HEALTH ECONOMICS
Health Econ. **22**: 1158–1174 (2013)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/hec.2972

PATIENT SAFETY IN HOSPITALS – A BAYESIAN ANALYSIS OF UNOBSERVABLE HOSPITAL AND SPECIALTY LEVEL RISK FACTORS†

XIAOHUI ZHANG[a], KATHARINA HAUCK[b] and XUEYAN ZHAO[c,*]

[a]*School of Management and Governance, Murdoch University, Perth, Australia*
[b]*Imperial College Business School, Imperial College London, South Kensington, London, UK*
[c]*Department of Econometrics and Business Statistics, Monash University, Melbourne, Australia*

ABSTRACT

This paper demonstrates how Bayesian hierarchical modelling can be used to evaluate the performance of hospitals. We estimate a three-level random intercept probit model to attribute unexplained variation in hospital-acquired complications to hospital effects, hospital-specialty effects and remaining random variations, controlling for observable patient complexities. The combined information provided by the posterior means and densities for latent hospital and specialty effects can be used to assess the need and scope for improvements in patient safety at different organizational levels. Posterior densities are not conventionally presented in performance assessment but provides valuable additional information to policy makers on what poorly performing hospitals and specialties may be prioritized for policy action. We use surgical patient administrative data for 2005/2006 for 16 specialties in 35 public hospitals in Victoria, Australia. We use posterior means for latent hospital and specialty effects to compare hospital performance in patient safety. Posterior densities and variances are also compared for different specialties to identify clinical areas with greatest scope for improvement. We also show that the same hospital may rank markedly differently for different specialties.

Received 29 May 2012; Revised 13 June 2013; Accepted 18 June 2013

KEY WORDS: patient safety; hospital acquired complication; random coefficient probit model; Bayesian estimation

1. INTRODUCTION

Patient safety is universally acknowledged as a priority concern in health care (de Vries *et al*., 2008; Vincent, 2010). In the USA, the famous 'To Err is Human' report popularized the statistic that 44 000–98 000 Americans die each year as a result of medical errors, which equals roughly to 'a jumbo jet per day' (Kohn *et al*., 2000). The most common of complications acquired during a hospital stay, infections, is estimated to lead to 100 000 deaths per year in the USA (Klevens *et al*., 2007; Stone *et al*., 2010). This makes hospital-acquired infections the sixth leading cause of death in the USA. Apart from preventable mortality and morbidity, the overall direct costs of hospital-acquired infections alone are estimated as between US $28bn and US$45bn (Stone, 2009; Kohn *et al*., 2000). Activities to monitor and improve safety in hospitals have been placed high on the political agenda over the past decade. Some countries have set up dedicated organizations to advance patient safety (ACSQHC, 2011; AHRQ, 2011; NPSA, 2011), and international organizations such as the World Health Organization (WHO, 2011) and the Organisation for Economic Co-operation and Development (Droesler, 2008) also support initiatives on patient safety. Obligatory safety procedures such

*Correspondence to: Department of Econometrics and Business Statistics, Monash University, Melbourne, Australia.
E-mail: xueyan.zhao@monash.edu
†Supporting information may be found in the online version of this article.

as the WHO surgical checklist (WHO, 2008) and the WHO Guidelines on Hand Hygiene in Health Care (WHO, 2009) are now common in most hospitals around the world to reduce in-hospital complications.

Over the last two decades, considerable research effort has been invested into analysing patient safety incidences and understanding why they occur and how they could be prevented (de Vries *et al*., 2008). It has been recognized that humans inevitably make errors and that it is more effective to direct efforts towards creating hospital environments that make errors less likely to occur or at least limit their negative consequences (Pronovost *et al*., 2006; Rivard *et al*., 2006; Cebul *et al*., 2008). This places great responsibility onto hospital management in creating such an environment (Pronovost *et al*., 2006).

However, although experts seem to agree that executive hospital management has a crucial role in any effort to increase patient safety and improve hospital performance, there is little solid quantitative evidence on the extent to which hospital level factors affect patient safety. Where such evidence is available, it is usually narrowly focused on one intervention targeted at one treatment, patient group or area of clinical practice. This provides no real guidance to decision makers on how to prioritize efforts to improve patient safety across clinical areas or across different departments within one hospital. For example, there is very little evidence whether policies should be targeted at hospital department or top management level and whether this differs across clinical areas. Better evidence on how to prioritize patient safety initiatives is urgently required because many are costly and require significant additional staff time and resources (Graves, 2004). A major challenge in any such investigation is that hospital level risk factors, such as poorly implemented safety procedures, bad communication between medical teams, substandard hygiene and ineffective hospital management, are often unobserved.

The objectives of this paper are to examine in what clinical areas and at what organizational levels there is greatest scope for potential improvements in the quality of hospital care and how individual hospitals are placed for such potentials. We examine the patient, specialty and hospital level risk factors associated with the occurrence of hospital-acquired complications by estimating a Bayesian three-level random intercept probit model using administrative hospital data from the state of Victoria, Australia, for the year 2005/2006. The model exploits the multilevel nature of the data where each observation is associated with a patient within a specialty within a hospital. We decompose the unexplained error term into hospital effect, specialty within hospital effect and remaining random variations. The multilevel random effect specification allows for the unobservable effects for all episodes within the same hospital and the same specialty to be correlated, as these unobservable effects could be affected by factors that are common to all episodes in the same hospital and same specialty, such as clinical equipment and infrastructure, expertise and operation of medical teams, and specific routines and procedures. After controlling for observable patient profiles, we can interpret the unobservable hospital and specialty level effects as an indication of hospital managers and medical staff's potential in improving patient safety at those organizational levels, following the common interpretation of such effects in the literature on organizational performance (Smith *et al*., 2009).

Another focus of the paper is to exploit the rich information contained in the multilevel error decomposition, using a Bayesian hierarchical model framework, to evaluate individual hospital performance and to determine the specialties that have the largest scope for reducing complications via benchmarking across hospitals. The Bayesian algorithm, using Gibbs sampling with data augmentation, avoids the evaluation of multidimensional integrals in a likelihood function and is computationally more expedient than maximum likelihood estimation methods (Geweke *et al*., 1997). Our hierarchical modelling approach (Zeger and Karim, 1991; Daniels and Gatsonis, 1999) quantifies unobserved propensity for complications based on the rich information contained in the large number of episodes observed for 16 major specialties and 35 hospitals in this research. As remarked in McCulloch and Rossi (1994), ‘the Bayesian approach combines the smoothing advantages of the frequentist random effects model with the richness of the fixed effects approach’. In fact, the difference between fixed effect and random effect model is trivial under a Bayesian framework and can be defined through the prior distribution for the error components (Koop *et al*., 1997). The advantage of using a Bayesian hierarchical approach in this research is that hospital and hospital-specialty effects can be easily presented by the whole posterior density distributions of the error components using draws from the Gibbs sampler for making inference for individual hospitals and specialties. Posterior means of the unobservable risks can be used to compare performance across hospitals in general and across hospitals for each speciality. Additionally, the variance of unobservable propensity for complications for a particular specialty can be used to indicate potentials for reducing complications for one hospital against another. By benchmarking hospitals against their peers, it is assumed that greater variations in complications imply greater scope for hospital managers and medical teams to improve their performance. These results allow policy makers and healthcare managers to judge in which specialities there is greatest opportunity for improvements in patient safety and target policy efforts towards the hospitals that are underperforming in those specialties. Interventions can be prioritized towards clinical areas where they can make greatest difference to patient safety and hospital quality.

