
A Grid Implementation for Profiling Hospitals Based on Patient Readmissions

Eren Demir¹, Thierry J. Chausalet¹, Noam Weingarten², and Tamas Kiss²

¹ Health and Social Care Modelling Group, University of Westminster, London, United Kingdom

² Centre for Parallel Computing, University of Westminster, London, United Kingdom

Abstract. Generally, high level of readmission is associated with poor patient care, hence its relation to the quality of care is plausible. Frequent patient readmissions have personal, financial and organisational consequences. This has motivated healthcare commissioners in England to use emergency readmission as an indicator in the performance rating framework. A statistical model, known as the multilevel transition model was previously developed, where individual hospitals propensity for first readmission, second readmission, third (and so on) were considered to be measures of performance. Using these measures, we defined a new performance index. During the period 1997 and 2004, the national (England) hospital episodes statistics dataset comprise more than 5 million patient readmissions. Implementing a statistical model using the complete population dataset could possibly take weeks to estimate the parameters. Moreover, it is not statistically sound to utilise the full population dataset. To resolve the problem, we extract 1000 random samples from the original data, where each random sample is likely to lead to differing hospital performance measures. For computational efficiency a Grid implementation of the model is developed. Using a stand-alone computer, it would take approximately 500 hours to estimate 1000 samples, whereas in the Grid implementation, the full 1000 samples were analysed in less than 24 hours. Analysing the output from the full 1000 sample, we noticed that 4 out of the 5 worst performing hospitals treating cancer patients were in London.

Keywords: emergency readmissions, Grid, multilevel modelling, hospital performance.

1 Introduction

Emergency hospital readmission has been a matter of concern due to their implications for both cost [1] and quality of hospital care [2]. Healthcare commissioners in England use emergency readmission as an indicator in the performance rating framework, where hospitals are rated based on their levels of readmission. Since it was introduced, 42% of hospitals have reduced emergency readmission to hospital following treatment for a fractured hip and 49% for those following treatment for stroke [3]. Therefore, emergency readmission is seen as an important part of the government's commitment to improve the quality of care that patients receive. In this context, patients and the public have the right to know how well different National Health Service (NHS) Trusts are performing. Trust is a term used in the UK, which can either be recognised as a single hospital or one or more hospitals combined within a small region.

Performance monitoring and the profiling of hospitals can be useful to generate feedbacks to hospitals for possible intervention policies to reduce levels of readmission.

Furthermore, it could benefit patients' selection of hospitals. Patient choice has been at the heart of the UK government's public service reform agenda to empower patients and reduce inequalities in access to healthcare [4]. However, negative implications include the ranking of hospitals through various sources of the media, and this apparently sheds light to those hospitals with the 'best' and 'worst' ranks. Hence, it is important that an appropriate quantitative modelling framework captures the variability in the readmission process between NHS Trusts. We have identified a number of concerns with the current adopted methodology.

The current method by the healthcare commissioners, profiles hospitals based on the estimates of the standardised percentage of emergency admissions within 28 days of a discharge from hospital (patients aged 16 and over). An emergency admission within 28 days of discharge from hospital (respectively greater than 28 days) is classified as readmitted (respectively non-readmitted). The process takes into account differences between types of patients by their age and gender. This process is known as the risk adjustment step to account for possible differences in patient case mix. For each hospital, the observed number of readmissions is determined. Given the case mix, the expected number of readmissions for hospital k is estimated (\hat{y}_k). Here, \hat{y}_k is determined by multiplying the number of emergency admissions to hospital k by the national average rate of readmission. The ratio of the observed to expected readmissions is interpreted as the hospital's standardised readmission rate (SR_k), which is multiplied by the overall readmission rate of patients in England (ϕ), providing an indicator value (as a %) known as the 'risk-adjusted readmission rate' ($\mu_k = 100 \times SR_k \times \phi$), which forms the basis for comparisons between hospitals. There are a number of issues related to this method.

First, \hat{y}_k could be misleading, where the number of emergency admissions to hospital k is multiplied by the national average rate of readmission. In some cases, such as Barts and The London NHS Trust, who provides diverse range of treatments and services to complex patients, which is inevitable to expect a higher numbers of emergency admissions. On the other hand, trusts treating less severe patients could experience a lower numbers of emergency admissions. Therefore, the estimation of \hat{y}_k is biased towards less active hospitals.

Second, in the literature, age and gender are the two predictors that are mostly insignificant to account for variations among patients, including the Hospital Episodes Statistics (HES) dataset, which is the data used by healthcare commissioners. For instance, [5,6] derived that patient ages from 75 to 79 had the highest rate of readmissions. However, [7] found that patients less than 60 years of age were more likely to be readmitted. [8] suggested a significant increase for patients 80 years and older, [9] for patients between 66 and 75, whereas others found that age was not a significant factor [10-13]. We notice conflicting results with regards to age. So, the case-mix process does not necessarily take into account variations between populations treated by individual hospitals. Other crucial failures include the inadequacy for accounting patient severity, not to mention numerous other anomalies such as the socioeconomic status and social factors related to the patient. A study by [7] found that more deprived patients has a higher risk of readmission.

