dummies which are intended to capture any contemporaneous shocks that may influence quality.

In any model that includes a lagged dependent variable there is an inherent problem of autocorrelation, which is magnified when the time-series dimension of the data is small (Nickell, 1981). The problem arises because of the correlation of the lagged dependent variable and the error term, and results in making the estimators inconsistent. Including additional regressors does not remove this bias, and if they are correlated with the lagged dependent variable their coefficients may also be seriously biased. This problem can be addressed by estimating a dynamic panel data model, which uses the first differenced Generalized Method of Moments (GMM) estimator (Arellano and Bond, 1991). However, the Arellano-Bond estimator may not perform well if the autoregressive parameters are too large and the time series observations are moderately small. This problem is addressed by the later work of Arellano and Bover (1995) and Blundell and Bond (1998) which impose additional restrictions on the initial conditions process. The Blundell-Bond estimator is used in this analysis, given the small time-series component available in the newly constructed panel dataset. In addition the Blundell-Bond estimator is able to incorporate lagged levels as well as lagged differences, which increases the efficiency of the model by allowing us to add additional instruments such as hospital characteristics.

The `xtabond2` command in statistical package STATA is used to perform the analysis. A two-step estimator was used, as it is asymptotically more efficient. Robust standard errors were requested, which ensures that the `xtabond2` command includes a finite-sample

correction to the two-step covariance matrix derived by Windmeijer (2005), which can make two-step robust estimations more efficient than one-step robust, especially for system GMM, with lower bias and lower standard errors. Following the recommendations made by Roodman (2006) the model is specified so that every repressor is included in instrument matrix, with endogenous variables, such as the lagged dependent variable, specified from two lags, and exogenous predetermined variables specified from one or two lags. Finally, the `xtabond2` command reports the results of the Arellano-Bond AR(1) and AR(2) tests for autocorrelations, as well as the Sargan and Hansen test statistics which indicate how well specified the model is.

3 Data

The data used to conduct this analysis is Hospital Episode Statistics (HES) which documents hospital activity in England. The HES database has been in existence since 1987 and is/has been used by the Department of Health to provide performance information at the hospital level (Spiegelhalter et al., 2002). Hospital episode statistics (HES) contain records for all NHS patients admitted to English hospitals in each financial year (April 1 to March 31), with information on all medical and surgical specialties, including private patients treated in NHS hospital trusts. The HES data holds over 15 million patient records each year, stored according to the financial year in which the period of care was completed. Each NHS hospital is required to submit data items for each episode in every patient's stay in that hospital. Data is entered from patient's notes onto hospital administration systems by trained clinical coders (Aylin et al., 2007).

Diagnosis of patients are coded using ICD-10 (international statistical classification of diseases, tenth revision) codes while procedures use the UK Office of Population Censuses and Surveys classification (OPCS4). Since the introduction of the internal market, HES data has also been used for contracting between purchasers and providers. The data available in the HES database contains patient characteristic data (e.g. gender, age), clinical information (e.g. diagnoses, procedures undergone), mode of admission (emergency, elective), outcome data (mortality, readmission, discharge location) as well as details on the amount of time spent in contact with the health system (waiting times, date of admission, date of discharge) and details of which hospital the patient was treated in. The HES data we used was accessed through Dr. Foster Intelligence, an independent association dedicated to providing high quality health information.

While HES data is a rich source of information, it requires some manipulation in order to ensure that the total care received by a patient is measured under the same episode. HES data measures the care received under one consultant during the course of the patient's treatment, in the case that the patient is treated by more than one consultant it is important to identify such patients and link their records of care to provide a complete picture of their care experience. Dr. Foster has done the matching within the HES data and is able to provide information on the complete patient experience. In addition they have linked to other data sources such as the death registries, to provide additional information such as death rates at different intervals (30-days and yearly), readmission rates and further details on the patient, such as further information on their co-morbidities and on some socioeconomic characteristics.

Data on gender and age are used as explanatory variables in the analysis, as is a variable indicating whether the treatment undergone was an elective procedure. The Charlson co-morbidity index which predicts the 1 year mortality for a patient who may have a range of co-morbid conditions was used to control for severity of patients. This index is constructed by assigning a score to each condition depending on the risk of dying associated with it, and summing these scores up (Charlson et al., 1987). Finally, socio-economic status was measured using the Carstairs index of deprivation. This index is based on four census indicators: low social class, lack of car ownership, overcrowding and male unemployment, which are combined to create a composite score. The deprivation score is divided into seven separate categories which range from very low to very high deprivation.

We requested data for seven conditions for the financial years 1996-2008, namely: Acute Myocardial Infarction (AMI), Myocardial Infarction (MI), other Ischemic Heart Disease (IHD), Congestive Cardiac Failure (CCF), Stroke, Transient Ischemic attack (TIA) and Hip Replacement. While the analysis is conducted in this paper, only the results for AMI and Hip Replacement are presented as these are common conditions and the results are similar for the other conditions. Some statistics for the other five conditions can be found in Appendix A. Seven conditions were chosen in order to be able to evaluate how well the quality measures performed for different clinical areas, different sample sizes and different type of admissions. The data for these conditions was extracted based on the ICD-10 and OPCS 4.3 classification codes indicated in Table 1 and Table 5. In most cases there were problems with the sample sizes of some of the years before 2000, and so these years were not included in the analysis. Any hospital trust that had less than 10 admissions throughout the entire period of analysis was dropped from the analysis. Moreover, any primary care trusts, private trusts acting as NHS providers and social care trusts were also excluded. For the sample of patients admitted with AMI, only emergency admissions

were examined, and only for patients with a length of stay greater than two days. For the patients admitted with Stroke and Congestive Cardiac Failure, all patients admitted as day cases were excluded.

AMI otherwise referred to as a heart attack, refers to an acute blockage of an artery that provides blood to the heart muscle, it is a major health event that almost always results in hospitalization. AMI is commonly used for quality assessment purposes because of the large sample size available, the established link between treatment and survival, and because it is usually an emergency admission which makes patient selection by providers difficult. Hip Replacement occurs in instances where the hip joint is surgically replaced by a prosthetic implant. Hip Replacement surgery can be performed as total replacement or a half replacement. Total Hip Replacement is defined as the surgical removal of the entire hip joint (ball and socket) and its replacement by a prosthetic (metal or synthetic plastic) implant. A half replacement or hemiarthroplasty, being a slightly less invasive procedure, refers to the replacement of only the ball part of the joint (femoral head). Hip Replacements can be undertaken as elective conditions, with the vast majority of such cases being undertaken to treat Arthritis. The procedure may be the result of an emergency to replace a broken hip. This condition was selected as it is not dominated by emergency admissions, indeed elective hip arthroplasty are extremely common and extremely successful. Also Hip Replacement is an easily identifiable condition, which will be easily coded and provide us with a large sample of patients to work from.

Tab. 1: Summary Statistics of the Sample.

Condition	ICD-10/OPCS4.3 codes	Years Analyzed	Mean cases per year	Number of hospitals
AMI	ICD-10: 121	2000-2008	399,560	139
Hip	OPCS4.3: W37-W39 W46-W48 W58	1996-2008	40,564	125

To undertake the first analysis outcomes are defined through the use of mortality and readmission rates at different intervals are used as dependent variables. These variables are directly reported in the data, and represent noisy estimates of hospital outcome. The model was estimated for 30-day within hospital mortality rates, designed to measure if a patient dies within up to 30 days in hospital after their initial admission for treatment (with a value of 1 in the case of death and 0 otherwise), 365-day overall mortality rates, 28-day readmission rates which measure whether a patient is readmitted for the same condition in a 28 day period (with a value of 1 for readmission and 0 otherwise), and 365 day readmission rates.

A trust code is used to distinguish each hospital in the data, allowing us to identify which hospitals are performing better or worse. Data on gender and age are used as explanatory variables in the analysis, as is a variable indicating whether the treatment undergone was an elective procedure. The Charlson co-morbidity index was used to control for patient co-morbidity, and the Carstairs index of deprivation was used to control for socio-economic status.

The second part of the analysis considers how explanatory variables at the hospital level influence the latent quality measures. In order to do this we take the latent measures constructed for every outcome measure, for every year and create a new panel data set at the hospital level. This data set contains the newly created latent outcome measures for each year, along with the hospital characteristic information that was available in the individual data. In addition, it is possible to estimate a set of aggregate variables for each year in the data corresponding to the average socioeconomic and demographic characteristics of the patients treated by each hospital, such as mean co-morbidity of patients treated, mean deprivation of patients treated, number of admissions, mean length of stay and mean waiting times. Again for the sake of brevity, and as the results are generally similar we concentrate on two sets of results in the paper (AMI and Hip Replacement and report the others in Appendix A.

4 Results

This paper examines the measurement of quality at the hospital level using the two models described previously. Model 1 refers to the model used to construct the latent measures (equation (1)) and Model 2 to the analysis of the latent measures (equations (2)–(5)). The results are presented for the two models, separately by condition. The results for AMI, and Hip Replacement, are presented in the results section. The results for the remaining five conditions are similar in that they demonstrate the methodology is reproducible for all areas. Some basic results presented in Appendix A, and discussed in the conclusions. The results for the other five conditions are summarised in Appendix A and full details can be forwarded if required.

4.1 Model 1

In order to compare how the latent variables perform relative to the raw outcome data we first present a graphical representation of the average value of the four outcome measures

being studied (30-day and 365-day mortality; and 28-day and 365-day readmissions) for each condition. These are indicative of raw outcomes across hospitals, over time, before any analysis is undertaken. In each of these figures, the dashed lines represent the 95% confidence intervals of the estimates representing variation among hospitals. These graphs are useful in order to visually represent the difference between the latent measures and the raw measures.

In Model 1, the unit of analysis is the individual patient, and the main focus of interest is the relationship between individual patient's death rates/ readmission rates and the quality of the hospital at which they were treated, controlling for patient characteristics. In the first model a linear regression is used (estimated separately for each year and each outcome indicator) to measure the hospital specific effects that contribute to mortality and readmission rates, controlling for age, gender, deprivation, co-morbidities and whether the procedure undertaken was elective or emergency. The model is estimated as a regression-through-the-origin, thus R^2 value is not meaningful and not reported. The results of these regressions are presented for each condition and indicate how each of the control variables influence the outcomes measured as the dependent variables. The sign and magnitude for each variable in each regression is as expected, such that age, gender, co-morbidity, deprivation and type of admission significantly influence outcomes for most conditions.