The literature on performance measurement and benchmarking commonly uses models with error decomposition after observable factors are controlled. The notion is that once the effects of observable factors are controlled, the remaining unexplained effects in the error associated with individual firms or organizations can be used as indicators for performance. A popular approach is stochastic frontier analysis (see reviews in Kumbhakar and Lovell 2000 and Coelli *et al*., 2005 and health economic applications in Hollingsworth, 2003) and, more recently, multilevel modelling when multilevel data are available. In the context of organizational performance measurement, it explicitly models the hierarchical nature of the organizational structure by decomposing the unexplained error term into symmetric random error components attributable to different levels of the hierarchy. The estimated error components are interpreted as measures of performance associated with different levels, and errors that deviate markedly from zero are interpreted as worse or better performance than would be expected given observable exogenous factors that affect the outcome (Jacobs *et al*., 2006). Multilevel models have been championed for some time as valuable tools for performance assessment in health care (Rice and Leyland, 1996; Rice and Jones, 1997; Duncan *et al*., 1998) and have been used to analyse mortality in heart failure patients (Merlo *et al*., 2001), heart attack patients (Gupta *et al*., 2003; Schreyoegg *et al*., 2011), patients in intensive care (Moreno *et al*., 2005) or women giving birth (Karlsen *et al*., 2011), and to analyse disease-specific outcome measures for stroke patients (Hinchey *et al*., 2008; Reeves *et al*., 2010), heart attack patients (Gupta *et al*., 2003), trauma patients (Huseynova *et al*., 2009) or abdominal aortic surgery (Pronovost *et al*., 2001). Most studies to date have focused on particular patient groups and disease-specific outcomes. A few papers focus more broadly on all hospital patients, but they pool them for analysis and do not compare across patient groups or clinical areas (Cho *et al*., 2003; Glance *et al*., 2008; Hauck and Zhao, 2011; Hauck *et al*., 2012).

The paper is organized as follows. In the next section, we describe the three-level random intercept probit model and other measures of interest. Data used in this study are described in Section 3. Section 4 presents and discusses the results, and Section 5 concludes.

2. THE ECONOMETRIC MODEL

We specify a three-level random intercept probit model to relate a binary dependent variable for hospital-acquired complication to observable patient risk factors and random error components of hospital effects, hospital-specialty effects and remaining random variations. Let y^*_{ish} be a latent variable that is proportional to the propensity of suffering one or more complications for the i-th episode in the s-th specialty and the h-th hospital and be given by the following equation:

$$\begin{aligned} y^*_{ish} &= x_{ish}\beta + e_h + u_{sh} + \varepsilon_{ish}, \\ e_h &\sim N\left(0, \sigma^2_H\right), u_{sh} \sim N\left(0, \sigma^2_S\right), \varepsilon_{ish} \sim N(0, 1), \end{aligned} \tag{1}$$

with $y^*_{ish} > 0$ mapped to an observable binary variable $y_{ish} = 1$ if the patient suffered at least one complication during the episode and $y^*_{ish} \leq 0$ to $y_{ish} = 0$ otherwise, $i = 1, 2, \ldots, \sum_{h=1}^{H}\sum_{s=1}^{S_h} N_{sh}$, $s = 1, 2, \ldots, S_h$ and $h = 1, 2, \ldots, H$. Here, N_{sh} represents the number of episodes in the s-th specialty and the h-th hospital. Let $N = \sum_{h=1}^{H}\sum_{s=1}^{S_h} N_{sh}$ represent the number of all episodes in the dataset. The term x_{ish} is a $1 \times (k+1)$ vector of covariates representing observable patient risk factors, with the first element being unity, and β is a $(k+1) \times 1$ vector of coefficients representing effects of observable risk factors, whose first element represents the fixed part of the intercept applying to all episodes. The error terms e_h represent unexplained variation across hospitals and are assumed to be independently and identically distributed (i.i.d.) normal for all h. The error u_{sh} is the hospital-specialty level effect that is i.i.d. normal for all h and s, and ε_{ish} is the error term that embodies remaining unobserved individual random effect, which is assumed to i.i.d. follow a standard normal distribution for identification.[1] All β's and σ's are coefficients to be estimated.

We extend the Bayesian algorithm for a single error probit model presented in Koop (2003, p. 215) and devise a Gibbs sampler with data augmentation for our three-level error component probit model. The unobservable latent variable y^*_{ish} and the error terms e_h and u_{sh} are treated as unknown parameters with values being drawn from their conditional posterior densities. Then, conditional on y^*_{ish}, e_h and u_{sh}, the three-level random intercept probit model reduces to the standard linear regression model, facilitating draws from the conditional posterior densities for the parameters. Technical details for the Gibbs sampling algorithm and Bayesian estimation, including the assumed prior distributions, the conditional posterior densities for all unknown parameters and latent variables, and other simulation details are presented in Zhang *et al*. (forthcoming). The Gibbs sampler proceeds by iteratively sampling from the full conditional distributions of all unknown parameters, that is, β, σ^2_H and σ^2_S, and latent variables, that is, y^*_{ish}, e_h and u_{sh}. After discarding draws from the burn-in period, the parameter draws obtained from the conditional posterior densities are used to estimate posterior densities of the parameters and other measures of interest.

The first set of measures of interest relates to the partition of the total variance in the latent dependent variable into contributions of observable patient characteristics and components of unobservable hospital, hospital-specialty and remaining random variations. Proportion of total variation that is explained by observable patient characteristics can be estimated by an extension of the pseudo-R^2 measure proposed by McKelvey and Zavoina (1975) for binary dependent variable models, which relates to the ratio of regression variation and total variation in the latent dependent variable:

$$R^2_{MZ} = \frac{\frac{1}{N}\sum\limits_{i,s,h}\left(x_{ish}\hat{\beta} - \bar{x}\hat{\beta}\right)^2}{1 + \hat{\sigma}^2_S + \hat{\sigma}^2_H + \frac{1}{N}\sum\limits_{i,s,h}\left(x_{ish}\hat{\beta} - \bar{x}\hat{\beta}\right)^2}$$

We use variance partition coefficients (VPCs) to measure proportions of total unexplained variation attributable to hospital specific, specialty-hospital specific and remaining random errors. Following the literature on organizational performance, we interpret unexplained conditional variations associated with higher levels of an organization as potential indicators for managerial effort (Jacobs *et al*., 2006). This implies that variations in complication rates across hospitals, which remain after taking account of differences in observable factors and random errors, give an indication of the extent to which complications may be amenable to interventions by the hospitals. Controlling for the impact of patient risk factors, the proportion of unexplained variance in the propensity of suffering a complication due to hospital level variation is given by

$$VPC_H \big| \sigma^2_H, \sigma^2_S = \frac{\sigma^2_H}{1 + \sigma^2_S + \sigma^2_H}; \tag{2}$$

the proportion due to hospital and specialty level variation is given by

$$VPC_{HS} \big| \sigma^2_H, \sigma^2_S = \frac{\sigma^2_S + \sigma^2_H}{1 + \sigma^2_S + \sigma^2_H}; \tag{3}$$

[1]We assume normal distribution following convention. We also experimented with Uniform distribution for the errors following a referee's comment and found that the ranking of hospitals in our empirical results almost unchanged with rank correlation of 0.999.

and the proportion of unexplained variance due to specialty level variation within a given hospital is measured by

$$VPC_{S|H}\left|\sigma_S^2 = \frac{\sigma_S^2}{1+\sigma_S^2}\right. . \tag{4}$$

Posterior densities for these VPC measures can be estimated using draws of parameters σ_H^2 and σ_S^2 from the Markov chain Monte Carlo (MCMC) simulation.

Another objective of the paper is to measure and benchmark performance of individual hospitals and specialties, taking full advantage of the rich information in our Bayesian simulation of the multilevel error components that may not be available when using commercial software. In particular, posterior densities for the unobservable error components for individual hospitals and individual specialties within hospitals can be obtained from the MCMC draws. Comparison of the posterior means and variance of these distributions provides valuable insights. Various measures are presented and discussed in Section 4 that evaluate performance of individual hospitals in general and by specialties. Specialties are also compared in terms of the potential in reducing complications.