Third, the healthcare commissioner's performance ratings framework defines re-admission for adults as an emergency or unplanned admission to the same hospital within 28 days following discharge. The time window for defining readmission varied according to the purpose of the study, generally from 30 [14] to 90 [15] days, but some studies have used readmissions following certain surgeries, for shorter [16] (14 days) or longer [17] time window (1 year). A number of authors have expressed their concerns in the appropriate choice of time windows in defining readmission [18,19]. Therefore, the definition of readmission for the classification process (readmitted and non-readmitted patients) plays a vital role. The choice of time windows will inevitably affect the outcome of profiling hospitals.

We have seen many published articles that profile hospitals for coronary artery bypass grafting or acute myocardial infarction patients [20-23]. However, little attention has been found for profiling hospitals based on patient readmissions. Modern statistical approaches, such as multilevel logistic regression [24,25] are frequently used in the profiling process. These methods may not capture within and between hospital variability through the random intercept or random slope, but may not necessarily indicate a good (or bad) performing hospital. Second, the majority of healthcare databases are highly administrative, and in the absence of detailed clinical data, particularly to determine individual patient's disease severity, it is difficult to derive the relevant risk adjusting factors. Third, multilevel modelling in the context of profiling hospitals is generally used to determine the expected number of outcomes for each hospital, given hospital specific parameter estimates; and the difference between the observed outcome is said to be a measure of quality of care. If additional patient or hospital characteristics are incorporated into the model, the expected number of readmissions for each hospital could change, which questions the accuracy of the model.

Having recognised the limitations of the currently adopted method by the healthcare commissioners in England, and some of the statistical concerns in the literature, we developed a new framework for profiling hospitals based on patient readmissions. This article's context is illustrated in Figure 1.

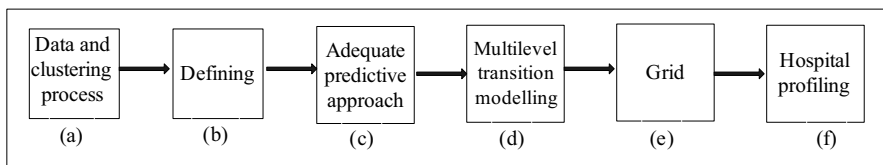


Fig. 1. This articles context and the process from data to hospital profiling

In section 2, we describe the data and cluster patients that experience similar length of stay (LoS) in the community before readmission (Figure 1 (a)). In doing so, we decompose a large number of clinical conditions into several sub-groups, where each group can be examined separately. Instead of using the 28 days defined by Department of Health (subjective definition), we objectively defined readmission based on a modelling approach that we have previously developed [26], where the optimal time window is determined for each of the clustered patient sub-groups (Figure 1 (b)).

Using the estimated time window, patients with a LoS in the community within the time window (respectively greater than the time window) are classified into the high risk group (respectively low risk). We then dichotomise as 1 for the high risk and 0 for the low risk group. This classification argument provides an approach to the transformation of the time spent (continuous variable) in the community of a patient into a binary response.

In the interest of profiling hospitals, identification of a method that provides an adequate predictive approach (Figure 1 (c)) is essential, ensuring that our model provides a good discrimination and accuracy in explaining variations in patient readmissions. Patient's past history of readmissions is known to be one of the most significant variable in determining the risk of readmission [7, 15]. In section 3, we introduce the transition model, which allows the incorporation of patient's past history of readmissions along with additional covariates. A transition model is selected for each of the clustered patient sub-groups, which is further used in the multilevel framework for hospital profiling purposes.

The effectiveness of the transition model further facilitated us with the extension to the multilevel transition model (MTM) (Figure 1 (d)), to determine individual hospital's propensity for a patient to be in the high risk group of readmission, given their past history of readmissions. We assumed that every hospital has its own propensity for first, second, third (and so on) readmissions. Hence, hospital specific estimates, known as the multilevel transition effects can be interpreted as indicators of hospitals performance. For instance, some hospitals may have higher propensity for first readmissions than others. Section 4 briefly describes the MTM and defines a new performance index.

Each clustered patient sub-group comprises more than three hundred thousand patient readmissions. Implementing a statistical model, such as multilevel modelling, using the complete population dataset could possibly take weeks to estimate the parameters. Moreover, it is not statistically sound to utilise the full population dataset. To resolve the problem, we extract 1000 random samples from the original data, where each random sample is likely to lead to differing hospital performance measures. For computational efficiency a Grid implementation of the model is developed (Figure 1 (e)). Section 5 describes the Gridification process in greater detail. Section 6 illustrates the results (Figure 1 (f)) and conclusion and further works are in section 7.

2 Data and the Clustering Process

In this section, we introduce the data used for profiling hospitals and the process of clustering patients that experience similar LoS in the community before readmission.