From the Model 1 regressions, the hospital intercepts are extracted and saved, and used as a 'latent outcome measures' of the unobservable hospital specific effect on mortality and readmissions. The latent outcome measure averaged across hospitals is graphically illustrated over time, separately for each outcome measure and condition studied. In these figures, the solid line represents the mean latent outcome measure of all hospitals in the sample, over the time period being evaluated, while the dashed lines indicate the 95% confidence intervals for these estimates representing variation among hospitals. If we imagine hospital quality to be the true underlying signal, surrounded by random, systematic and measurement error, than each hospital's intercept indicates the slope of that curve when graphed over time. As the latent variables are the hospital coefficients in each of the outcome equations, a negative value indicates a fall in the raw outcome attributable to unobserved hospital effects, while a positive value indicates a rise in the raw outcome. While the 95% confidence intervals represent the variation across hospitals. These diagrams allow us to visualise the average latent measures over time and draw some conclusions about trends in quality. The latent measures show a clearer change in quality over time than apparent from the noisy outcome data. This is consistent with the trends as shown by other risk adjusted measures, such as HSMRs, which indicate that quality is improving (Hawkes, 2010b).

In order to observe the trend in latent outcome measures at the hospital level, the time trend of latent values are illustrated for a selection of four hospitals in each condition, represented in four panels. Of the four hospitals included, there is always a small hospital (upper left), a large hospital (lower right), and two midsize hospitals, where size of hospital is determined relative to the average caseload per year per condition. These hospitals are not a random sample, but chosen to illustrate the data in different settings. The solid line in each of these panels indicates the latent outcome measure estimated from the linear model for the selected hospital, while the two dashed lines indicate the 95% confidence intervals of these estimates. The latent outcome measures have been normalized such that the mean aggregate outcome value of all hospitals for each year is equal to zero, and any deviation from the mean indicates above or below average performance for that hospital. Negative values indicate lower mortality than average, and positive values indicate higher mortality than average, controlling for patient characteristics. The solid lines can be interpreted as absolute outcome differences, for example a value of 0.02 indicates that the hospital's mortality was 2% above the average hospital in that year. These figures allow an interpretation of how the individual hospitals are performing relative to their peers in all areas evaluated. They indicate that the latent measures are easy to use as performance indicators at the individual hospital level.

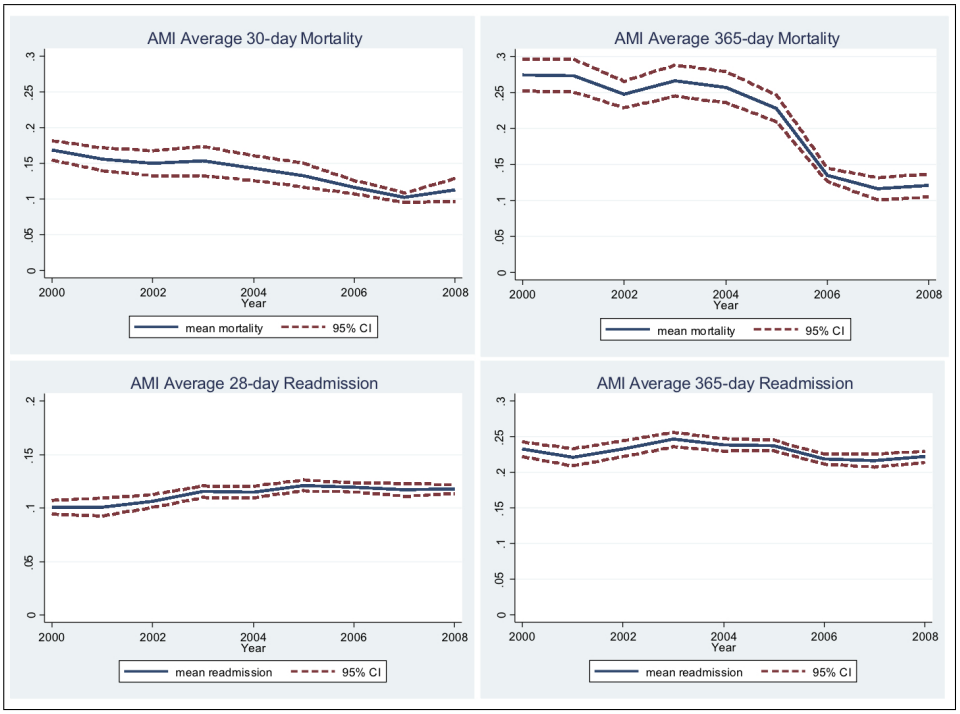
4.1.1 AMI

Figure 1 indicates the trends over time in raw average mortality and readmission rates for AMI. The trend in average AMI 30-day mortality across hospitals is downward, with short term mortality falling gradually over time, at a gradual pace. Average 365-day mortality is also falling, yet while this trend is gradual for most years there is a large sudden drop from 2005 to 2006, which is not present in the 30-day mortality trend. The trend in average 28-day readmissions indicates an almost negligible increase over time. While, average 365-day readmissions stay relatively constant, rising and falling marginally over the time period studied, such that readmissions at the end of the time period are around the same level as they were in the beginning.

The regression results from Model 1, presented in Table 2, indicate that patient characteristics such as age, gender, deprivation and co-morbidities are almost always significant at high levels for all four outcome indicators. Gender significantly impacts mortality and readmissions, such that women have a marginally higher mortality, and higher readmission rates, than men. Existing co-morbidities, as measured by the Charlson Co-morbidity index, significantly increase mortality and readmissions, as do increased deprivation as

measured by the Carstairs score. Only emergency admissions are considered for AMI, and so type of admission is not controlled for. Trust dummies included for each hospital and year are highly significant for all four outcome measures.

Fig. 1: Trends across years in average AMI outcome measures across hospitals.



Tab. 2: Regression results for AMI Model 1.

Year	N (total)	Age	Gender	Carstairs Score	Co-morbidity	Trust dummies
30-Day Mortality						
		(0.000)	(0.003)	(0.000)	0.001	
2001	43986	0.004***	0.011***	3.82E-04	0.037***	yes
		(0.000)	0.003	(0.000)	-0.001	
2002	44619	0.004***	0.007***	0.001***	0.039***	yes
		(0.000)	0.003	(0.000)	0.001	
2003	44160	0.004***	0.009***	8.16e-04*	0.040***	yes
		(0.000)	0.003	(0.000)	0.001	
2004	43426	0.004***	0.009***	-2.18E-05	0.035***	yes
		(0.000)	0.003	(0.000)	0.001	
2005	40186	0.004***	0.019***	5.35E-04	0.035***	yes
		(0.000)	0.003	(0.000)	0.001	

Year	N (total)	Age	Gender	Carstairs Score	Co-morbidity	Trust dummies
2000	32950	0.004***	0.006*	0.001**	0.043***	yes
2006	37743	0.003***	0.002	0.001***	0.031***	yes
		(0.000)	0.003	(0.000)	0.001	
2007	36240	0.003***	0.001***	1.61E-04	0.032***	yes
		(0.000)	0.003	(0.000)	0.001	
2008	33607	0.003***	0.001	4.29E-04	0.031***	yes
		(0.000)	0.003	(0.000)	0.001	
365-Day Mortality						
2000	37346	0.011***	0.010**	0.004***	0.090***	yes
		(0.000)	-0.004	-0.001	-0.002	
2001	49780	0.011***	0.016***	0.003***	0.086***	yes
		(0.000)	-0.004	-0.001	-0.002	
2002	50711	0.011***	0.017***	0.004***	0.084***	yes
		(0.000)	-0.004	-0.001	-0.002	
2003	50202	0.011***	0.027***	0.004***	0.086***	yes
		(0.000)	-0.004	-0.001	-0.002	
2004	49762	0.011***	0.014***	0.003***	0.078***	yes
		(0.000)	-0.004	-0.001	-0.002	
2005	46914	0.010***	0.020***	0.004***	0.075***	yes
		(0.000)	-0.004	-0.001	-0.002	
2006	45133	0.006***	0.012***	0.003***	0.042***	yes
		(0.000)	-0.003	-0.001	-0.002	
2007	44026	0.005***	0.017***	0.002***	0.040***	yes
		(0.000)	-0.003	-0.001	-0.001	
2008	41474	0.005***	0.009***	0.001**	0.039***	yes
		(0.000)	-0.003	-0.001	-0.001	
28-Day Readmission						
2000	32950	4.80e-04***	0.007*	0.002***	0.010***	yes
		(0.000)	-0.004	-0.001	-0.002	
2001	43986	5.71e-04***	0.013***	0.002***	0.009***	yes
		(0.000)	-0.003	-0.001	-0.002	
2002	44619	5.40e-04***	0.011***	0.001**	0.011***	yes
		(0.000)	-0.003	-0.001	-0.002	
2003	44160	6.64e-04***	0.022***	0.001**	0.008***	yes

Year	N (total)	Age	Gender	Carstairs Score	Co-morbidity	Trust dummies
2000	32950	0.004***	0.006*	0.001**	0.043***	yes
		(0.000)	-0.003	-0.001	-0.002	
2004	43426	9.75e-04***	0.008***	0.002***	0.011***	yes
		(0.000)	-0.003	-0.001	-0.002	
2005	40186	0.001***	0.017***	0.002***	0.013***	yes
		(0.000)	-0.004	-0.001	-0.002	
2006	37743	0.001***	0.015***	0.001	0.014***	yes
		(0.000)	-0.004	-0.001	-0.002	
2007	36240	0.001***	0.016***	0.003***	0.015***	yes
		(0.000)	-0.004	-0.001	-0.002	
2008	33607	0.001***	0.017***	0.002***	0.013***	yes
		(0.000)	-0.004	-0.001	-0.002	
365-Day Readmission						
2000	37346	0.001***	0.008*	0.005***	0.018***	yes
		(0.000)	-0.005	-0.001	-0.003	
2001	49780	0.001***	0.022***	0.007***	0.020***	yes
		(0.000)	-0.004	-0.001	-0.002	
2002	50711	0.001***	0.017***	0.005***	0.021***	yes
		(0.000)	-0.004	-0.001	-0.002	
2003	50202	0.002***	0.026***	0.005***	0.018***	yes
		(0.000)	-0.004	-0.001	-0.002	
2004	49762	0.002***	0.023***	0.005***	0.020***	yes
		(0.000)	-0.004	-0.001	-0.002	
2005	46914	0.001***	0.026***	0.005***	0.022***	yes
		(0.000)	-0.004	-0.001	-0.002	
2006	45133	0.001***	0.030***	0.004***	0.022***	yes
		(0.000)	-0.004	-0.001	-0.002	
2007	44026	0.002***	0.027***	0.006***	0.022***	yes
		(0.000)	-0.004	-0.001	-0.002	
2008	41474	0.002***	0.029***	0.004***	0.020***	yes
		(0.000)	-0.004	-0.001	-0.002	