3. DATA

In this analysis, we use a routine administrative hospital dataset, the Victorian Admitted Episodes Dataset (VAED). It contains detailed information on all patient episodes in all public hospitals in the state of Victoria, Australia. The data are naturally clustered into three levels: patients/episodes within specialties/departments within hospitals. The VAED contains detailed information on patients' diagnoses, treatments and other episode characteristics, and we supplement these data with secondary data on patients and hospitals.

Our sample consists of 67 129 inpatient elective surgical episodes in 35 public hospitals and 37 232 inpatient emergency surgical episodes from 34 hospitals in the year 2005/2006. We exclude small rural health centres that performed less than 200 surgical procedures over the year, as these small centres have very different activities to the larger regional, city and teaching hospitals, and we further exclude medical episodes, including dialysis, radiology, chemotherapy and dental episodes, and patients younger than 18 years. We estimate the random intercept probit model separately for elective and emergency surgical patients. Definitions of all variables and summary statistics are given in Table I.[2]

The dependent variable in this analysis indicates whether the patient experienced one or more complications during admission. We code 'complications' as a binary variable because different complications for the same patient during one episode may not be independent events to be modelled as counts.

Victorian coding standards record whether a condition was present on admission (Jackson *et al*., 2006); this allows to unambiguously attribute complications to the treating hospitals, a necessary condition if they are to be used as measures of performance. The Victorian coding of complications has been validated by a number of studies (Ehsani *et al*., 2006; Jackson *et al*., 2006; Moje *et al*., 2006; Ehsani *et al*., 2007; McNair *et al*., 2009; Michel *et al*., 2009). Of the 67 129 elective inpatients, 20.79% suffered at least one complication, and 18.52% of the emergency inpatients experienced one or more complications during their admission.

The independent variables are risk factors pertaining to the patient and the episode of care. The dummies 'wies1' to 'wies4' represent the originally continuous relative cost weight (weighted inlier equivalent separation) on the basis of which hospitals receive reimbursement for treatment of individual patients. The cost weight is based on the patients' diagnosis-related groups (DRG), and higher values characterize more complex

[2]More detailed summary statistics by type of hospital are available in the working paper version of this paper, Zhang *et al*. (forthcoming).

DOI: 10.1002/hec

Table I. Variable definitions and summary statistics

Variable	Definition	Elective Mean	Elective SD	Emergency Mean	Emergency SD
y	1 if suffer at least one complication, 0 otherwise	0.2079	0.41	0.1852	0.39
Episode characteristics					
wies1	1 if the weighted inlier equivalent separation is less than 1, 0 otherwise (referenced group)	0.3144	0.46	0.1981	0.40
wies2	1 if the weighted inlier equivalent separation is between 1 and 4, 0 otherwise	0.5210	0.50	0.5386	0.50
wies3	1 if the weighted inlier equivalent separation is between 4 and 7, 0 otherwise	0.1224	0.33	0.1565	0.36
wies4	1 if the weighted inlier equivalent separation is greater than 7, 0 otherwise	0.0422	0.20	0.1069	0.31
weekenda	1 if admitted on a Saturday, Sunday or public holiday, 0 otherwise	0.0369	0.19	0.2863	0.45
numberdiag	Number of diagnosis	3.2713	1.49	4.1117	1.34
numberop	Number of treatments and interventions	3.4830	1.24	3.9417	1.25
transep	1 if transferred to another hospital or hospital department, 0 otherwise	0.0562	0.23	0.1879	0.39
Patient characteristics					
age	Age of patient divided by 10	5.5858	1.80	5.5733	2.12
age2	Square of patients' age divided by 10 000	0.3445	0.20	0.3555	0.24
female	1 if female, 0 if male	0.5267	0.50	0.4573	0.50
private	1 if paid privately for the episode, 0 otherwise	0.0962	0.29	0.0980	0.30
seifa	Index of social dis/advantage, based on postcode of patient	0.9875	0.07	0.9986	0.08
Charlson co-morbidites					
ami	1 if has acute myocardial infarction, 0 otherwise	0.0128	0.11	0.0909	0.29
chf	1 if has congestive heart failure, 0 otherwise	0.0127	0.11	0.0475	0.21
pvd	1 if peripheral vascular disease, 0 otherwise	0.0323	0.18	0.0401	0.20
cevd	1 if has a cerebrovascular event, 0 otherwise	0.0115	0.11	0.0290	0.17
dementia	1 if has dementia, 0 otherwise	0.0032	0.06	0.0271	0.16
copd	1 if has chronic obstructive pulmonary disease, 0 otherwise	0.0097	0.10	0.0223	0.15
pud	1 if has peptic ulcer, 0 otherwise	0.0011	0.03	0.0089	0.09
mild_ld	1 if has mile liver disease, 0 otherwise	0.0069	0.08	0.0096	0.10
diab	1 if has diabetes, 0 otherwise	0.0347	0.18	0.0488	0.22
diab_com	1 if has diabetes and complications, 0 otherwise	0.1007	0.30	0.1147	0.32
hp_papl	1 if has hemiplegia or paraplegia, 0 otherwise	0.0029	0.05	0.0149	0.12
renal_di	1 if has renal disease, 0 otherwise	0.0313	0.17	0.0562	0.23
cancer	1 if has cancer, 0 otherwise	0.1349	0.34	0.0642	0.25
aids	1 if has AIDS, 0 otherwise	0.0003	0.02	0.0004	0.02

Table II. Estimated variances and variance partitions

	Elective inpatients	Emergency inpatients
σ_H	0.338(0.048)	0.311(0.050)
σ_S	0.406(0.024)	0.563(0.032)
Variation partition (%): observable factors and unobservable factors		
R^2_{MZ}: % explained by $X\beta$	38.13	26.12
% explained by errors	61.87	73.88
VPCs [mean (SD), %]		
VPC_H	9.0(2.3)	7.0(2.1)
VPC_{HS}	25.6(3.0)	38.7(4.1)
$VPC_{S\backslash H}$	14.2(1.4)	24.0(2.1)

R^2_{MZ} is the proportion of the total variation in the propensity of complication that can be explained by observable patient characteristics, with the remaining variation attributed to unobservable factors (% explained by errors). VPC_H represents the proportion of total unexplained variation that is attributable to hospital specific factors. VPC_{HS} represents the proportion of total unexplained variation that is attributable to hospital and speciality specific factors together. $VPC_{S\backslash H}$ represents the proportion of total unexplained variation that is attributable to speciality specific factors.

patients and consequently attract higher reimbursement. Additional measures of patient complexity are the number of co-morbidities and operations and medical interventions administered during the episode. We further control for whether admission was on a weekday or weekend, and for destination of the patient at discharge. Other patient-level indicators of medical need in our models are age, gender, the 2001 ‘Socio-Economic Indexes for Areas’ (Australian Bureau of Statistics, 2001) and ‘private’ indicating whether a patient paid privately for the stay in hospital. We further adjust for patients' medical complexity by including separately all comorbidities comprised in the Charlson index (Charlson *et al*., 1987; Sundararajan *et al*., 2004).[3] To guarantee anonymity of the hospitals in our study, we do not disclose their names, but we classify them into four types according to their geographical location, teaching status and whether they are specialized or general hospitals. This allows comparing hospitals with their respective peers.

4. RESULTS AND DISCUSSION

We discard draws of 5000 iterations in a burn-in period in the MCMC simulation and use the 10 000 iterations after the burn-in to estimate the posterior densities. We use the simulation inefficiency factor (SIF; Kim *et al*., 1998) to assess mixing performance, and we find that SIF values are small for all parameters, which strongly suggest convergence of the sampler. Marginal effects that inform on the patient-level factors' impact on the likelihood of experiencing complications and complexity-adjusted complication rates for hospitals are estimated but are not presented here due to space restrictions.[4] In the following sections, we will discuss results on the effects of unobservable risk factors, the main focus of our paper.