2.1 Data

The Department of Health in the UK releases annually its national database, the Hospital Episode Statistics (HES). The HES dataset contains personal, medical and administrative details of all patients admitted to, and treated in, NHS hospitals in England. There are approximately 12 million records for each financial year. A financial year is from 1st April to 31st March the following year. The HES dataset captures all

the consultant episodes of patients during their stay in hospital. During a hospital stay a patient might encounter several successive episodes, collectively known as a spell. Spells ending with discharge by death are excluded, as death within 30 days of discharge is used as another performance indicator, and should not be combined with re-admission. As a result, we limit our data selection to patients who had a subsequent admission following a discharge. Since our aim is to study patients who are admitted to hospital soon after their discharge, this data selection procedure is justified.

The HES data was provided in a disc format, including 7 financial year periods between 04/1997 - 03/2004 (approximately 80 million episodes in total). The data was in a flat file and necessary steps were taken to restore the data in to a relational database. Each financial year was imported into MySQL version 5.0 and a longitudinal dataset was constructed using all patient episodes joining across financial years.

The covariates used in the transition model and in the multilevel transition model are the Charlson index of comorbidity [27], age, gender, previous LoS in hospital, and Index of Multiple Deprivation (IMD) 2004 [28]. The Charlson index of comorbidity is a measure of patient severity, which is based on ICD-10 diagnosis codes, where various weights are attached to the presence of conditions such as congestive heart failure and cancer. The Charlson index of comorbidity is further reduced to a categorical variable with eight levels (0-7+), i.e. 0 indicates no severity and 7+ indicates high severity. The IMD is a weighted index based on seven factors of deprivation, which can be recognised and measured separately, and are related to: income; employment; health and disability; education, skills and training; barriers to housing and services; living environment and crime. Charlson comorbidity index and IMD are not available in the HES dataset and a number of steps had to be taken to incorporate this information.

2.2 Clustering Similar Length of Stay

In the context of performance ratings framework, profiling hospitals for each clinical condition may not be practical, as there are hundreds of clinical conditions. NHS hospitals are rated using aggregated levels of readmission. Previously in [29] we demonstrated that the use of a single number (i.e. 28 days) across all clinical conditions may not be an appropriate choice of a time window. To resolve this problem, we take a high level overview of all inpatient readmissions, and cluster all chapters according to LoS distribution quantiles in the community before readmission. This method allows the computation of time windows for several clustered patient sub-groups, rather than determining time windows for each clinical condition separately. Based on each clustered sub-group, we profile hospitals using their unique estimated optimal time windows.

Using the HES dataset from 1997 to 2004, we extracted 73,576,883 episodes from patients who had primary diagnosis codes corresponding to one of 22 chapters, where each chapter covers many clinical conditions. Each chapter was derived using the patients ICD-10 codes. For example, ICD-10 codes A00-B99 refer to chapter 1, where A00-A79 are bacterial infections, other intestinal infectious diseases, and sexually transmitted diseases, A80-B34 are viral infections, B35-B89 are infections caused by fungi, protozoans, worms, and infestations and B90-B99 refer to sequelae, and diseases classified elsewhere. We group these patients as chapter 1 and proceed similarly for the remaining chapters. Chapter 5 (mental and behavioural disorders), chapter 20

(external causes of morbidity and mortality) and chapter 22 (codes for special purposes) are missing, as no data were available. A set of 66,336,588 spells (or patient admissions) was derived (see Table 1).

Using the time window of 28 days, we observed that in the case of diseases of the respiratory system (chapter 10), 26% of all readmissions occurred within 28 days of

Table 1. Live discharges and readmissions between 1997 and 2004 in the HES dataset. Levels of readmissions are the proportion of readmissions within a 28 day time window. Data are not available for chapters 5, 20 and 22.

Chapter number	Chapter name	No. of spells	No. of readmissions	Levels of readmissions
16	Certain conditions originating in the perinatal period	243187	6592	0.89
15	Pregnancy, childbirth and the puerperium	4029152	212250	0.54
8	Diseases of the ear and mastoid process	603949	5646	0.50
17	Congenital malformations, deformations & chromosomal abnormalities	514160	9100	0.47
21	Factors influencing health status and contact with health services	4176837	53214	0.45
7	Diseases of the eye and adnexa	2563736	16350	0.43
2	Neoplasms	8484428	278773	0.43
1	Infectious and parasitic diseases	918549	63128	0.40
3	Diseases of the blood & blood forming organs	1008047	89814	0.39
13	Diseases of the musculoskeletal system and connective tissue	4305451	95130	0.34
14	Diseases of the genitourinary system	5081217	185166	0.32
19	Injury, poisoning & certain other consequences of external causes	4598407	555920	0.30
12	Diseases of the skin and subcutaneous tissue	1661306	82928	0.30
11	Diseases of the digestive system	7780287	436526	0.29
4	Endocrine, nutritional and metabolic diseases	879568	112166	0.26
6	Diseases of the nervous system	1423243	140032	0.26
9	Diseases of the circulatory system	5874495	870815	0.26
18	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	8162033	991419	0.26
10	Diseases of the respiratory system	4271723	826425	0.26

discharge, whereas, 89% for conditions originating in the perinatal period. This is a clear indication that using a single number in defining readmission across all clinical conditions may not be an appropriate choice of a time window.