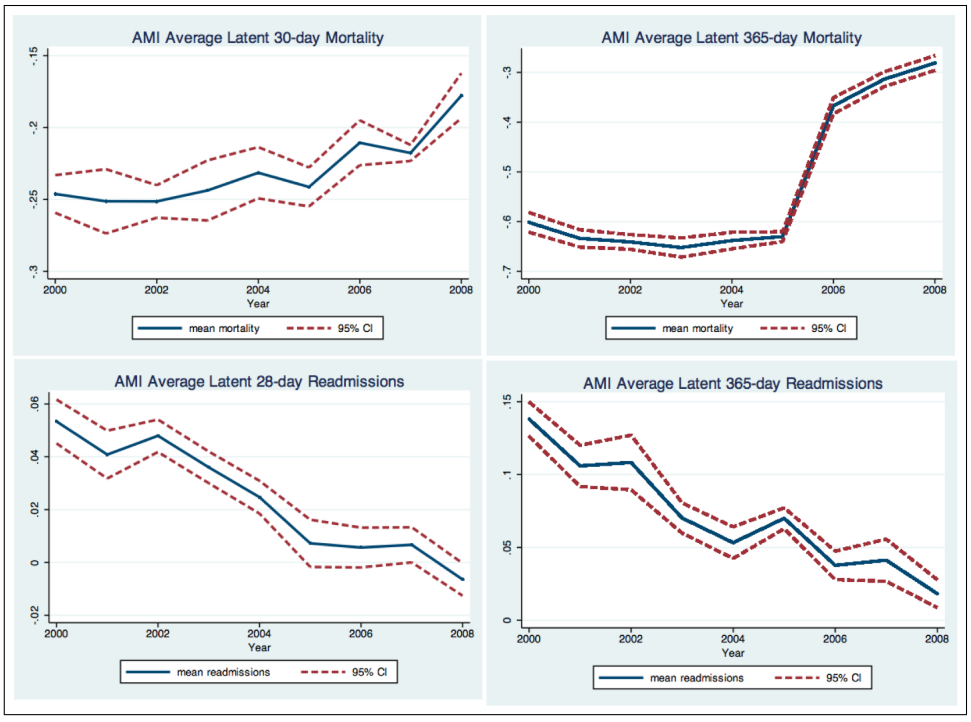
* Significant at $p \leq 0.1$

** Significant at $p \leq 0.05$

*** Significant at $p \leq 0.01$

Figure 2 shows the trends over time in average latent mortality and readmission rates for AMI, such that the curve in each panel represents the rate of change in the raw outcomes over time controlling for patient characteristics. The mean hospital latent mortality outcomes are negative for both 30-day and year-long mortality, indicating that on average rate of change in mortality attributable to each hospital is decreasing. However, for both outcomes the values are becoming more positive over time, meaning that the mean is decreasing at an increasing rate. The confidence intervals for 30-day mortality show a variation of just under 2% in the beginning of the sample, which narrow after 2005 to about 1%. This indicates less variation in the quality of hospitals towards the end of the time period. The panels for readmissions are mostly positive, indicating increasing readmissions attributable to hospital performance are increasing, on average. However, in both readmission panels the points are approaching zero indicating that they are increasing at a decreasing rate. Indeed for 28-day readmissions the value falls below zero for 2007-2008, indicating decreasing readmissions in this period.

Fig. 2: Trends across years in average latent AMI outcome measures across hospitals.



Figures 3 and 4, show the trend in AMI latent 30-day and 365-day mortality for four selected hospitals. The confidence intervals for both figures, show more variation in latent mortality within hospitals than indicated by either of the averages plotted in Figure 2. For both short term and long term mortality, estimates within hospitals range from over 5% above average to more than 7% below average. There hospital specific panel indicate year-

to-year variations of performance, commonly around 3-4% in either direction. Figures 5 and 6 show AMI latent 28-day and 365-day readmissions for the same four hospitals. The confidence intervals for both readmission figures show about the same degree of within hospital variation among latent readmissions than was observed for latent mortality, of about 5%. The magnitude of the year-to-year variation, is also similar, but varies according to the hospital.

The latent outcome measures graphed by individual hospital allow quality comparisons to be made between different providers, and examination of their own quality trajectory through time. For example, in Figure 3, the midsize hospital in the bottom left hand panel is clearly providing above average quality for the entire time period, as the estimate and the entire 95% confidence interval lie below 0, or the average mortality rate of all hospitals. However, the trend of the estimate over suggests that while mortality rates are below average, they are steadily increasing relative to its peers. Looking at the latent long term mortality rates for the same hospital (Figure 4) indicates that the provider performs relatively worse on this outcome measure in the later time periods. While 365-day latent mortality rates started off below average in 2000, they have steadily increased throughout the sample until they are unequivocally above average in 2008.

Fig. 3: Trends across years in latent AMI 30-day mortality for selected hospitals.

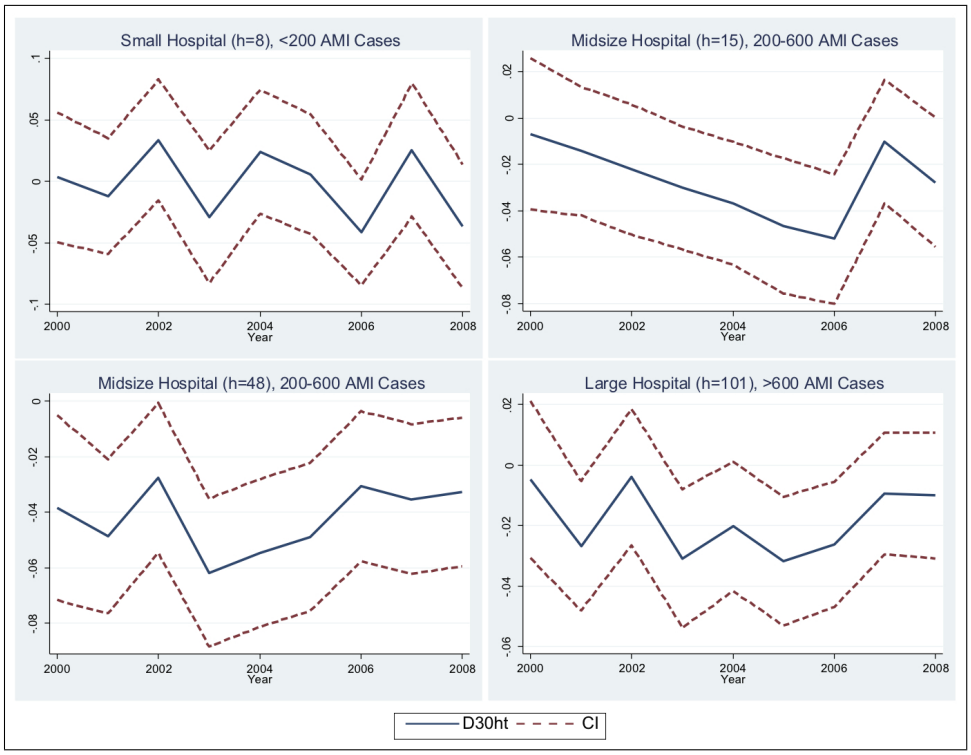


Fig. 4: Trends across years in latent AMI 365-day mortality for selected hospitals.

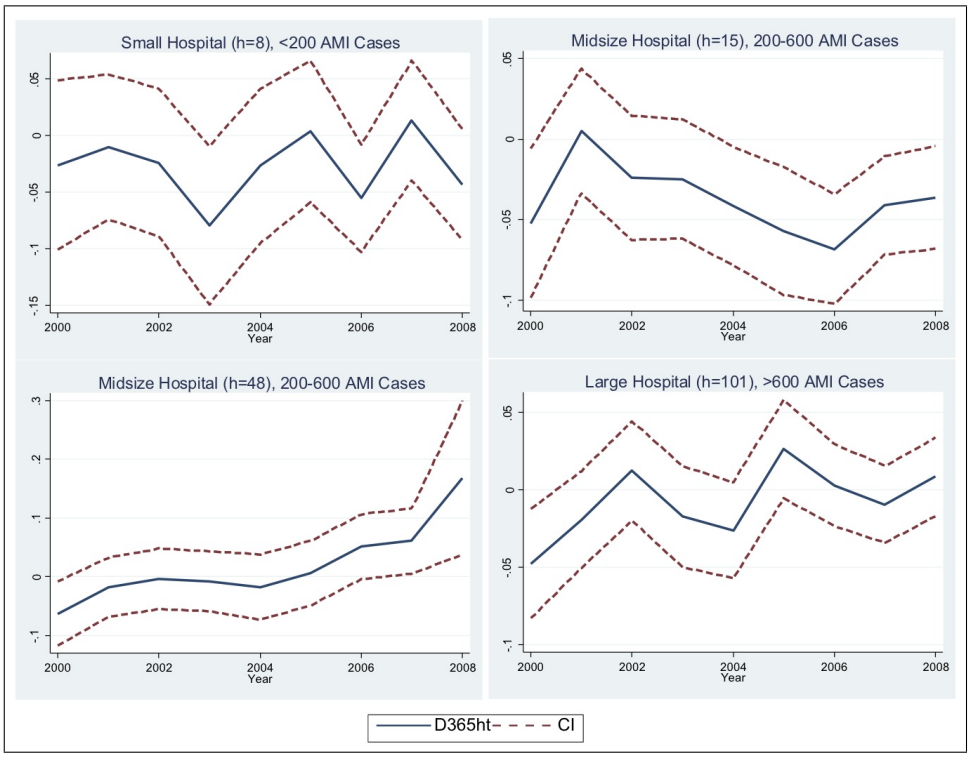


Fig. 5: Trends across years in Latent AMI 28-day readmissions for selected hospitals.