4.1. Variance partition coefficients

Table II presents the estimated standard deviations of hospital and specialty level error terms, percentage contribution of observable patient profiles to total variation in the latent dependent variable (pseudo-R^2_{MZ}) and three VPCs. Relative to the standard deviation of 1 for the random error term ε_{ish}, the estimates of 0.338 and 0.406 for σ_H and σ_S, respectively, are significantly above 0. The results for R^2_{MZ} show that 38% and 26% of the total variation in the propensity of complication can be explained by observable patient characteristics for elective

[3]Because of high degree of collinearity, we combine liver disease and severe liver disease, and cancer and metastatic cancer into two joint categories.

[4]These results are presented in Zhang, *et al*. (forthcoming).

DOI: 10.1002/hec

and emergency inpatients, respectively. The fact that the estimates for the two error components' standard deviations are statistically significant also indicates that the multilevel error component model is preferred over the single error probit model.

Of the remaining 62% and 74% unexplained variances, the VPCs represent proportions of this total unexplained variation attributable to hospital-specific, hospital-specialty-specific and random effects. The VPC_h shows that of the variation in complications across elective episodes that cannot be explained by observable risk factors, 9% (for elective patients) and 7% (for emergency patients) are due to variation at the hospital level. These values can be interpreted as the potential impact of differences in hospital level management on performance. We also calculate the VPC_{hs} for hospital and hospital-specialty variation together. If the VPC_h is as large as the VPC_{hs}, then all variation would be attributable to the higher managerial level and none to the lower department or speciality level. We find that this is not the case at all. The VPC_{hs} is 26% for elective patients and even higher at nearly 39% for emergency patients. This implies that management efforts in hospitals on both higher executive and lower specialty level together account for about one quarter of unexplained variation in elective and nearly 40% in emergency complications, whereas the remainder is due to random variation at patient level. Although the VPC_h's are significantly different from zero, they are still markedly smaller than the VPC_{hs}. This result implies that interventions that are only targeted at higher managerial levels are unlikely to be very effective in reducing complications; they need to be combined with interventions targeted at department level. For example, the US Centers for Medicare & Medicaid Services (CMS) have excluded eight complications that are deemed preventable from reimbursement (CMS, 2007; CMS, 2011). Managers at higher executive level are probably more concerned with the financial viability of their hospital than individual doctors. Thus, our results suggest that practical interventions as a result of changes in reimbursement policy, such as implemented by the CMS, are more likely to be effective if targeted at departmental level.

Within each given hospital, $VPC_{s|h}$ shows that 14% (elective) and 24% (emergency) of unexplained variation in complication probabilities can be attributed to specialty level effects. Specialty level variation seems to play a more important role in explaining variation in the risk of a complication for emergency patients than for elective patients. A possible explanation might be that medical teams in emergency departments on a lower decision-making level often need to make quick and ad hoc decisions on treatments and medications, and hospital managers on a higher executive level may have less influence over such decisions. An alternative explanation is that hospitals have greater discretion over selecting elective patients and may strategically select elective patients of lower medical complexity and associated lower treatment costs. It is plausible that such financial considerations impact more profoundly on executive managers' decision making, applying the aforementioned reasoning.

We have found only five studies that have estimated VPCs for the impact of hospitals on quality of care. Although it is problematic to compare results across studies because of great differences in outcome measures, patient groups and error structures, our results are broadly in line with the literature. Previous studies have found that VPC values range widely between less than 1% and 18% depending on disease area and outcome measure (Merlo *et al.*, 2001; Huseynova *et al.*, 2009; Reeves *et al.*, 2010). Two studies use three-level models and found that less variation is attributable to the higher level, confirming our findings (Hauck *et al.*, 2003; Hekkert *et al.*, 2009).

4.2. Hospital level performance

The three-level random intercept model allows us to exploit various approaches and methods of measuring hospital performance based on posterior densities for the unobservable error components at hospital and specialty level. The MCMC from the Bayesian estimation provides separate draws for the error components of e_h for each hospital and u_{sh} for each specialty in each hospital. The estimated error components indicate whether a hospital or hospital-specialty lies above, or below, the complication rate that would be expected according to observable characteristics of the patients they treat. In particular, the smaller (greater) e_h or u_{sh}

for a specific hospital or speciality, the better (worse) this hospital performs in general or in a particular specialty, conditional on observable patients' risk factors.

We first focus on hospital level variation. Table III presents hospital level performance, grouped by types of hospitals, with the posterior mean, standard deviation and 95% credibility interval of e_h for each hospital. According to the posterior means of e_h, in treating elective patients, the majority of regional hospitals (10 of 14) and city hospitals (6 of 11) perform better than the expectation according to the patients' observable characteristics. All teaching hospitals have higher probability to lie above the expected complication rate according to observable episode characteristics. However, teaching hospitals perform better in treating emergency patients than elective patients; for example, teaching hospital 7 is more likely to lie below the expected complication rate.

To further illustrate the interpretation of posterior densities for e_h of individual hospitals, we plot the whole posterior density distributions for two particular hospitals in Figure 1(a) and (b): the 'best' and 'worst' performing teaching hospitals (with the lowest and highest posterior means of e_h of all teaching hospitals).

Table III. Hospital performance: posterior summary statistics for e_h

Elective inpatients					Emergency inpatients				
Hospital	Posterior mean	SD	2.5%	97.5%	Hospital	Posterior mean	SD	2.5%	97.5%
Specialty					Specialty				
5	−0.266	0.181	−0.618	0.064	5	−0.433	0.254	−0.988	0.015
4	−0.044	0.176	−0.408	0.276	4	0.048	0.229	−0.412	0.468
36	0.030	0.150	−0.256	0.328	36	0.094	0.218	−0.332	0.529
1	0.182	0.202	−0.232	0.561	1	0.141	0.236	−0.319	0.606
Regional					Regional				
35	−0.497	0.173	−0.841	−0.154	35	−0.363	0.194	−0.761	0.011
28	−0.330	0.169	−0.671	−0.003	28	−0.138	0.222	−0.575	0.295
27	−0.218	0.165	−0.536	0.089	20	−0.125	0.158	−0.446	0.179
21	−0.213	0.131	−0.480	0.032	32	−0.124	0.202	−0.516	0.278
25	−0.201	0.148	−0.482	0.105	27	−0.115	0.195	−0.496	0.265
20	−0.200	0.125	−0.459	0.034	34	−0.053	0.203	−0.459	0.350
24	−0.185	0.149	−0.481	0.104	25	−0.040	0.192	−0.421	0.333
29	−0.077	0.174	−0.432	0.267	6	−0.036	0.223	−0.465	0.410
26	−0.055	0.148	−0.340	0.234	22	0.005	0.155	−0.303	0.302
23	−0.026	0.173	−0.371	0.318	26	0.010	0.185	−0.359	0.362
32	0.017	0.166	−0.316	0.334	29	0.018	0.228	−0.423	0.463
34	0.052	0.148	−0.233	0.344	23	0.039	0.199	−0.366	0.428
22	0.105	0.125	−0.139	0.337	24	0.048	0.192	−0.317	0.435
6	0.218	0.157	−0.088	0.524	21	0.068	0.168	−0.269	0.393
City					City				
17	−0.342	0.175	−0.694	−0.008	30	−0.268	0.158	−0.581	0.032
16	−0.308	0.172	−0.654	0.007	16	−0.140	0.213	−0.551	0.274
30	−0.299	0.137	−0.576	−0.041	12	−0.100	0.178	−0.458	0.240
12	−0.145	0.153	−0.439	0.151	17	−0.050	0.207	−0.442	0.363
10	−0.033	0.136	−0.306	0.229	10	−0.039	0.177	−0.396	0.306
33	−0.015	0.174	−0.359	0.313	19	0.048	0.214	−0.368	0.474
19	0.151	0.154	−0.172	0.444	9	0.093	0.165	−0.218	0.415
13	0.152	0.148	−0.133	0.448	11	0.100	0.153	−0.179	0.430
14	0.167	0.136	−0.091	0.446	14	0.133	0.147	−0.153	0.416
9	0.191	0.142	−0.099	0.465	33	0.184	0.208	−0.217	0.610
11	0.219	0.122	−0.005	0.463	13	0.274	0.224	−0.168	0.724
Teaching					Teaching				
7	0.037	0.126	−0.216	0.277	7	−0.092	0.150	−0.373	0.197
8	0.150	0.120	−0.107	0.383	8	0.035	0.154	−0.252	0.351
2	0.196	0.125	−0.046	0.467	2	0.249	0.154	−0.053	0.540
18	0.434	0.137	0.166	0.694	15	0.258	0.151	−0.021	0.564
3	0.510	0.220	0.087	0.948	18	0.358	0.158	0.048	0.684
15	0.536	0.122	0.301	0.774					