We now illustrate a method of clustering patients that experience similar LoS in the community. Here, the objective is to avoid the computation of time windows for individual clinical conditions. In doing so, we decompose a large number of clinical conditions into several sub-groups, where each group can be examined separately. This approach could be beneficial for healthcare commissioners in the UK, where hospitals can be rated for each clustered subgroups with their unique estimated time windows. First, we extract LoS for each of the chapters of the ICD-10 codes illustrated in Table 1. For each chapter, we derive quantiles between 0% and 100% in steps of 10% [30]. Then, a hierarchical cluster analysis is performed to group all chapters with similarities based on the shape of the cumulative distribution function of LoS.

We used the average linkage of the Euclidean distance for clustering [31]. Figure 2 illustrates four clustered sets of sub-groups with similar LoS in terms of distributions. The clustered groups were identified as follows: Cluster 1 (chapter 16); Cluster 2 (chapters 2, 3, 17); Cluster 3 (chapters 9, 11, 12, 13, 14, 15, 18, 19); and Cluster 4 (chapters 1, 4, 6, 7, 8, 10, 21). Furthermore, we tested the data with quantiles of 5%, 7.5%, 12.5% and 15% and found that there were no changes in the clustered set of sub-groups. So, the clustering process is not sensitive to the chosen quantiles. Hence, the 22 chapters are decomposed into 4 sub-groups of patients that experience similar LoS. We notice that the main condition in cluster 2 is cancer (neoplasms), diseases of

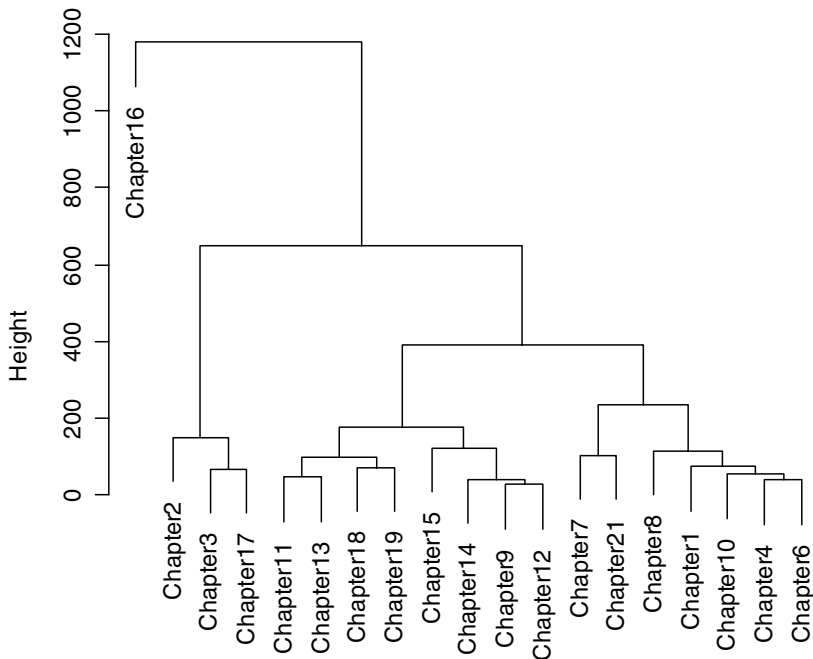


Fig. 2. Clustered LoS using chapters extracted from the HES dataset

the circulatory system for cluster 3, and diseases of the respiratory system is the dominant one for cluster 4 patient sub-group.

Throughout this article, we do not consider cluster 1 patients. Chapter 16 is the only condition in cluster 1. This chapter is a condition that originates during the perinatal period, i.e. 5 months before birth and 1 month after. It comprises females only and profiling hospitals based on a single sex may not be appropriate; the readmission time interval is very short (maximum 180 days); a large number of hospitals have very few perinatal readmissions, where the sample size becomes a major issue; and the total number of readmissions in England over the 7 financial year period is 6592, which is very low in comparison to the other clustered sub-groups.

2.3 Time Window for the Clustered Patient Sub-groups

We determined the optimal time window for the clustered patient sub-group datasets using a modelling approach that we have previously developed [26]. The estimated time windows are 62, 36, and 35 days for clusters 2, 3 and 4, respectively. Using the estimated time windows, patients with a length of stay in the community for the three clustered patient sub-groups within 62, 36 and 35 days of discharge from hospital (respectively greater than 62, 36 and 35 days) are classified into the high risk group (respectively low risk). We then dichotomise the repeated measurements as 1 for the high risk and 0 for the low risk group. This classification argument provides an approach to the transformation of the time spent in the community of a patient into a binary response.

In the following section, we describe a “transition model” which allows the incorporation of patient’s past history of readmissions along with additional covariates.

3 An Adequate Predictive Approach

In the interest of profiling hospitals, identification of a method that provides an adequate predictive approach is essential, ensuring that our model provides a good discrimination and accuracy in explaining variations in patient readmissions. In this section, we evaluate the effectiveness of using readmission data, with particular emphasis on patient’s previous readmissions to identify those who are at high risk of readmission. The effectiveness of this model further facilitated the extension to the multilevel analogue to determine individual hospital’s propensity for a patient to be in the high risk group of readmission given their past history of readmissions.