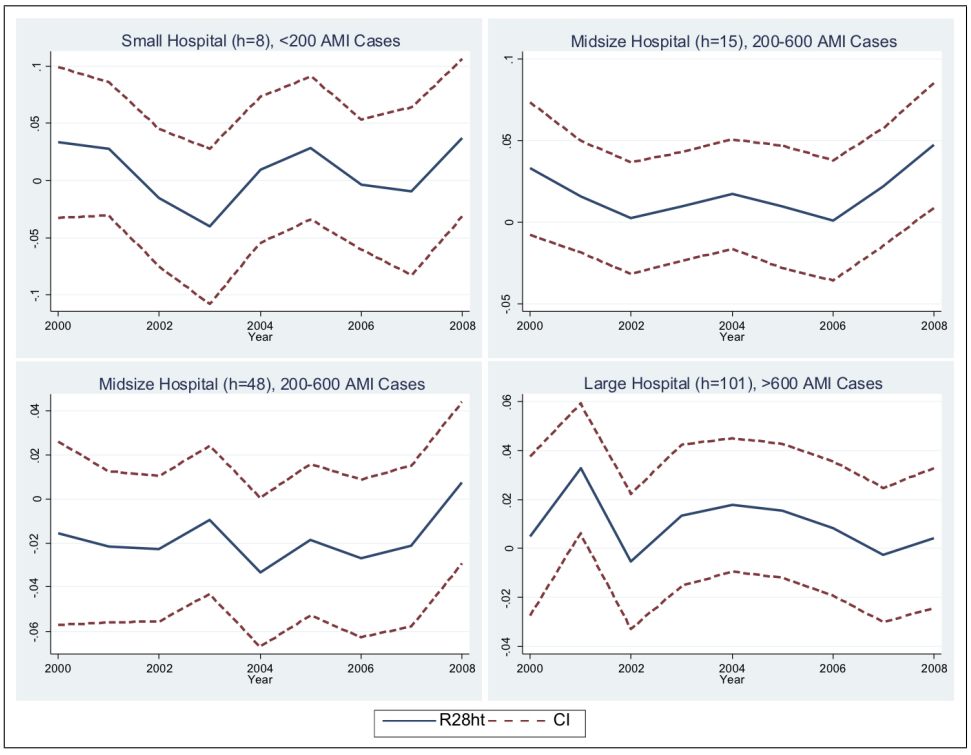
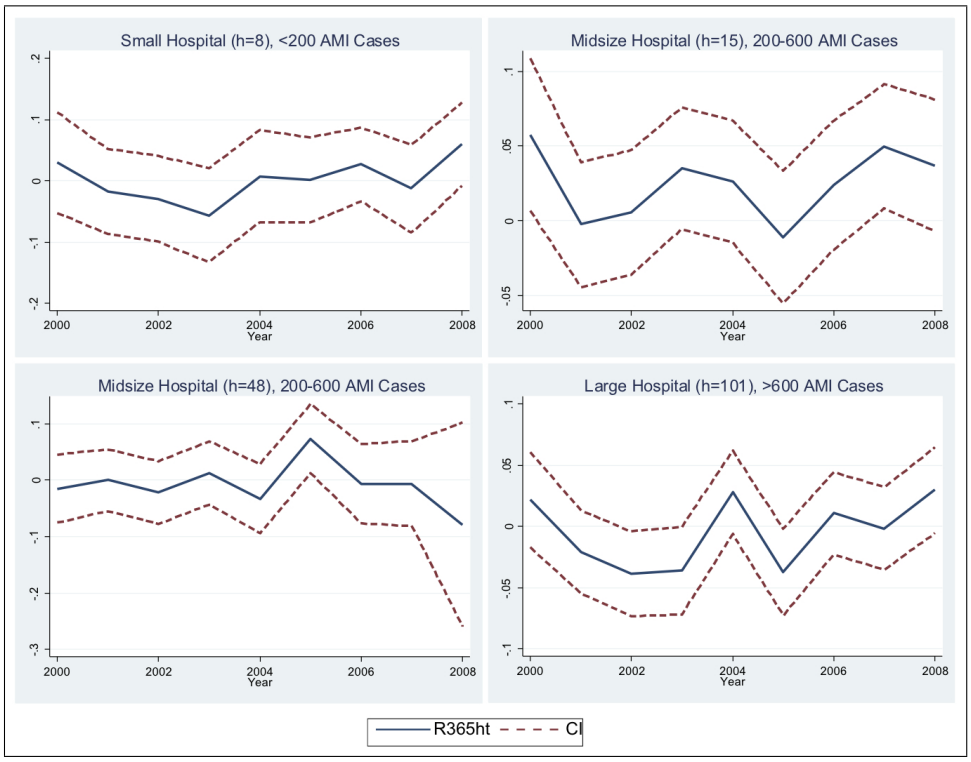


Fig. 6: Trends across years in latent AMI 365-day readmissions for selected hospitals.



4.1.2 Hip Replacement

Average 30-day mortality for Hip Replacement in the years 1996-2008, presented in Figure 7 exhibits a relatively constant trend during the entire time period. The confidence intervals surrounding the estimate narrow slightly from 2004 onwards, suggesting less variation in outcomes among hospitals from that point onwards. The average 365-day mortality rates for Hip Replacement shown in Figure 7 do not display a constant trend. Instead there is a noticeable increase from the year 1999 which is sustained until 2006, where mortality returns to its 2008 level. Similar to Figure 1, the confidence intervals surrounding the estimate narrow from 2005 onwards. Average 28-day and 365-day readmissions for Hip Replacement both show a slight upwards trend throughout the 2000-2008 period. In both figures there is a widening of confidence intervals in the year 2001, suggesting a larger variation in readmissions amongst hospitals for that year. For average 365-day readmissions only, there is sharp increase from 2006-2007.

Fig. 7: Trends across years in average Hip Replacement outcome measures across hospitals.

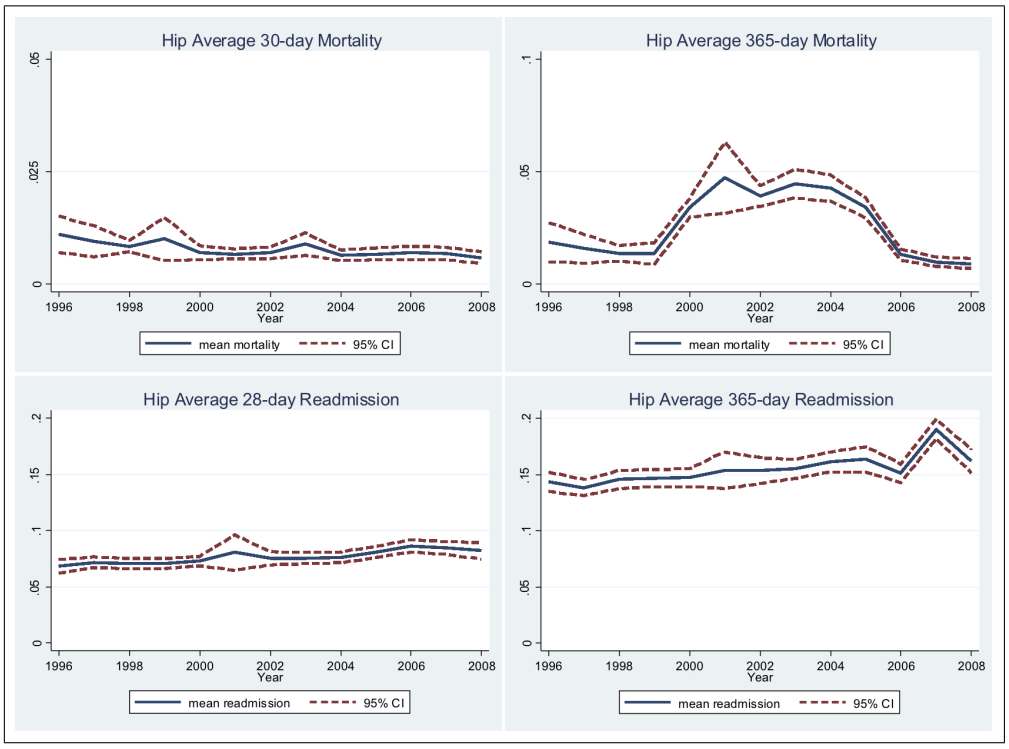


Table 3 presents the regression results from Model 1 for all four outcome indicators. In both mortality regressions age, gender, co-morbidity and type of admission is significant such that higher age and co-morbidity leads to increased mortality. In addition, women have a marginally higher mortality than men, and non-elective admissions have higher mortality than elective admissions. Deprivation is significant for only some of the years in both regressions. Where significant, higher deprivation is associated with higher mortality. The trust dummies were always significant. All the explanatory variables included in the 28-day and 365-day readmissions models were significant, such that older patients, more deprived patients and patients with co-morbidities had higher rates of readmissions in addition to women and patients that were admitted for elective procedures.

Tab. 3: Regression results for hip Model 1.

Year	N (total)	Age	Gender	Carstairs Score	Co-morbidity	Elective	Trust dummies
30-Day Mortality							
1996	25835	4.29e-04***	0.004***	-2.12e-04	0.021***	0.015***	yes
		(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	

Year	N (total)	Age	Gender	Carstairs Score	Co-morbidity	Elective	Trust dummies
1997	29952	4.12e-04***	0.002**	-2.48e-04	0.019***	0.017***	yes
		(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	
1998	34559	5.34e-04***	0.004***	2.44e-04	0.020***	0.015***	yes
		(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	
1999	36527	4.41e-04***	0.002**	1.81e-04	0.020***	0.019***	yes
		(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	
2000	36864	4.55e-04***	0.004***	5.31e-05	0.017***	0.014***	yes
		(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	
2001	38745	4.47e-04***	0.002**	4.20e-04**	0.015***	0.016***	yes
		(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	
2002	41502	4.39e-04***	0.002**	4.73e-04***	0.017***	0.015***	yes
		(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	
2003	44759	4.68e-04***	0.001	9.38e-05	0.017***	-0.011***	yes
		(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	
2004	47124	3.87e-04***	0.003***	2.24e-04	0.012***	-0.012***	yes
		(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	
2005	46507	4.06e-04***	0.002***	1.52e-04	0.013***	-0.013***	yes
		(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	
2006	45438	3.76e-04***	0.003***	1.39e-04	0.013***	-0.013***	yes
		(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	
2007	47232	3.64e-04***	0.002***	1.26e-04	0.011***	-0.011***	yes
		(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	
2008	48243	3.35e-04***	0.002***	1.29e-04	0.008**	-0.008***	yes
		(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	
365-Day Mortality							
1996	25835	0.001***	0.005***	-1.64e-04	0.033***	0.030***	yes
		(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	
1997	29952	0.001***	0.003***	-1.59e-04	0.032***	0.028***	yes
		(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	
1998	34559	0.001***	0.005***	3.10e-04	0.028***	0.026***	yes
		(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	
1999	36527	0.001***	0.002**	3.41e-04	0.026***	0.029***	yes
		(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	