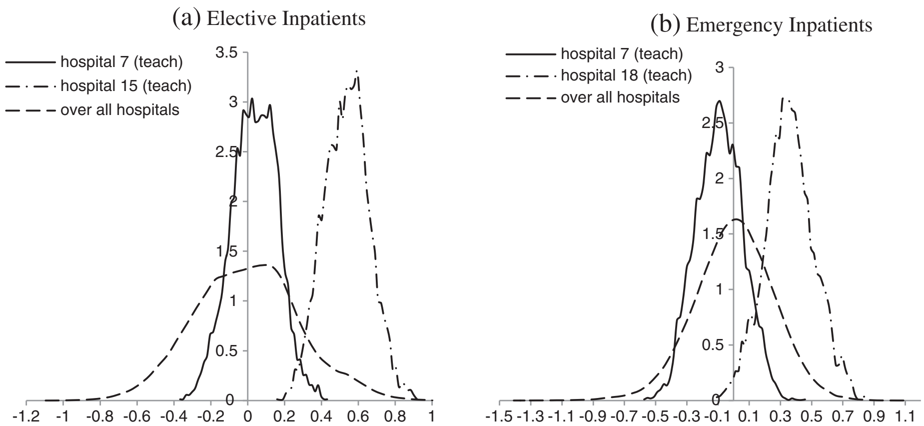

Figure 1. Posterior densities for e_h: over all hospitals and selected hospitals

For comparison, the posterior density for e_h of *all* hospitals is also presented. It is not surprising that e_h of all hospitals approximately follows a normal distribution with mean of zero and standard deviation of 0.34 (elective) and 0.31 (emergency); see the two estimated σ_H in Table II. The posterior densities for the best and worst hospitals, as two of the hospitals contributing to the overall distribution, in Figure 1 indicate very different distributions for the hospital level propensity for complications. For elective inpatients, the best-performing teaching hospital has e_h ranging from -0.4 to 0.4 and concentrating around 0.04; possible values of e_h for the worst-performing teaching hospital are concentrated around 0.54, with almost no probability to be negative. For emergency inpatients, the best-performing teaching hospital has a smaller value for e_h compared with the best-performing teaching hospital for elective episodes, ranging from -0.6 to 0.5 and concentrating around -0.09; the worst-performing teaching hospital has smaller values for e_h relative to those in elective patients, which are concentrating around 0.36.

When comparing posterior distributions across elective and emergency episodes, we find that the variance of hospital performance for elective patients is greater than for emergency patients, which suggests that there is greater scope for improving performance in the care of elective patients. It further suggests that hospital management at the higher executive level has greater scope for improving performance for elective than emergency patients. As discussed before, a possible explanation for this result is that treatment decisions for elective patients can be taken under less time pressure than for emergency patients; hence, there is greater opportunity for hospital managers to impact on those decisions. Alternatively, for elective patients, hospital management may have greater scope in selecting patients with particular characteristics, for example, low medical complexity.

4.3. Which specialty has more scope for improvement?

Posterior variances for each specialty across all hospitals can be a useful indicator for improvement potentials. Episodes in our sample fall into 16 specialties. The estimated posterior density for u_{sh} for each specialty-hospital pair indicates unexplained propensity to complications for a particular specialty in a particular hospital. This allows comparing hospitals with respect to the same specialty. Table IV presents the posterior mean and standard deviation of u_{sh} for each specialty over all hospitals. The results in Table IV show that there are significant differences in latent complication propensity across different specialties. The posterior means identify specialties with higher complication propensity and provide useful information for

DOI: 10.1002/hec

Table IV. Posterior summary statistics for u_{sh} by specialty over all hospitals

Elective inpatients			Emergency inpatients		
Speciality	Posterior mean	SD	Speciality	Posterior mean	SD
Obstetrics	−0.417	0.466	Plastics	−0.580	0.506
Plastics	−0.387	0.354	Orthopaedics	−0.359	0.361
Cardiology	−0.113	0.345	Endocrinology	−0.250	0.510
Endocrinology	−0.090	0.386	Neurosurgery	−0.242	0.542
Respiratory	−0.050	0.358	Ophthalmology	−0.141	0.564
Orthopaedics	−0.015	0.317	ENT	−0.004	0.496
Gynaecology	0.015	0.314	Urology	0.005	0.445
Neurosurgery	0.022	0.399	Cardiology	0.084	0.365
Ophthalmology	0.025	0.365	Obstetrics	0.108	0.477
ENT	0.031	0.306	Vascular	0.133	0.464
Vascular	0.109	0.364	Respiratory	0.134	0.377
Haematology	0.149	0.432	Gynaecology	0.153	0.529
Urology	0.190	0.316	Haematology	0.273	0.601
Cardiothoracic	0.205	0.340	Cardiothoracic	0.313	0.510
General Surgery	0.225	0.242	General Surgery	0.352	0.427
Nephrology	0.265	0.437	Nephrology	0.660	0.729

hospitals with higher proportion of episodes in such high risk specialties. For example, Obstetric elective patients are least likely, and patients treated in Nephrology are most likely to experience a complication during their stay in hospital, on average. For emergency patients, the specialty of Plastics has the lowest complication rate and, again, Nephrology the highest.

The notion of comparing the variances of different specialties (across all hospitals) in Table IV is an interesting one. This allows policy makers to judge in which specialities there is greatest opportunity for improvements in patient safety and target policy efforts towards the hospitals that are underperforming in those specialties. To further illustrate, we plot the whole posterior density distribution of u_{sh} for specialties with highest and lowest posterior mean (Nephrology and Obstetrics for elective patients; Nephrology and Plastics for emergency patients) and highest and lowest standard deviation (Obstetrics and General Surgery for elective patients; Nephrology and Orthopaedics for emergency patients) in Figure 2.

In general, a specialty with *high* posterior mean and *high* variation, in comparison with other specialities, should receive priority attention. Such specialities have high complication rates and large scope for reducing them. An example is Nephrology for emergency episodes. The second question is of course whether there are effective and cost-effective policies to reduce complications in these clinical areas, but if there are, policy

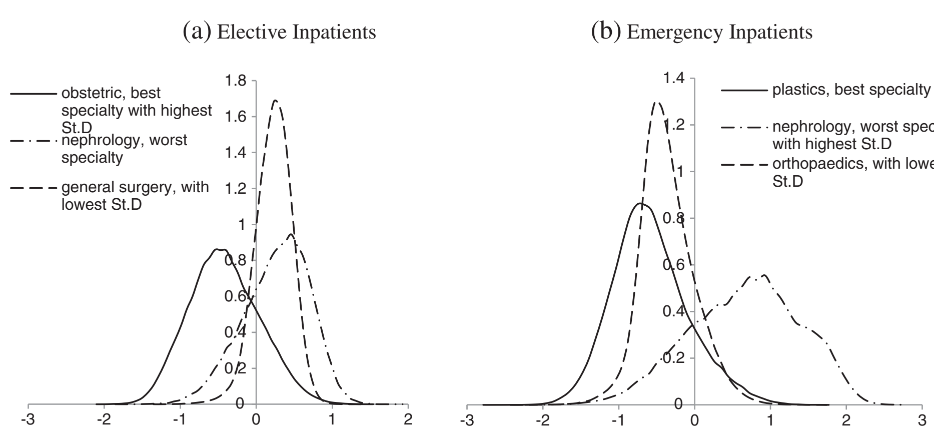

Figure 2. Posterior densities for u_{sh}: speciality with highest or lowest posterior mean, and highest or lowest standard deviation