Previously, we reported the effectiveness of the transition model for predicting the risk of readmission for chronic obstructive pulmonary disease patients [32]. Here, we apply the transition model for each of the clustered patient sub-groups. The objective is to select a transition model for each sub-group, i.e. a first, or second, or third (and so on) transition model. So, our goal here is to develop a parsimonious probability model that could further be used in the multilevel transition framework to determine individual hospital’s propensity for a patient to be in the high risk group of readmission.

A well established method of summarising the ability of the model to correctly discriminate between outcomes is the receiver operating characteristic (ROC) c statistic. This method is used to select the appropriate model for each of the clustered patient sub-groups.

3.1 An Overview of the Transition Model

In general patients may encounter several successive readmissions. Let Y_{ij} be the random variable for all patient readmissions, where $i = 1, \dots, m$ and $j = 1, \dots, n_i$ readmissions. Therefore, $Y_{ij} = 1$ if the i -th patient is in the high risk group at readmission j , and $Y_{ij} = 0$ otherwise. In the presence of covariate information, let

$f(y_{i1}, \dots, y_{in_i}) = \prod_{j=1}^{n_i} f(y_{ij} | H_{ij})$, where y_{ij} is the realization (0 or 1) of the random variable Y_{ij} and $H_{ij} = \{y_{i1}, \dots, y_{ij-1}\}$ is the history of past readmissions for the i -th patient available up to readmission j . Bernoulli random variable is $f(y_{ij} | H_{ij}) = (\mu_{ij}^c)^{y_{ij}} (1 - \mu_{ij}^c)^{1 - y_{ij}}$, where $\mu_{ij}^c = P(Y_{ij} = 1 | H_{ij})$ is the conditional mean of a patient being in the high risk group at occasion j , which depends on the patient's past readmissions through the vector H_{ij} . The conditional mean μ_{ij}^c on the logistic scale is a linear function of H_{ij} on a p -dimensional covariate vector \mathbf{x}_{ij} ,

$$h(\mu_{ij}^c) = \text{logit}(\mu_{ij}^c) = \log\left(\frac{\mu_{ij}^c}{1 - \mu_{ij}^c}\right) = \mathbf{x}_{ij}^T \boldsymbol{\beta}_q + \sum_{r=1}^s f_r(H_{ij}; \boldsymbol{\alpha}), \tag{3.1}$$

where $\boldsymbol{\beta}_q$ is a vector of unknown fixed coefficients, which describes a transition model of order $q \in \{1, \dots, q_{\max}\}$, where $1 \leq q_{\max} \leq n_i - 1$ is the maximal order. The role of $f_r(\cdot)$ is to account for previous readmissions. Here, we consider a *logit link* function for $f_r(\cdot)$, where $f_r(H_{ij}; \boldsymbol{\alpha}) = \alpha_r y_{ij-r}$. The transition model expresses the conditional mean μ_{ij}^c as a function of both covariates \mathbf{x}_{ij} and the past responses $y_{ij-1}, \dots, y_{ij-q}$. From (3.1), $h(\mu_{ij}^c)$ is a linear function of both $\boldsymbol{\beta}_q$ and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_s)$, hence the estimation proceeds as in the classical generalized linear model for independent data. We simply regress Y_{ij} on the $(p + s)$ -dimensional vector of extended explanatory variables $(\mathbf{x}_{ij}, f_1(H_{ij}), \dots, f_s(H_{ij}))$. Therefore, we apply the ordinary logistic regression maximum likelihood procedure. If the maximum number of readmissions per patient is 5, there are only four transition models to consider; the first, second, third and fourth order.

In the next section, we briefly describe the clustered patient sub-groups and illustrate six models, where past history of readmissions is incorporated sequentially to examine the impact on the ROC c statistic.

3.2 Application to Clustered Patient Sub-groups

From the 66,336,588 spells we extracted from HES, approximately 10 million of these belong to cluster 2, 40 million to cluster 3, and 15 million to cluster 4 patient sub-groups. From these spells, we further extracted 1.25, 18 and 5 million emergency

Table 2. A breakdown of the HES dataset based on the clustered patient sub-groups

Cluster	No. of spells	No. of emergency admissions discharged live	No. of electives (and other) discharged live	No. of emergency readmissions	No. of deaths
2	10,006,635	1,250,033	8,407,312	377,686	349,290
3	41,492,348	17,636,074	22,935,536	3,430,133	920,738
4	14,837,605	5,134,698	9,271,196	1,216,958	431,711

admissions, where patient readmissions comprised 0.4, 3.5, and 1.2 million for clusters 2, 3 and 4, respectively (refer to Table 2 for exact numbers).

From these patient readmissions a fair number of records were incomplete. For instance, incorrectly entered NHS Trust provider codes and missing local authority codes. Furthermore, for the hospital profiling process, hospitals with fewer than 500 readmissions were removed from the study, as 500 readmissions from a seven year dataset with hundreds of clinical conditions is rather small. This is an arbitrarily chosen number of readmissions, which could possibly change according to clinical expertise. Therefore, after the data cleansing process the number of records used in the transition model (and multilevel transition model) is 337,953, 3,270,325 and 1,133,398 for cluster 2, cluster 3, and cluster 4, respectively.