Year	N (total)	Age	Gender	Carstairs Score	Co-morbidity	Elective	Trust dummies
2000	36864	0.002***	0.010***	3.00e-04	0.057***	0.064***	yes
		(0.000)	(0.002)	(0.000)	(0.002)	(0.003)	
2001	38745	0.002***	0.006***	0.002***	0.059***	0.089***	yes
		(0.000)	(0.002)	(0.000)	(0.002)	(0.003)	
2002	41502	0.002***	0.004**	0.002***	0.054***	0.079***	yes
		(0.000)	(0.002)	(0.000)	(0.002)	(0.003)	
2003	44759	0.002***	0.003*	0.001***	0.058***	0.082***	yes
		(0.000)	(0.002)	(0.000)	(0.002)	(0.003)	
2004	47124	0.002***	0.007***	0.002***	0.058***	0.077***	yes
		(0.000)	(0.002)	(0.000)	(0.002)	(0.003)	
2005	46507	0.002***	0.005***	0.001***	0.044***	0.070***	yes
		(0.000)	(0.002)	(0.000)	(0.001)	(0.002)	
2006	45438	0.001***	0.004***	2.88e-04	0.019***	0.028***	yes
		(0.000)	(0.010)	(0.000)	(0.001)	(0.001)	
2007	47232	0.001***	0.003***	2.46e-04	0.014***	0.021***	yes
		(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	
2008	48243	0.001***	0.002***	3.76e-04**	0.010***	0.023***	yes
		(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	
28-Day Readmission							
1996	25835	4.63e-04***	0.014***	0.001	0.008***	0.053***	yes
		(0.000)	(0.003)	(0.001)	(0.003)	(0.005)	
1997	29952	5.13e-04***	0.009***	0.002**	0.009***	0.046***	yes
		(0.000)	(0.003)	(0.001)	(0.003)	(0.004)	
1998	34559	4.36e-04***	0.018***	0.002**	0.008***	0.053***	yes
		(0.000)	(0.003)	(0.001)	(0.003)	(0.004)	
1999	36527	6.01e-04***	0.013***	1.23e-04	0.014***	0.055***	yes
		(0.000)	(0.003)	(0.001)	(0.003)	(0.004)	
2000	36864	5.99e-04***	0.012***	0.2e-06***	0.008***	0.052***	yes
		(0.000)	(0.003)	(0.001)	(0.003)	(0.004)	
2001	38745	9.41e-04***	0.0017**	0.002***	0.006***	0.059***	yes
		(0.000)	(0.003)	(0.001)	(0.002)	(0.004)	
2002	41502	7.54e-04***	0.013***	0.002***	0.006***	0.070***	yes
		(0.000)	(0.002)	(0.001)	(0.002)	(0.001)	

Year	N (total)	Age	Gender	Carstairs Score	Co-morbidity	Elective	Trust dummies
2003	44759	8.63e-04***	0.014***	0.002***	0.011***	0.067***	yes
		(0.000)	(0.003)	(0.001)	(0.002)	(0.001)	
2004	47124	9.01e-04***	0.016***	0.002***	0.006***	0.068***	yes
		(0.000)	(0.002)	(0.001)	(0.002)	(0.001)	
2005	46507	9.55e-04***	0.018***	0.002***	0.012***	0.066***	yes
		(0.000)	(0.003)	(0.001)	(0.002)	(0.001)	
2006	45438	0.001***	0.019***	0.002***	0.011***	0.068***	yes
		(0.000)	(0.003)	(0.001)	(0.002)	(0.001)	
2007	47232	0.001***	0.014***	0.003***	0.013***	0.075***	yes
		(0.000)	(0.003)	(0.001)	(0.002)	(0.001)	
2008	48243	0.001***	0.014***	0.003***	0.011***	0.076***	yes
		(0.000)	(0.002)	(0.001)	(0.002)	(0.001)	
365-Day Readmission							
1996	25835	0.002***	0.013***	0.002**	0.026***	0.164***	yes
		(0.000)	(0.005)	(0.001)	(0.005)	(0.007)	
1997	29952	0.002***	0.017***	0.003***	0.021***	0.147***	yes
		(0.000)	(0.004)	(0.001)	(0.004)	(0.006)	
1998	34559	0.002***	0.025***	0.004***	0.023***	0.152***	yes
		(0.000)	(0.004)	(0.001)	(0.004)	(0.006)	
1999	36527	0.002***	0.016***	0.002***	0.023***	0.159***	yes
		(0.000)	(0.004)	(0.001)	(0.004)	(0.006)	
2000	36864	0.002***	0.011***	0.003***	0.022***	0.167***	yes
		(0.000)	(0.004)	(0.001)	(0.004)	(0.005)	
2001	38745	0.002***	0.0124***	0.004***	0.018***	0.165***	yes
		(0.000)	(0.004)	(0.001)	(0.003)	(0.005)	
2002	41502	0.002***	0.011***	0.003***	0.018***	0.185***	yes
		(0.000)	(0.004)	(0.001)	(0.003)	(0.005)	
2003	44759	0.002***	0.012***	0.004***	0.027***	0.184***	yes
		(0.000)	(0.003)	(0.001)	(0.003)	(0.005)	
2004	47124	0.002***	0.015***	0.005***	0.017***	0.178***	yes
		(0.000)	(0.003)	(0.001)	(0.003)	(0.005)	
2005	46507	0.002***	0.017***	0.004***	0.022***	0.164***	yes
		(0.000)	(0.003)	(0.001)	(0.003)	(0.005)	

Year	N (total)	Age	Gender	Carstairs Score	Co-morbidity	Elective	Trust dummies
2006	45438	0.002***	0.013***	0.003***	0.018***	0.165***	yes
		(0.000)	(0.003)	(0.001)	(0.003)	(0.005)	
2007	47232	0.003***	0.006*	0.006***	0.026***	0.186***	yes
		(0.000)	(0.003)	(0.001)	(0.003)	(0.005)	
2008	48243	0.002***	0.015***	0.005***	0.024***	0.174***	yes
		(0.000)	(0.003)	(0.001)	(0.002)	(0.005)	

* Significant at $p \leq 0.1$

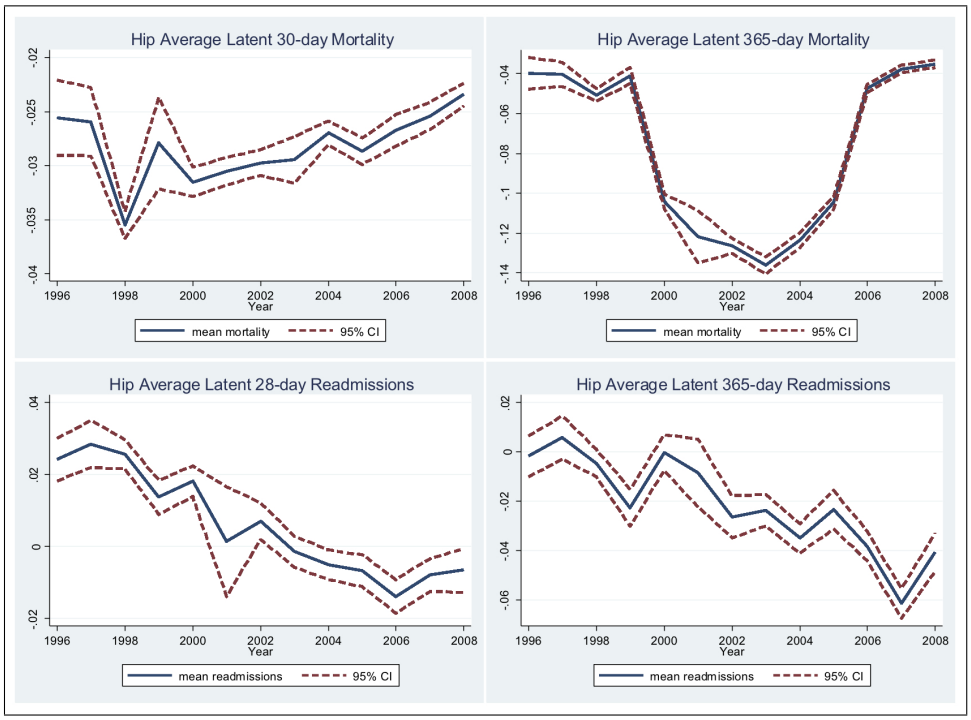
** Significant at $p \leq 0.05$

*** Significant at $p \leq 0.01$

Figure 8 shows the average Hip Replacement 30-day latent outcome estimates across hospitals in the time period 1996-2008. Both mortality estimates are negative throughout the period being investigated. This indicates that on average, the rate of change in mortality, controlling for patient characteristics is falling. For both short and long term mortality, the trajectory of the average hospital intercept indicates that initially the rate of change is decreasing at an increasing rate, but after 2000 it begins to decrease at a decreasing rate. This pattern is much more pronounced for year-long mortality than 30-day mortality. It is also apparent in both panels that the confidence intervals at the end of the period are narrower than in the first years of the sample.

The hospital intercepts for both readmission estimates vary around zero. The average latent outcome for 28-day readmissions indicates increasing average readmissions attributable to hospital performance, but at a declining rate. Around 2001, the average becomes negative, indicating that hospitals are contributing towards declining readmissions. The average latent outcomes estimated from 365-day readmissions are negative for all years but 1997, indicating that throughout the period readmissions are falling. As the values are becoming increasingly more negative we can say that they are falling at a decreasing rate.

Fig. 8: Trends across years in average latent Hip Replacement outcome measures across hospitals.



Figures 9 – 12 show the latent mortality and readmission estimates for four selected hospitals treating Hip Replacement patients. The within hospital and year-to-year variation in latent mortality is smaller than all other conditions aside from Hip Replacement, ranging around $1-5\%$ below and above both 30-day aggregate mortality and 28-day readmission measures, and around $5-10\%$ below and above the 365-day aggregate mortality and aggregate readmission measures. While there is year-to-year variation this is usually around $2-4\%$ in either direction. There is wider variation in the confidence interval and the year-to-year variation for the small hospital for all conditions.

Fig. 9: Trends across hospitals in latent Hip 30-day mortality for selected hospitals.

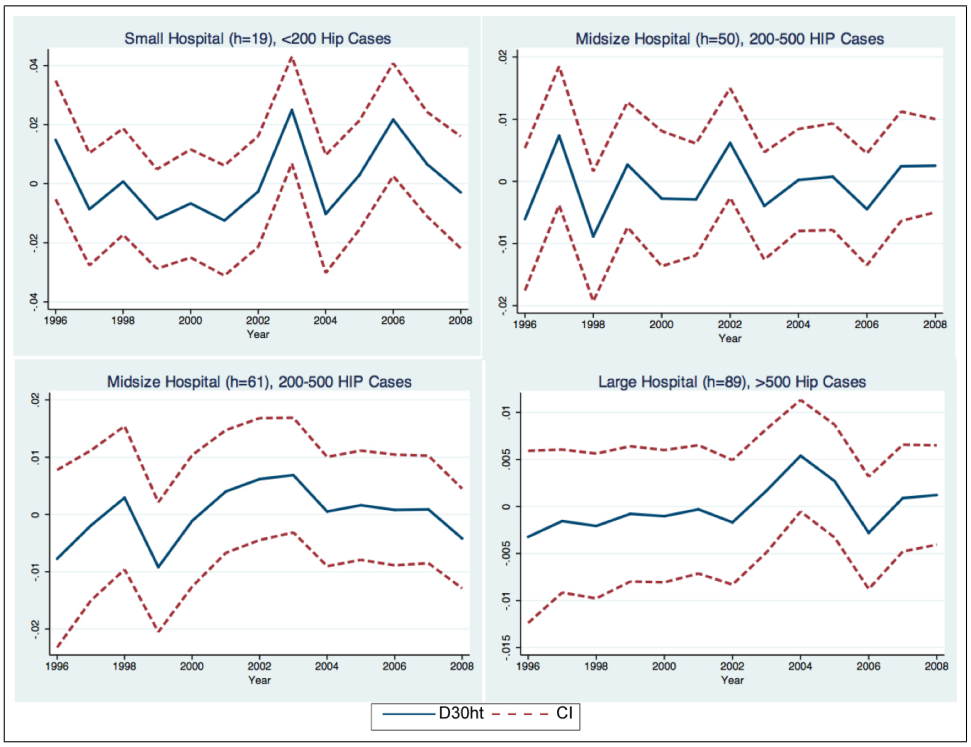


Fig. 10: Trends across years in latent Hip 365-day mortality for selected hospitals.