Table V. Hospital performance within selected specialties: posterior mean for $(e_h + u_{sh})$ – elective episodes

General surgery			Obstetrics			Nephrology		
Hospital	Posterior mean	Ranking	Hospital	Posterior mean	Ranking	Hospital	Posterior mean	Ranking
Specialty			Specialty			Specialty		
5	−0.430	1	4	−0.490	1	21	−0.222	1
36	−0.125	2	1	0.229	2	22	0.029	2
4	0.019	3	Regional			City		
1	0.410	4	25	−1.116	1	10	−0.037	1
Regional			24	−1.095	2	11	0.200	2
35	−0.358	1	20	−0.923	3	Teaching		
28	−0.096	2	21	−0.845	4	8	0.450	1
20	−0.025	3	26	−0.638	5	7	0.481	2
25	0.024	4	35	−0.539	6	2	0.790	3
27	0.081	5	34	−0.521	7	18	0.831	4
21	0.115	6	23	−0.464	8	15	1.290	5
22	0.121	7	32	−0.448	9			
24	0.221	8	27	−0.405	10			
29	0.222	9	22	−0.371	11			
34	0.264	10	28	−0.368	12			
26	0.273	11	29	−0.268	13			
32	0.309	12	6	−0.190	14			
23	0.378	13	City					
6	0.640	14	16	−1.197	1			
City			9	−0.934	2			
30	−0.074	1	17	−0.719	3			
17	0.005	2	30	−0.578	4			
12	0.137	3	10	−0.575	5			
16	0.145	4	33	−0.305	6			
13	0.224	5	14	0.103	7			
33	0.390	6	11	0.191	8			
11	0.394	7	13	0.248	9			
10	0.479	8	19	0.301	10			
19	0.511	9	Teaching					
9	0.516	10	2	−0.745	1			
14	0.521	11						
Teaching								
7	0.203	1						
8	0.207	2						
18	0.454	3						
15	0.498	4						
3	0.525	5						
2	0.602	6						

action in these specialities are likely to lead to an overall reduction in complications and measurable improvements in patient safety. Specialties with *low* posterior mean and *low* variation are possibly low priority areas for action, because complications are comparably low, and there is small scope to reduce those further. Examples are elective episodes in General Surgery and emergency episodes in Orthopaedics. Specialties with *high* posterior mean and *low* variation have high complication rates, but there is unfortunately small scope for reducing them. These clinical areas are of concern, and they should be prioritized for further research into effective strategies that help reduce complications. Lastly, specialities with *low* posterior mean and *high* variation are in general not specialities of prime concern, but specific hospitals with high complication rates for this specialty should be specifically targeted because there is scope for improving their performance. An example is Obstetrics for elective episodes.

4.4. Hospital performance for each specialty

Are well-performing hospitals doing well across all their specialities, or do these hospitals have specialities where complications may give rise to concern? Performance of a hospital in one particular speciality may differ

DOI: 10.1002/hec

from its performance in another speciality. We next compare hospital performance for individual hospitals for the same specialty. For a particular specialty S_0, the posterior mean and density of $(e_h + u_{S_0h})$ for a specific hospital informs how this hospital performs in S_0. To illustrate, we choose three specialties: a specialty that all hospitals in our sample have, General Surgery; the specialty with the lowest mean complication rates, which is Obstetrics for elective inpatients and Plastics for emergency inpatients; and the specialty with the highest mean complication rates for both elective and emergency inpatients, Nephrology.

Tables V and VI show that the same hospital can have marked differences in performance across the three specialties. For example, for elective patients, teaching hospital 2 has high complication rates in General Surgery and Nephrology but very low rates in Obstetrics. Hospital 4, on the other hand, is performing very well in Obstetrics but not in General Surgery. For emergency patients, hospital 5 is performing very well in both General Surgery and Plastics, whereas hospital 23, a regional hospital, is performing very poorly in these two specialties. Figure 3(a) and (b) presents the estimated posterior densities for $(e_h + u_{S_0h})$ of three particular hospitals and of over all hospitals for comparison, for elective episodes in Obstetrics and emergency episodes in Nephrology. The densities indicate how much variation there is in complication rates across patients treated in a

Table VI. Hospital performance within selected specialties: posterior mean for (e_h+u_{sh}) – emergency episodes

General surgery			Plastics			Nephrology		
Hospital	Posterior Mean	Ranking	Hospital	Posterior Mean	Ranking	Hospital	Posterior Mean	Ranking
Specialty			Specialty			City		
5	−0.739	1	5	−1.121	1	11	−0.180	1
36	−0.446	2	4	−0.048	2	Teaching		
1	−0.311	3	1	−0.024	3	7	0.087	1
4	−0.297	4	36	0.499	4	8	0.479	2
Regional			Regional			2	1.202	3
35	−0.093	1	22	−1.412	1	18	1.370	4
20	0.157	2	35	−1.075	2	15	1.909	5
28	0.188	3	20	−0.778	3			
27	0.297	4	21	−0.610	4			
29	0.359	5	28	−0.610	5			
25	0.371	6	29	−0.575	6			
21	0.377	7	24	−0.558	7			
32	0.382	8	6	−0.549	8			
22	0.432	9	26	−0.539	9			
6	0.461	10	34	−0.529	10			
34	0.568	11	27	−0.486	11			
26	0.614	12	23	−0.248	12			
24	0.732	13	25	−0.248	13			
23	0.751	14	32	0.051	14			
City			City					
19	0.256	1	17	−1.132	1			
30	0.270	2	30	−1.106	2			
13	0.369	3	16	−0.911	3			
11	0.481	4	12	−0.886	4			
17	0.627	5	9	−0.873	5			
14	0.657	6	10	−0.860	6			
9	0.659	7	14	−0.605	7			
16	0.722	8	11	−0.434	8			
10	0.729	9	33	−0.352	9			
12	0.797	10	13	0.100	10			
33	0.810	11	19	0.137	11			
Teaching			Teaching					
7	−0.126	1	7	−1.211	1			
15	0.407	2	15	−1.010	2			
8	0.415	3	8	−0.757	3			
18	0.545	4	18	−0.669	4			
2	0.615	5	2	−0.197	5			

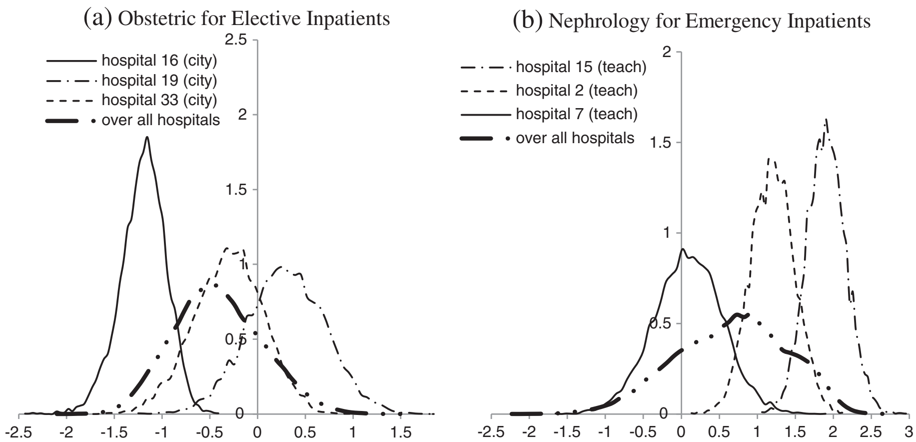

Figure 3. Posterior densities for $(e_h + u_{sh})$: over all hospitals and selected hospitals for selected specialties

speciality in a particular hospital. Comparably high variation for a speciality in one hospital, in comparison with variation for the same specialty in another hospital, may indicate that there is greater scope for improvement. Again, this interpretation relies crucially on that other exogenous sources of variation, notably patient complexity, is accounted for. For Obstetrics, hospital 16 has only negative values for $(e_h + u_{S_0h})$ coupled with comparably small variation around the mean, which indicates that there may be little scope to improve performance even further, whereas for hospitals 19 and 33, which have higher complication rates, variations are also greater, implying that there may be greater scope for improvement. For Nephrology, the two hospitals with high complication rates have at the same time small variations, which may indicate that there may only be small scope for improvements in their poor performance.