We now illustrate six different models for each of the clustered patient sub-groups, where higher order transition models are considered. These models illustrate the use of logistic regression to fit simple transition models. The estimated coefficients for these models are in Tables 3 - 5. The first model predicts the risk of being in the high risk group using covariates, where the patient’s history of readmission effect is set to 0 ($\alpha_r = 0$). The estimated coefficients (β is independent from q) for each of the clustered patient sub-groups in Tables 3 – 5 can be interpreted as the usual logistic regression. However, this interpretation for β_q is not valid for the transition model. The coefficient β_q gives the per unit change in the corresponding x_{ij} in the log odds of being in the high risk group among patients at their q previous visits. This interpretation changes with transition order. The estimated ROC c statistics is 0.63 for cluster 2, 0.62 for cluster 3 and 0.60 for cluster 4 patient sub-group. Suggesting that the first model is not much better than chance in determining risk of readmission.

In models 2 to 6 we study patient’s previous readmission effects, where past history of readmissions is incorporated sequentially to examine the impact on the ROC c statistic. Here, we test the model up to five previous readmissions for each of the clustered patient sub-groups and observe that the highest impact on the ROC c statistic is in model 2, where the incorporation of a patient’s previous readmission increases the discrimination dramatically. For example, from Table 3 (cluster 2) the ROC c statistics for model 1, 2 and 3 are 0.63, 0.69 and 0.72, respectively. So, the highest increase is from model 1 to model 2, i.e. from a model with no previous readmissions (model 1) to a model incorporating patient’s one previous readmission

Table 3. Logistic regression coefficients and p-values (square parentheses) for the six models fitted to the (“high” and “low” risk patients) transitions for patients in cluster 2. LoS refers to patient’s prior length of stay and ROC is the receiver operating characteristic. [*] and [^] refers to p-values < 0.05 and > 0.05, respectively. Model 1 refers to no previous readmission effects. Models 2 - 6 incorporate the effects of patient’s previous readmissions sequentially. CC_Index refers to Charlson comorbidity index.

Variable	1	2	3	4	5	6
Intercept	0.551[*]	0.038 [*]	-0.320[*]	-0.639[*]	-0.871[*]	-1.061[*]
IMD	-0.003[^]	-0.005[^]	0.001[*]	0.004[*]	0.007[*]	0.009[*]
CC_Index1	-0.250[*]	-0.192[*]	-0.155[*]	-0.166[*]	-0.175[*]	-0.179[*]
CC_Index2	0.294[*]	0.232[*]	0.147[*]	0.079[*]	0.026[^]	-0.047[*]
CC_Index3	0.138[*]	0.158[*]	0.127[*]	0.083[^]	0.085[^]	0.085[^]
CC_Index4	0.161[*]	0.180[*]	0.072[*]	0.011[^]	0.025[^]	0.011[^]
CC_Index5	0.260[*]	0.276[*]	0.219[*]	0.156[*]	0.044[^]	0.002[^]
CC_Index6	0.197[*]	0.213[*]	0.247[*]	0.253[*]	0.082[^]	0.015[^]
CC_Index7+	0.194[*]	0.256[*]	0.288[*]	0.259[*]	0.204[*]	0.279[*]
Age	-0.006[*]	-0.009[*]	0.001[*]	0.003[*]	0.003[*]	0.004[*]
LoS	-0.004[*]	-0.004[*]	-0.004[*]	-0.003[*]	-0.003[*]	-0.003[*]
Gender	-0.059[*]	-0.045[*]	-0.026[^]	-0.014[^]	-0.027[^]	-0.033[^]
$Y_{ij-1} (\alpha_1)$		0.868[*]	0.898[*]	0.912[*]	0.920[*]	0.907[*]
$Y_{ij-2} (\alpha_2)$			0.628[*]	0.589[*]	0.601[*]	0.617[*]
$Y_{ij-3} (\alpha_3)$				0.566[*]	0.593[*]	0.484[*]
$Y_{ij-4} (\alpha_4)$					0.471[*]	0.429[*]
$Y_{ij-5} (\alpha_5)$						0.425[*]
ROC c	0.63	0.69	0.72	0.73	0.74	0.75

(model 2). Similar results are found for clusters 3 and 4 (see Tables 4 and 5). The previous two readmissions have higher effect (Y_{ij-1}, Y_{ij-2}) than readmissions occurring in the distant past, e.g. in model 4 (Table 5) for cluster 4 patient sub-group, the estimated past history effects are $\alpha_1 = 0.7608$, $\alpha_2 = 0.5729$ and $\alpha_3 = 0.4911$. Thus, a decreasing trend is noticeable. These results are also supported by the remaining clustered sub-groups.

Patients who are consistently observed to be in the high risk group have a much greater chance to be in the high risk group in their next readmission. This illustrates the cumulative effect of previous behaviour on readmission history. Patients who were in the high risk group in their previous five readmissions for cluster 2 have 17.5 (respectively 7 and 10 for clusters 3 and 4) times the odds of being in the high risk group at their next readmission compared to those who were in the low risk group at their previous five readmissions.