Fig. 11: Trends across years in latent Hip 28-day readmissions for selected hospitals.

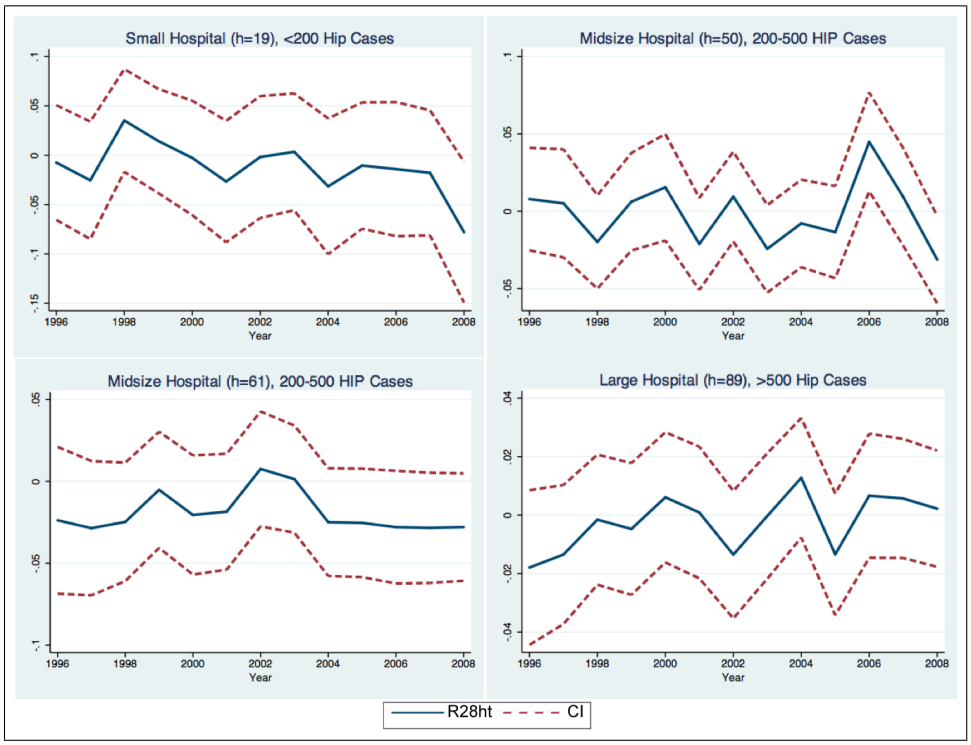
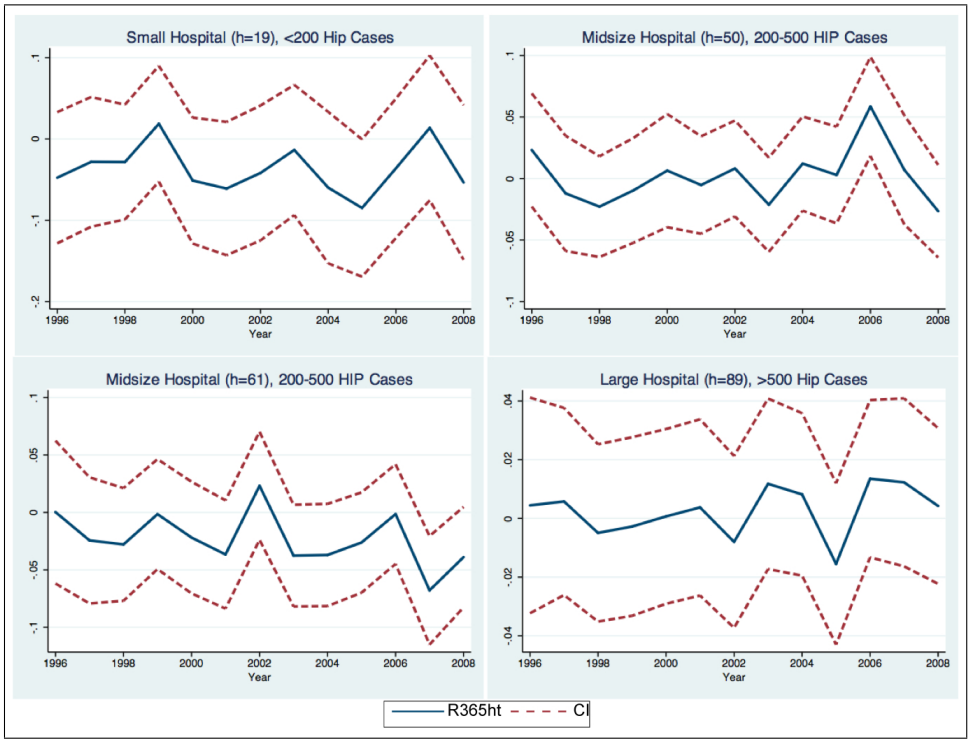


Fig. 12: Trends across years in latent Hip 365-day readmissions for selected hospitals.



4.2 Model 2

The second model uses the aggregate outcome measures as dependent variables to test for fundamental relationships amongst outcomes and hospital characteristics. Table 4 presents the results for the regressions estimated for Model 2, one for each of the aggregate outcome measures, for each condition, run separately for AMI and Hip Replacement. The number of instruments for each model is reported together with the regression results. As only an emergency AMI conditions were included in the sample, waiting times and lagged waiting times were not included in any of the AMI models. Additionally, different specifications of the models were run for the different conditions, and the one which best met model fit criteria is reported. For this reason lagged waiting times and length of stay are sometimes not included in selected models. Most models passed the AR(1) test for autocorrelation with over 95% confidence. All models also passed the Sargan test for instrument validity with over 95% confidence, rejecting the null hypothesis that the overidentifying assumptions are valid. The models indicate that few hospital characteristics are significant in influencing the change in latent outcomes over time. However, for most conditions some element of performance is dynamic – demonstrating that change does not occur instantaneously but is incremental.

Tab. 4: Model 2 regression results for AMI and Hip Replacement.

	AMI	Hip	AMI	Hip	AMI	Hip	AMI	Hip
	$D30_{ht}$	$D30_{ht}$	$D365_{ht}$	$D365_{ht}$	$R28_{ht}$	$R28_{ht}$	$R365_{ht}$	$R365_{ht}$
L. Latent	0.961***	-0.603***	-0.194	-0.450***	-0.0775	-0.0459	-0.448***	-0.821
Mortality	(0.209)	(0.103)	(0.181)	(0.111)	(0.0589)	(1.247)	(0.149)	(0.527)
L. Latent	-	-	-	-	0.210	0.0894	-0.220**	0.0497
Readmissions					(0.313)	(0.205)	(0.110)	(0.183)
L. LOS	0.00738***	-0.00147*	0.00471***	-	-0.000251	-0.00113	0.00132**	-0.00107
				0.00421***				
	(0.00113)	(0.000815)	(0.00100)	(0.00149)	(0.000389)	(0.00131)	(0.000560)	(0.00235)
L. Cases	-0.000188	6.09e-05**	0.000250	0.000103	-1.52e-05	3.66e-05	-5.91e-06	3.14e-06
	(0.000394)	(3.01e-05)	(0.000185)	(7.49e-05)	(1.60e-05)	(6.13e-05)	(2.74e-05)	(2.09e-05)
L. Waiting	-	2.56e-05	-	0.000144	-	-5.10e-05	-	-4.33e-05
Times	-	(4.97e-05)		(9.69e-05)		(0.000133)	-	(0.000164)
Cases	-0.000329	-5.47e-05	-0.00110**	-0.000138	1.93e-05	5.87e-05	3.56e-05	2.53e-05
	(0.000525)	(3.85e-05)	(0.000529)	(8.79e-05)	(1.56e-05)	(0.000102)	(2.40e-05)	(4.06e-05)
$Cases^2$	5.92e-07	-7.12e-09	7.83e-07*	3.34e-08	-	-	-	-
	(6.63e-07)	(2.42e-08)	(4.35e-07)	(4.65e-08)				

	AMI	Hip	AMI	Hip	AMI	Hip	AMI	Hip
	$D30_{ht}$	$D30_{ht}$	$D365_{ht}$	$D365_{ht}$	$R28_{ht}$	$R28_{ht}$	$R365_{ht}$	$R365_{ht}$
LOS	-0.0111***	0.00240**	0.00506**	0.00780***	-0.000906	0.00517**	-0.00206	0.00740
	(0.00144)	(0.000940)	(0.00231)	(0.00211)	(0.00114)	(0.00241)	(0.00215)	(0.00642)
Waiting	-	-2.04e-05	-	-0.000146	1.77e-05	-3.11e-05	3.87e-05	1.20e-05
Times	-	(5.11e-05)	-	(0.000104)	(1.47e-05)	(6.60e-05)	(2.49e-05)	(2.15e-05)
Specialist	0.0603	0.00400	-0.0487	0.0105	-0.000906	0.00295	-0.00135	0.00484
Trust	(0.0723)	(0.00523)	(0.0475)	(0.00698)	(0.00114)	(0.00230)	(0.00177)	(0.00346)
Foundation	-0.00120	-0.000743	-0.00101	-0.00107	0.00507	6.80e-05	-0.000908	3.45e-05
Trust	(0.00936)	(0.00163)	(0.00959)	(0.00346)	(0.00359)	(0.000156)	(0.00424)	(0.000192)
University	-0.00149	-0.00114	-0.0332*	-0.00126	-0.0264	-0.00946	-0.00519	-0.0169
Hospital	(0.0171)	(0.00205)	(0.0190)	(0.00423)	(0.0207)	(0.0108)	(0.0273)	(0.0212)
Constant	0.142	-0.0484***	-0.269**	-0.0809***	-0.0138	-0.0283	-0.117**	-0.110***
	(0.121)	(0.00971)	(0.106)	(0.0185)	(0.0215)	(0.0442)	(0.0568)	(0.0350)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Instruments	22	47	22	47	30	42	25	41
N	986	1,047	986	1,047	986	1,047	986	1,047
Groups (hospitals)	132	121	132	121	132	121	132	121

* Significant at p $\leq$ 0.1

** Significant at p $\leq$ 0.05

*** Significant at p $\leq$ 0.01

4.2.1 AMI

Table 4 presents the results for the regressions run for Model 2, using the AMI latent outcome measures as dependent variables. The results suggest that an increase in lagged 30-day mortality is associated with higher 30-day mortality and lower 1-year readmissions. Aside from lagged mortality, the only other variables found to be significantly related to 30-day mortality were length of stay and lagged length of stay. Where lagged length of stay has a significant negative association, such that higher LOS and length of stay a significant positive association. In the model for long term AMI mortality, length of stay and lagged length of stay were also significant, but both had a positive association, such that higher length of stay is associated with higher 1-year mortality estimates. Lagged caseload is negatively associated with year long mortality, such that more cases are associated with higher 1-year mortality. In addition, teaching hospitals were significantly related to lower 1-year mortality than acute care trusts, but with 90% significance. The results

using short term readmissions as a dependent variable show no significant associations. While the results for the year-long readmissions model indicate that lagged readmissions, lagged length of stay and lagged mortality are all associated with the dependent variables. The direction of the results suggests that higher lagged readmissions will lead to lower readmissions, and higher lagged length of stay will lead to higher 1-year readmissions.