5. CONCLUSION

This study contributes to the literature on the performance of hospitals and public sector organizations in general. We estimate a three-level random intercept probit model to decompose the effects of unobservable risk factors for hospital-acquired complications into hospital effects, hospital-specialty effects and remaining random effects using a large-scale patient-level administrative hospital dataset from public hospitals in Victoria, Australia. Our main innovative contributions are that we adopt a broader perspective than most previous studies because we jointly analyse outcomes for patients across 16 clinical areas. Second, our use of hospital-acquired complications is an extension to the more frequently used mortality or readmission rates as measures of quality. Third, we estimate a three-level model that can disentangle effects by hospital and specialty within hospital levels. Lastly, and most importantly, our study contributes to the wider literature on hospital performance by demonstrating the advantage of a Bayesian hierarchical modelling approach when faced with large multilevel datasets. Our full Bayesian approach allows easy access to the whole posterior distributions of the effects of latent risk factors specific to a particular hospital and specialty, rather than just the means of the various error components. We demonstrate that variances of the error posterior densities can offer useful insights. A comparably larger variance for a specialty or department suggests a greater scope for improvements in performance. This provides valuable insights on areas which should be prioritized for interventions aimed at improving patient safety. To our knowledge, it is the first time that a study on hospital performance can present this kind of evidence to policy makers.

Of the variation in complication propensity unexplained by observable patient characteristics, we find that a larger proportion is attributable to the lower specialty or departmental level, in particular for emergency patients. This suggests to policy makers that although efforts to implement patient safety initiatives should ideally involve both executive hospital management and medical teams on specialty level, action on the lower

DOI: 10.1002/hec

level is of prime importance. Variation in complications rates at hospital level as measured by the posterior densities for the hospital level effects confirms this result. These variations are greater for elective than for emergency patients, suggesting that hospital management at higher executive level has greater scope for improving performance in elective than emergency episodes.

Variation in complication rates at specialty level across all hospitals as measured by the posterior densities for the specialty-hospital effects provides some interesting results on identifying specialties that should be prioritized for policy action. We find that the specialties Nephrology and General Surgery give cause for concern because they have both high complications rates. However, whereas Nephrology has high variation in rates across hospitals, General Surgery has comparably small variation. This suggests that policy action should first focus on Nephrology, because there is greater scope for reducing these high rates than in General Surgery. Obstetrics has comparably low rates but coupled with high variation. This suggests that although Obstetrics is not a clinical area of great concern, policy action may focus on selected hospitals that have high rates in this specialty because results indicate that there is scope for improvement. Orthopaedics has low complication rates coupled with low variation, which suggests that this specialty is not a priority area for action.

Our results also answer the interesting question whether well-performing hospitals do well across all their specialties and vice versa. We find that there is surprisingly little correlation in performance across specialties within one hospital. Furthermore, we can use posterior means of hospital and specialty effects to assess the need for action for specialties in specific hospitals and couple this information with the posterior densities as measures of scope for improvements in performance. This allows us to prioritize hospitals for action not only on the basis of their average performance but also whether they can reasonably be expected to improve their performance.

A potential limitation of our study, and in fact of most other studies in this area, is that adjustments for casemix complexity may be inadequate because of unobservable, systematic differences in patients' medical complexity across hospitals not captured in casemix adjustment (Iezzoni, 1997). Another limitation is that we do not have information to shed light on the unobserved hospital and department-specific factors such as safety procedures and protocols that impact on unobserved variations in complication rates.

It would be counterproductive to use our results to punish poorly performing hospitals. Poor performance on patient safety should in the first instance lead to a careful investigation of the organizational and economic constraints in which the affected hospital has to operate and a constructive search by all levels of government for ways of working within these realities to help hospitals to improve performance. The results of our research will hopefully contribute to the implementation of effective measures to improve patient safety in hospitals, reduce costs of hospital care and ultimately, save many patients from unnecessary harm.

ACKNOWLEDGEMENTS

We thank the discussant and participants of the Australasian Workshop on Econometrics and Health Economics, Sydney, 2012, for useful comments and Don Poskitt, Monash University, for the valuable discussions.

REFERENCES

ACSQHC. 2011. Australian Commission on Safety and Quality in Health Care. (Available from: http://www.safetyandquality.gov.au/, Retrieved 27 October 2011)

AHRQ. 2011. Agency for Healthcare Research and Quality. (Available from: http://www.ahrq.gov/, Retrieved 27 October 2011)

Australian Bureau of Statistics. 2001. Socio-economic indexes for areas. *Information Paper Census of Population and Housing*, Canberra.

Cebul RD, Rebitzer JB, Taylor LJ, Votruba M. 2008. Organizational fragmentation and care quality in the U.S. health care system. No. w14212. National Bureau of Economic Research.