An advantage of this model is that the probability of a patient being in the high risk group can be estimated using both the past history of readmissions and covariates. For

Table 4. Logistic regression coefficients and p-values (square parentheses) for the six models fitted to the (“high” and “low” risk patients) transitions for patients in cluster 3. LoS refers to patient’s prior length of stay and ROC is the receiver operating characteristic. [*] and [^] refers to p-values < 0.05 and > 0.05, respectively. Model 1 refers to no previous readmission effects. Models 2 - 6 incorporate the effects of patient’s previous readmissions sequentially. CC_Index refers to Charlson comorbidity index.

Variable	1	2	3	4	5	6
Intercept	-0.152[*]	-0.589[*]	-0.784[*]	-0.889[*]	-0.967[*]	-1.039[*]
IMD	-0.005[^]	-0.002[^]	-0.004[^]	0.002[^]	0.004[^]	0.008[^]
CC_Index1	-0.087[*]	-0.057[*]	-0.048[*]	-0.033[*]	-0.034[*]	-0.026[^]
CC_Index2	-0.070[*]	-0.009[^]	-0.008[^]	0.015[^]	0.012[^]	0.014[^]
CC_Index3	-0.110[*]	-0.003[^]	0.039[*]	0.043[*]	0.024[^]	0.011[^]
CC_Index4	-0.075[*]	0.027[^]	0.055[*]	0.075[*]	0.070[*]	0.021[^]
CC_Index5	0.042[^]	0.124[*]	0.134[*]	0.114[*]	0.107[^]	0.077[^]
CC_Index6	0.027[^]	0.366[^]	0.075[^]	0.111[^]	0.083[^]	0.169[^]
CC_Index7+	0.169[*]	0.268[*]	0.292[*]	0.307[*]	0.303[*]	0.336[*]
Age	-0.009[*]	-0.006[*]	-0.006[*]	-0.004[*]	-0.004[*]	-0.003[*]
LoS	0.001[*]	0.003[*]	0.0031[*]	0.002[*]	0.001[*]	0.007[^]
Gender	0.050[*]	0.034[*]	0.049[*]	0.016[*]	0.015[^]	0.015[^]
$Y_{ij-1} (\alpha_1)$		0.704[*]	0.680[*]	0.673[*]	0.667[*]	0.672[*]
$Y_{ij-2} (\alpha_2)$			0.485[*]	0.467[*]	0.457[*]	0.453[*]
$Y_{ij-3} (\alpha_3)$				0.365[*]	0.347[*]	0.338[*]
$Y_{ij-4} (\alpha_4)$					0.302[*]	0.285[*]
$Y_{ij-5} (\alpha_5)$						0.236[*]
ROC c	0.62	0.69	0.70	0.71	0.72	0.73

example, the probability that a patient from cluster 2 (model 3) is in the high risk ($Y_{ij} = 1$) group given their past two prior readmissions ($Y_{ij-1} = 1, Y_{ij-2} = 1$) fell in the high risk group, has an IMD score of 40 (deprived), Charlson comorbidity index of 2, age of 65, LoS of 3 days and is a male is estimated to be 81%.

The strength of this formulation is the ease of adding additional predictors to the model. Despite the fact that the index of multiple deprivation and Charlson comorbidity index were included, the ROC c statistic show that patient’s past readmissions provide a better discrimination between “high” and “low” risk group of readmission as opposed to using solely covariates (model 1). A limitation of the attached IMD scores is that it is based at the local authority level (wider area). If the full postcode of the patient was available, the IMD scores could have been incorporated at the super output area level, i.e. minimum population of a 1000 within an area. Accurate IMD scores are likely to have more impact on the transition effects than the ones we have used here.

Table 5. Table 5 Logistic regression coefficients and p-values (square parentheses) for the six models fitted to the (“high” and “low” risk patients) transitions for patients in cluster 4. LoS refers to patient’s prior length of stay and ROC is the receiver operating characteristic. [*] and [^] refers to p-values < 0.05 and > 0.05, respectively. Model 1 refers to no previous readmission effects. Models 2 - 6 incorporate the effects of patient’s previous readmissions sequentially. CC_Index refers to Charlson comorbidity index.