4.2.2 Hip Replacement

Table 4 shows the Model 2 results for the Hip Replacement latent outcome measures. The results indicate that 30-day mortality has a negative and significant relationship with past mortality. Current and lagged length of stay are also significantly associated with 30-day mortality, such that higher lagged length of stay is associated with lower 30-day mortality, while higher length of stay is associated with higher 30-day mortality. In addition caseload is significant, such that it has a positive association with 30-day mortality. The results for year-long mortality indicate similar results. Lagged year-long mortality is significantly associated with year-long mortality, such that an increase in lagged mortality results in lower mortality. Length of stay and lagged length of stay are also significantly associated with year-long mortality, indicating a positive and negative association respectively. The regression results for readmissions, indicate fewer significant explanatory variables. The only significant variable in the 28-day readmission model is length of stay where higher length of stay is associated with higher readmissions. There are no significant variables in when the year-long readmission outcome is used as the dependent variable.

5 Discussion

This paper focuses on the need to develop more sensitive indicators to measure the quality of health care providers. Performance measurement is increasingly being used by stakeholders such as managers, politicians, regulators, researchers, service users and the general public to inform their decision making – whether this be individual choices or broader health policy decisions. While increasingly process and structure measures are being used in quality measurement, outcome measures remain lucrative in their ability to present a simple, all-encompassing measure of health care efforts. Hitherto, stakeholders have relied on raw or risk-adjusted outcome measures as indicators of organizational quality. Yet, these types of indicators are largely determined by exogenous factors such as patient characteristics, and often risk-adjustment techniques are highly sensitive to technical choices

which can bias measures in different directions depending on the risk adjustment method selected.

The method applied in this paper employs a systematic approach for evaluating hospital quality using outcome indicators. This approach assumes a latent hospital level variable that is unmeasured but includes all of the unobserved factors influencing hospital quality. It is captured by vectors of hospital level intercepts estimated from individual patient level equations measuring determinants of outcomes for seven different conditions across hospitals. These intercepts, or 'latent outcome measures' will thus exclude known confounding patient level variables, but reflect factors such as unmeasured resources, environmental and organizational characteristics as well as random error and data imperfections. They can then be used to explain how much unobserved hospital effects contribute to changes in mortality and readmission in any given year. Closely observing the trend of the latent outcome measures over time allows for a better examination of the rate of change in hospital quality than the examination of raw outcome measures, that can be of practical use to individual providers or policy makers.

The average latent mortality estimates, plotted over time, show the rate of change in outcomes over time, controlling for patient characteristics. The AMI and Hip Replacement outcomes indicated that for most years, average mortality attributable to hospital quality is declining. This is also the case for the other conditions studied, as indicated in Appendix A. In most cases, the rate of decline appears to slow from 2005 onwards. This is especially pronounced for the year-long mortality measures in all conditions, and for 30-day mortality intercepts in AMI and Hip Replacement. The trends in readmissions vary more by condition. The average of the hospital intercepts suggest increasing readmissions for AMI and decreasing readmissions for Hip Replacement. While this paper has not used these measures to investigate the possible factors that could be responsible for these changes, it is likely that policy changes in this time associated with the introduction of Payment by Results and increased competition play some role, either in terms of the effects they have on quality through incentives or more likely on the effect they have on coding of mortality. Under this payment scheme, hospital revenue is very closely linked to coding. This may result in better coding, but also can lead to adverse behaviours such as miscoding or fraud. Discrepancies in coding practices have recently been reported in the literature, such as hospitals coding deaths as palliative care in order to reduce mortality rates (Hawkes, 2010a). Further in-depth investigation is necessary to draw any conclusive results. Given the concerns about mortality coding, future estimates using this methodology would benefit from considering palliative outcomes in addition to mortality

and readmission rates at different levels, or controlling for palliative care by including it in the explanatory variables of the patient level regressions.

Another interesting observation for many of the average latent estimate plots is that the difference between the coefficients from one year to the next, which indicate the absolute rate of change in mortality. Thus an increase of 0.02 in the average mortality intercept from one year to the next means that the mortality attributable to hospital quality increased by 2% over that time period. This information can allow us to make important conclusions about changes over time, controlling for patient factors. Finally, the confidence intervals often become narrower for the mortality intercepts, and either wider or narrower for the readmissions estimates. This suggests that the variation amongst relative hospital performance in any given year is also changing. Again this may be linked to policies occurring during this time period, and perhaps indicate a need to examine possible explanatory factors.

The latent outcome measures can also be examined at the individual hospital level. Investigating the outcome measures at this level of analysis provides more information on how hospitals are performing relative to each other. The results indicate that where there is sufficient sample size these estimates can be informative and quite precise in distinguishing quality trends over time, and relative to peers. However, in cases where there is a small number of patients treated annually, the estimates will be subject to wide year-to-year fluctuations in quality, and surrounded by wide confidence intervals. Both factors make it harder to draw conclusive results about the quality of provider, making it difficult to draw any conclusions with an adequate degree of certainty. This suggests that the best sample with which to conduct this type of analysis at the hospital level is for conditions where there are large number of patients, provided at many hospitals in the country.

The final part of the analysis of this paper examined the latent outcome measures to determine how much they can be explained by known hospital characteristics. One of the main areas of interest of this section was to determine how dynamic hospital outcomes are. The regression results produced mixed findings. Both AMI and Hip Replacement indicate that past 30-day mortality is an important determinant of current mortality, but only for Hip Replacement is year-long mortality also dynamic. In the readmission models, only long term AMI readmissions are influenced by lagged readmissions.

In general, there appear to be two ways in which past performance can influence current performance: through positive association or a negative association. The first way, a positive association, occurs when good/bad past performance is a predictor of continuing

good/bad performance. This can be best explained through a notion of path dependency, where hospitals performing a certain way in one period are likely to continue to do so regardless of the characteristics of the patients admitted. The second way, as indicated by a negative association, occurs where past good/bad past performance is associated with bad/good current performance. This can be regarded as a process of change. Where bad performance precedes good performance, we can conceptualise this as a process of improvement, where poor outcomes in one period motivate internal change so that outcomes will improve in the next period. Where good performance precedes bad performance, we can consider this change an indication of decline.

In the 30-day mortality model AMI lagged AMI and current AMI had a positive association, suggesting path dependency. For the 30-day Hip Replacement model the association between past and current mortality is negative, suggesting there has been change over time. The average trend in latent 30-day mortality indicated in Figure 8 suggest that mortality is improving over the time period studied, thus the negative association is indicative of an improvement. The results from the year-long mortality regression model, and Figure 8 also indicate a dynamic effect indicating an improvement in performance.

Unlike the mortality models, the short and long term readmission models were estimated using two lagged outcome measures to test for a dynamic effect; past readmissions and past mortality. The positive and negative associations for the lagged readmission variables can be interpreted in terms of path dependency and change as explained above. However, the sign on the past mortality variables indicates the association between mortality and readmissions. Common sense would suggest that high readmissions are an indication of poor quality, as are high mortality rates. However, this may not always be the case. High readmissions may also be indicative of good quality where severe patients have been saved, or even an indication of other factors such as patient lifestyle and behaviour after discharge.

Only in the AMI model for year-long readmissions were both lagged outcome measures significant for the same condition. In this case, the negative sign on the lagged readmission measure suggests change over time. The average latent trend for this AMI outcome, plotted in Figure 2 indicates increasing readmissions in the beginning of the sample, followed by decreasing readmissions in the final year. Thus the negative sign on the lagged measure is indicating the improvement in readmissions. The negative sign on the lagged mortality variable indicates that 365-day mortality and 365-day readmissions are negatively correlated, meaning that higher year-long readmissions may not be indicative of worse performance for this condition. The results for the Hip Replacement model show

no association between lagged mortality and readmissions, or lagged readmissions and current readmissions.

The results from Model 2 do not suggest any conclusive findings on how hospital type influences performance. There are very few specialist trusts in the sample[1]. However, specialist trusts were not associated with quality in any of the AMI or Hip Replacement Models. Indeed, the only type of hospital that was associated with better mortality were teaching hospitals, but only for year-long AMI outcomes and only with 10% significance. Of the control variables length of stay and lagged length of stay were significantly associated with all mortality models, but only a few of the readmissions models. Where significant in the readmissions models both indicators were always indicative of higher readmissions. Higher lagged length of stay is associated with worse AMI mortality but improved hip mortality. While higher same period length of stay is associated with better 30-day mortality for AMI and worse 30-day mortality for Hip, and worse year-long mortality in both conditions.

Caseload also has mixed effect in the different models, this could be because good hospitals get more cases or because too many cases may lead to lower quality. Higher lagged caseload is associated with higher Hip Replacement 30-day mortality. However, but same period caseload is positively associated with AMI year-long mortality. Interestingly caseload does not appear to have a significant effect on readmissions for either condition.

Overall we find that it is difficult to attribute changes in performance to hospital characteristics and exogenous factors using only this analysis. However we are able to identify areas where past performance is an important predictor of future performance, as well as areas where the association between different outcome measures may be informative as to their correct interpretation. Moreover, our results suggest that we can distinguish between two different types of dynamic effect, path dependency and change, and identify the conditions where one occurs instead of the other. Indeed, in the areas where we see change and not path dependency there is room for further analysis to discover what factors are motivating this change, and whether it constitutes an improvement or decline in performance.

We conclude that while the latent variable technique used in this paper does not offer the solution to quality measurement it is an important addition to the tool box of methods with which to better understand the hospital contributions to quality of care. The latent outcome measures provide an interesting way to examine hospital performance over time

[1] There were four specialist trusts in the sample and of those two are orthopaedic.

and relative to one another, provided that they are used for conditions with large sample size. While the results in this paper only report the results for two of the seven conditions studied, the findings are consistent across conditions. Thus we support that latent outcomes generated through the latent variable approach provide useful indicators to use in further analysis of hospital performance. These indicators can be useful in correcting for methodological bias that arises from using statistical models that combine aggregate variables (such as hospital mortality rates) with individual observations to determine relationships. In conclusion, while these types of variables may only be only risk adjusted outcome measures, they provide a straightforward way to present, observe and analyze outcome data as well as to provide directions for further research.