DOI: 10.1002/hec

Charlson ME, Pompei P, Ales KL, Mackenzie CR. 1987. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Disease* **40**: 373–383.
Cho S-H, Ketefian S, Barkauskas VH, Smith DG. 2003. The effects of nurse staffing on adverse events, morbidity, mortality, and medical costs. *Nursing Research* **52**(2): 71–79.
CMS. 2007. Medicare program: changes to the hospital inpatient prospective payment systems and fiscal year 2008 rates.
CMS. 2011. Hospital-acquired conditions (present on admission indicator). (Available from: https://www.cms.gov/HospitalAcqCond/06_Hospital-Acquired_Conditions.asp#TopOfPage, Retrieved 11/3/2011)
Coelli TJ, Rao DSP, O'Donnell CJ, Battese GE. 2005. An Introduction to Efficiency and Productivity Analysis. (2nd edn). Springer: New York.
Daniels MJ, Gatsonis C. 1999. Hierarchical generalized linear models in the analysis of variations in health care utilization. *Journal of the American Statistical Association, American Statistical Association* **94**: 29–42.
Droesler S. 2008. Facilitating cross-national comparisons of indicators for patient safety at the health-system level in the OECD countries. OECD: Paris.
Duncan C, Jones K, Moon G. 1998. Context, composition and heterogeneity: using multilevel models in health research. *Social Science & Medicine* **46**(1): 97–117.
Ehsani JP, Jackson T, Duckett SJ. 2006. The incidence and cost of adverse events in Victorian hospitals 2003–04. *The Medical Journal of Australia* **184**: 551–555.
Ehsani J, Duckett S, Jackson T. 2007. The incidence and cost of cardiac surgery adverse events in Australian (Victorian) hospitals 2003–2004. *The European Journal of Health Economics* **8**: 339–346.
Geweke JF, Keane MP, Runkle DE. 1997. Statistical inference in the multinomial multiperiod probit model. *Journal of Econometrics* **80**: 125–165.
Glance L, Li Y, Osler T, Mukamel D, Dick A. 2008. Impact of date stamping on patient safety measurement in patients undergoing CABG: experience with the AHRQ Patient Safety Indicators. *BMC Health Services Research* **8**(1): 176.
Graves N. 2004. Economics and preventing hospital-acquired infection. *Emerging Infectious Diseases* **10**(4): 561–566.
Gupta M, Chang W-C, Van de Werf F, Granger CB, Midodzi W, Barbash G, … Armstrong PW. 2003. International differences in in-hospital revascularization and outcomes following acute myocardial infarction: a multilevel analysis of patients in ASSENT-2. *European Heart Journal* **24**(18): 1640–1650.
Hauck K, Zhao X. 2011. How dangerous is a Day in hospital?: a model of adverse events and length of stay for medical inpatients. *Medical Care* **49**(12): 1068–1075.
Hauck K, Rice N, Smith PC. 2003. The influence of health care organizations on health system performance. *Journal of Health Services Research & Policy* **8**: 68–74.
Hauck K, Zhao X, Jackson T. 2012. Adverse event rates as measures of hospital performance. *Health Policy* **104**: 146–154.
Hekkert KD, Cihangir S, Kleefstra SM, van den Berg B, Kool RB. 2009. Patient satisfaction revisited: a multilevel approach. *Social Science & Medicine* **69**(1): 68–75.
Hinchey JA, Shephard T, Tonn ST, Ruthazer R, Selker HP, Kent DM. 2008. Benchmarks and determinants of adherence to stroke performance measures. *Stroke A Journal of Cerebral Circulation* **39**(5): 1619–1620.
Hollingsworth B. 2003. Non-parametric and parametric applications measuring efficiency in health care. *Health Care Management Science* **6**(4): 203–218.
Huseynova K, Xiong W, Ray JG, Ahmed N, Nathens AB. 2009. Venous thromboembolism as a marker of quality of care in trauma. *Journal of the American College of Surgeons* **208**(4): 547–552.
Iezzoni L. 1997. The risks of risk adjustment. *JAMA : The Journal of the American Medical Association* **278**: 1600–1607.
Jackson T, Duckett S, Shepheard J, Baxter K. 2006. Measurement of adverse events using ‘incidence flagged’ diagnosis codes. *Journal of Health Services Research & Policy* **11**: 21–26.
Jacobs R, Smith PC, Street A. 2006. Measuring Efficiency in Health Care: Analytical Techniques and Health Policy. Cambridge University Press: Cambridge.
Karlsen S, Say, L, Souza J-P, Hogue C, Calles D, Gulmezoglu AM, Raine R. 2011. The relationship between maternal education and mortality among women giving birth in health care institutions: analysis of the cross sectional WHO Global Survey on Maternal and Perinatal Health. *BMC Public Health* **11**(1): 606.
Kim S, Shephard N, Chib S. 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies* **65**: 361–393.
Klevens RM, Edwards JR, Richards CL, Horan TC, Gaynes RP, Pollock DA, Cardo DM. 2007. Estimating health care-associated infections and deaths in U.S. hospitals, 2002. *Public health reports (Washington, D.C.: 1974)* **122**: 160–6.
Kohn LT, Corrigan J, Donaldson MS (eds.) 2000. To Err is Human: Building a Safer Health System. Vol. 627 National Academies Press: Washington, D.C.
Koop G. 2003. Bayesian Econometrics. John Wiley & Sons Ltd: England.
Koop G, Osiewalski J, Steel MFJ. 1997. Bayesian efficiency analysis through individual effects: hospital cost frontiers. *Journal of Econometrics* **76**: 77–105.
Kumbhakar S, Lovell CAK. 2000. Stochastic Frontier Analysis. Cambridge University Press: New York.

DOI: 10.1002/hec

McCulloch R, Rossi PE. 1994. An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics* **64**: 207–240.

McKelvey R, Zavoina W. 1975. A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology Summer* 103–120.

McNair P, Borovnicar D, Jackson T, Gillett S. 2009. Prospective payment to encourage system wide quality improvement. *Medical Care* **47**: 272–278.

Merlo J, Östergren P-O, Broms K, Bjorck-Linné A, Liedholm H. 2001. Survival after initial hospitalisation for heart failure: a multilevel analysis of patients in Swedish acute care hospitals. *Journal of Epidemiology and Community Health* **55**(5): 323–329.

Michel JL, Nghiem HS, Jackson T. 2009. Using ICD-10-AM codes to characterise hospital-acquired complications. *Health Information Management Journal* **38**: 18–25.

Moje C, Jackson T, McNair P. 2006. Adverse events in Victorian admissions for elective surgery. *Australian Health Review* **30**: 333–343.

Moreno R, Metnitz PH, Almeida E, Jordan B, Bauer P, Campos R, … Le Gall J-R. 2005. SAPS 3—From evaluation of the patient to evaluation of the intensive care unit. Part 2: development of a prognostic model for hospital mortality at ICU admission. *Intensive Care Medicine* **31**(10): 1345–1355.

NPSA. (2011). National Patient Safety Agency. (Available from: http://www.npsa.nhs.uk/, Retrieved 27 October 2011)

Pronovost P, Garrett E, Dorman T, Jenckes M, Webb Iii T, Breslow M, … Bass E. 2001. Variations in complication rates and opportunities for improvement in quality of care for patients having abdominal aortic surgery. *Langenbeck's Archives of Surgery* **386**(4): 249–256.

Pronovost PJ, Berenholtz SM, Goeschel CA, Needham DM, Sexton JB, Thompson DA, Lubomski LH, Marsteller JA, Makary MA, Hunt E. 2006. Creating high reliability in health care organizations. *Health Services Research* **41**: 1599–1617.

Reeves MJ, Gargano J, Maier KS, Broderick JP, Frankel M, LaBresh KA, … Schwamm L. 2010. Patient-level and hospital-level determinants of the quality of acute stroke care: a multilevel modeling approach. *Stroke* **41**(12): 2924–2931.

Rice N, Jones A. 1997. Multilevel models and health economics. *Health Economics* **6**: 561–575.

Rice N, Leyland A. 1996. Multilevel models: applications to health data. *Journal of Health Services Research & Policy* **1**(3): 154–164.

Rivard PE, Rosen AK, Carroll JS. 2006. Enhancing patient safety through organizational learning: are patient safety indicators a step in the right direction? *Health Services Research* **41**: 1633–1653.

Schreyoegg J, Stargardt T, Tiemann O. 2011. Costs and quality of hospitals in different health care systems: a multi-level approach with propensity score matching. *Health Economics* **20**(1): 85–100.

Smith PC, Mossialos E, Papanicolas I, Leatherman S (eds.) 2009. Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects. Cambridge University Press: Cambridge.

Stone PW. 2009. Economic burden of healthcare-associated infections: an American perspective. *Expert Review of Pharmacoeconomics & Outcomes Research* **9**: 417–422.

Stone P, Glied S, McNair P, Matthes N, Cohen B, Landers T, Larson E. 2010. CMS changes in reimbursement for HAIs: setting a research agenda. *Medical Care* **48**: 433–439.

Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. 2004. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *Journal of clinical epidemiology* **57**(12): 1288–1294.

Vincent C. 2010. *Patient Safety*. West Sussex, BMJ Publishing Group Ltd: Chichester.

de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermaster MA. 2008. The incidence and nature of in-hospital adverse events: a systematic review. *Quality & Safety in Health Care* **17**: 216–223.

WHO. 2008. Implementation Manual Surgical Safety Checklist (1st edn). WHO (ed.), WHO: Geneva, Switzerland.

WHO. 2009. WHO Guidelines on Hand Hygiene in Health Care.

WHO. 2011. Patient safety. (Available from: http://www.who.int/patientsafety/en/, Retrieved 27 October 2011)

Zeger SL, Karim MR. 1991. Generalized linear models with random effects: a Gibbs sampling approach. *Journal of the American Statistical Association, American Statistical Association* **86**: 79–86.

Zhang X, Hauck K, Zhao X. forthcoming. Quantifying effect of unobservable risk factors on hospital acquired complications: a Bayesian analysis. Working paper, Department of Econometrics and Business Statistics, Moansh University, Australia.