A Results for Other Conditions

Tab. 5: Summary Statistics of the Sample.

Condition	ICD-10/OPCS4.3 codes	Years Analyzed	Mean cases per year	Number of hospitals
MI	ICD-10: 122,123	2000-2008	7,641	150
IHD	ICD-10: 120,125	2000-2008	142,638	119
CCF	ICD-10: I11.0, I13.0, I25.5, I50.0, I50.1, I50.9, J81X	2000-2008	3,717	122
Stroke	ICD-10: I60 - I67	2000-2008	66,866	167
TIA	ICD-10: G45.0-G45.4, G45.8-G45.9, G46.0-G46.8	2000-2008	12,433	139

Fig. 13: Average raw and latent 30-day mortality 2000-2008

Tab. 6: Model 2 regression results for 30-day Mortality.

	MI	IHD	CCF	STROKE	TIA
	$D30_{ht}$	$D30_{ht}$	$D30_{ht}$	$D30_{ht}$	$D30_{ht}$
L. Latent	-0.0711	0.488**	-0.0265	0.342**	0.0166
Mortality	(0.176)	(0.247)	(0.121)	(0.153)	(0.0939)
L. LOS	0.00380	-1.92e-05	-0.000525	0.000656	-1.32e-06
	(0.00543)	(0.000538)	(0.00233)	(0.000768)	(0.000744)
L. Cases	7.44e-05	3.95e-08	0.000558	-2.08e-05	-5.77e-05
	(0.000195)	(1.19e-06)	(0.000447)	(5.98e-05)	(7.07e-05)
L. Waiting	-	-	-	-	6.51e-07**
Times	-	-	-	-	(2.85e-07)
Cases	0.000404	-4.86e-07	-0.000966	3.94e-05	2.36e-05
	(0.00117)	(3.32e-06)	(0.00120)	(0.000283)	(8.79e-05)
$Cases^2$	-9.94e-07	1.84e-10	8.63e-06	-9.43e-09	1.08e-07
	(4.38e-06)	(9.40e-10)	(8.83e-06)	(1.96e-07)	(1.36e-07)
LOS	-0.00525	0.000815	0.00181	-	7.47e-05
	(0.0102)	(0.00122)	(0.00307)	-	(0.00155)
Waiting	6.53e-06***	-1.91e-06	6.80e-05	3.01e-06	4.81e-06
Times	(1.82e-06)	(2.44e-06)	(0.000148)	(1.24e-05)	(5.68e-06)
Specialist	-0.0373	0.000995	-0.0603***	0.0384	0.00990***
Trust	(0.105)	(0.00433)	(0.0203)	(0.0443)	(0.00368)
Foundation	0.00210	-0.000576	0.0167	0.00447	-0.000772
Trust	(0.0154)	(0.000766)	(0.0163)	(0.00474)	(0.00227)
University	0.0141	-0.000822	-0.0471***	-0.0133	0.000767
Hospital	(0.0187)	(0.00257)	(0.0173)	(0.00914)	(0.00143)
Constant	-0.331**	-0.0208	-0.237***	-0.179*	-0.0324***
	(0.134)	(0.0135)	(0.0617)	(0.0930)	(0.00861)
Year Dummies	Yes	Yes	Yes	Yes	Yes
Instruments	28	33	33	30	34
N	341	919	352	830	509
Groups (hospitals)	105	129	101	125	104

* Significant at p $\leq$ 0.1

** Significant at p $\leq$ 0.05

*** Significant at p $\leq$ 0.01

References

Arellano, M. and S. Bond (1991, April). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies 58*(2), 277 –297.

Arellano, M. and O. Bover (1995, July). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*(1), 29–51.

Audit Commission (2004, July). Introducing payment by results -.

Aylin, P., A. Bottle, and A. Majeed (2007, May). Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *BMJ 334*(7602), 1044.

Birkmeyer, J. D., A. E. Siewers, E. V. A. Finlayson, T. A. Stukel, F. L. Lucas, I. Batista, H. G. Welch, and D. E. Wennberg (2002). Hospital Volume and Surgical Mortality in the United States. *New England Journal of Medicine 346*(15), 1128–1137.

Blundell, R. and S. Bond (1998, August). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115–143.

Charlson, M. E., P. Pompei, K. L. Ales, and C. R. MacKenzie (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases 40*(5), 373–383. PMID: 3558716.

Dimick, J. B., H. G. Welch, and J. D. Birkmeyer (2004). Surgical Mortality as an Indicator of Hospital Quality. *JAMA: The Journal of the American Medical Association 292*(7), 847 –851.

Farrar, S., D. Yi, M. Sutton, M. Chalkley, J. Sussex, and A. Scott (2009). Has payment by results affected the way that english hospitals provide care? difference-in-differences analysis. *339*. PMID: 19713233 PMCID: 2733950.

Gujarati, D. N. (2003). *Basic econometrics*. McGraw Hill.

Hambleton, R. K. and L. L. Cook (1977, July). Latent Trait Models and Their Use in the Analysis of Educational Test Data. *Journal of Educational Measurement 14*(2), 75–96. ArticleType: research-article / Issue Title: Applications of Latent Trait Models / Full publication date: Summer, 1977 / Copyright © 1977 National Council on Measurement in Education.

Hawkes, N. (2010a, April). Patient coding and the ratings game. *BMJ 340*(apr23 2), c2153–c2153.

Hawkes, N. (2010b, April). The coding maze: mortality ratios and real life | Straight Statistics. http://www.straightstatistics.org/article/coding-maze-mortality-ratios-and-real-life.

Holland, W. W. (1988). *European Community atlas of "Avoidable Death"*. Oxford University Press.

Iezzoni, L. I. (1994, December). Using risk-adjusted outcomes to assess clinical practice: An overview of issues pertaining to risk adjustment. *The Annals of Thoracic Surgery 58*(6), 1822–1826.

Iezzoni, L. I. (1997, October). Assessing Quality Using Administrative Data. *Annals of Internal Medicine 127*(2), 666 –674.

Iezzoni, L. I., A. S. Ash, M. Shwartz, J. Daley, J. S. Hughes, and Y. D. Mackiernan (1996, October). Judging hospitals by severity-adjusted mortality rates: the influence of the severity-adjustment method. *Am J Public Health 86*(10), 1379–1387.

Jarman, B., A. Bottle, P. Aylin, and M. Browne (2005, February). Monitoring changes in hospital standardised mortality ratios. *BMJ 330*(7487), 329.

Jarman, B., S. Gault, B. Alves, A. Hider, S. Dolan, A. Cook, B. Hurwitz, and L. I. Iezzoni (1999, June). Explaining differences in English hospital death rates using routinely collected data. *BMJ 318*(7197), 1515 –1520.

Landrum, M. B., S. E. Bronskill, and S. T. Normand (2000). Analytic Methods for Constructing Cross-Sectional Profiles of Health Care Providers. *Health Services and Outcomes Research Methodology 1*(1), 23–47.

Lilford, R., M. A. Mohammed, D. Spiegelhalter, and R. Thomson (2004, April). Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *The Lancet 363*(9415), 1147–1154.

Maynard, A. (2008). Payment for Performance (P4P): International experience and a cautionary proposal for Estonia.

McClellan, M. and D. Staiger (1999, August). The Quality of Health Care Providers. *National Bureau of Economic Research Working Paper Series No. 7327*. published as McClellan, Mark and Douglas Staiger. "Comparing The Quality Of Health Care

Providers", Forum for Health Economics and Policy, 2000, v3, Article 6. Mark McClellan & Douglas Staiger, 2000. "Comparing the Quality of Health Care Providers," NBER Chapters, in: Frontiers in Health Policy Research, Volume 3, pages 113-136 National Bureau of Economic Research, Inc.

McKee, M., J. Coles, and P. James (1999, December). 'Failure to rescue' as a measure of quality of hospital care: the limitations of secondary diagnosis coding in english hospital data. *Journal of Public Health 21*(4), 453 –458.

McKee, M. and P. James (1997, October). Using Routine Data to Evaluate Quality of Care in British Hospitals. *Medical Care 35*(10), OS102–OS111. ArticleType: research-article / Issue Title: Supplement: Hospital Restructuring in North America and Europe: Patient Outcomes and Workforce Implications / Full publication date: Oct., 1997 / Copyright © 1997 Lippincott Williams & Wilkins.

Muthén, B. O. (1989, September). Latent variable modeling in heterogeneous populations. *Psychometrika 54*(4), 557–585.

Nickell, S. (1981, November). Biases in Dynamic Models with Fixed Effects. *Econometrica 49*(6), 1417–1426. ArticleType: research-article / Full publication date: Nov., 1981 / Copyright © 1981 The Econometric Society.

Nolte, E., M. McKee, N. T. for Research, and P. S. in Health Services (2004). *Does health care save lives?: avoidable mortality revisited*. Nuffield Trust.

Roodman, D. (2006, December). How to Do xtabond2: An Introduction to "Difference" and "System" GMM in Stata. Working Paper 103, Center for Global Development.

Shahian, D. M., S. Normand, D. F. Torchiana, S. M. Lewis, J. O. Pastore, R. E. Kuntz, and P. I. Dreyer (2001, December). Cardiac surgery report cards: comprehensive review and statistical critique. *The Annals of Thoracic Surgery 72*(6), 2155–2168.

Shahian, D. M., R. E. Wolf, L. I. Iezzoni, L. Kirle, and S. T. Normand (2010, December). Variability in the Measurement of Hospital-wide Mortality Rates. *New England Journal of Medicine 363*(26), 2530–2539.

Spiegelhalter, D. J., P. Aylin, N. G. Best, S. J. W. Evans, and G. D. Murray (2002, June). Commissioned analysis of surgical performance using routine data: lessons from the Bristol inquiry. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 165*(2), 191–221.

Tovey, D. (2007). *Clinical evidence handbook: the international source of the best available evidence for effective health care*. BMJ Publishing Group.

Treier, S. and S. Jackman (2008, January). Democracy as a Latent Variable. *American Journal of Political Science 52*(1), 201–217.

Windmeijer, F. (2005, May). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics 126*(1), 25–51.

For further information on this or any of the Health publications contact:

Naho Ollason
Managing Editor
LSE Health
The London School of Economics and Political Science
Houghton Street
London WC2A 2AE

Tel: + 44 (0)20 7955 3733
Fax: + 44 (0)20 7955 6090
Email: n.ollason@lse.ac.uk
Website: www.lse.ac.uk/collections/LSEHealth/