Variable	1	2	3	4	5	6
Intercept	-0.836[*]	-1.071[*]	-1.221[*]	-1.315[*]	-1.398[*]	-
IMD	-0.005[*]	0.002[*]	0.002[*]	0.003[*]	0.003[*]	0.003[*]
CC_Index1	-0.315[*]	-0.191[*]	-0.123[*]	-0.094[*]	-0.077[*]	-
CC_Index2	-0.242[*]	-0.131[*]	-0.076[*]	-0.059[*]	-0.038[^]	-
CC_Index3	-0.307[*]	-0.188[*]	-0.147[*]	-0.114[*]	-0.082[*]	-
CC_Index4	-0.356[*]	-0.195[*]	-0.155[*]	-0.116[*]	-0.099[^]	-
CC_Index5	-0.088[*]	-0.069[^]	-0.084[^]	-0.079[^]	-0.157[^]	-
CC_Index6	-0.059[^]	0.018[^]	-0.107[^]	-0.137[^]	-0.141[^]	-
CC_Index7+	0.214[*]	0.431[*]	0.310[*]	0.308[*]	0.276[^]	0.254[^]
Age	0.002[*]	0.003[*]	0.003[*]	0.003[*]	0.003[*]	0.003[*]
LoS	0.002[*]	0.006[*]	0.007[*]	0.008[*]	0.009[*]	0.010[*]
Gender	-0.004[^]	-0.009[^]	-0.009[^]	-0.009[^]	-0.006[^]	-
$Y_{ij-1} (\alpha_1)$		0.863[*]	0.799[*]	0.761[*]	0.743[*]	0.731[*]
$Y_{ij-2} (\alpha_2)$			0.634[*]	0.573[*]	0.531[*]	0.517[*]
$Y_{ij-3} (\alpha_3)$				0.491[*]	0.443[*]	0.424[*]
$Y_{ij-4} (\alpha_4)$					0.411[*]	0.372[*]
$Y_{ij-5} (\alpha_5)$						0.339[*]
ROC c	0.60	0.68	0.69	0.71	0.73	0.74

The ROC c statistics ranges from 0.69 - 0.77 for models three to six, producing a good discrimination, illustrating that such simple models have the potential of predicting outcomes based on patient’s past history of readmissions. However, for each of the clustered patient sub-groups, models three to six exhibit relatively small increases in the ROC c statistics. For the purpose of utilising a high number of patient observations in the profiling process and selecting a parsimonious probability model, a trade-off between the two is considered. In the transition model, as the transition order increases the number of patient observations decrease. Hence, the ROC c statistics for models 2, 3, and 4 for the cluster 2 patient sub-group are 0.69, 0.72 and 0.73, respectively (see Table 3). The ROC c statistics greater or equal to 0.70 [33] is regarded as a good discrimination between the high risk and low risk group patients. Thus, it seems appropriate to select model 3, as the increase in the ROC c statistics is negligible. Similarly, we select model 3 for cluster 3 (Table 4) and model 4 for cluster 4 patient sub-group (Table 5).

Additional information might improve our prediction accuracy, however, the model is sufficiently powerful without additional covariates. An extension of the transition

model described above is to incorporate individual hospital effects of patient transitions, i.e. the effect of α_1 (first order transition effect) is expected to vary between hospitals. In the next section, we briefly describe the multilevel transition model (MTM), which will be used in the profiling process. The multilevel framework will be implemented based on the selected transition models for each of the clustered patient sub-groups.

4 Multilevel Transition Model

The transition model assumed a conditional distribution on patient's readmissions given the individual patient's past history of readmissions. Hospital-to-hospital heterogeneity was not considered. In the context of the organisational consequences of patient readmissions, profiling of hospitals is an area of interest to most healthcare managers and healthcare purchasers. Accepting that each hospital has its unique transition effects (i.e. some hospitals may have higher propensity for first readmissions) we developed an extension of the transition model to the multilevel transition model [29].

The selected transition model for cluster 2 patients was a second order transition model, i.e. patient's previous two readmissions were considered. The population average first (α_1) and second (α_2) order transition effects were 0.898 and 0.628, respectively (see Table 3). The MTM is formulated such that it determines individual hospitals deviation from α_1 and α_2 , i.e. negative deviation from α_1 indicates that the hospitals propensity of being in the high risk group for first order transition effect (first readmissions) is lower than the population average first order transition effect, thus, a good performing hospital. The MTM also determines the deviation from α_2 . Hence, the final performance index is defined to be the sum of these two deviations, known as the sum effects. So, if a third order MTM is considered, then the hospitals deviation from α_1 , α_2 and α_3 is determined, where the sum of these deviations is the final performance index.

Cluster 2, cluster 3 and cluster 4 comprise 337,953, 3,270,325 and 1,133,398 patient readmissions, respectively. Implementing MTM using the open source program R [34] with the complete population dataset could possibly take weeks to estimate the parameters. Moreover it is not statistically sound to utilise the full population dataset. It can cause the acceptance of non-significant covariates (hypothesis testing). To resolve the problem of using the full dataset, we perform random sampling from the original data, where each random sample is likely to estimate differing sum effects. A large number of samples is required, and sampling over many samples using a standalone personal computer can be very time consuming and cumbersome. For computational efficiency a Grid implementation of MTM is considered. We perform one thousand random sampling with replacement. The average sum effect over the one thousand random samples is the final performance index.

5 Grid

The Grid is a collection of computers, storage devices, special services that can dynamically join and leave the Grid [35]. It is useful when a user has a complex problem

that requires many services (i.e. many samples) in order to reduce computation time. The University of Westminster has a 256 central processing unit (CPU) computing cluster, which is part of the UK National Grid Service. The term “cluster” in the Grid is defined as connected computers. Hence, this should not be confused with our clustered patient groupings. There are two basic cluster types: nondedicated clusters are simply a network of workstations; dedicated clusters, where all the computers are connected for high performance (computers are parallelised). We used dedicated clusters.

5.1 The Gridification Process

The P-Grade Portal 2.5 [36] with parameterisation was used to prepare a prototype workflow. The P-Grade portal is an easy to use yet highly functional, web-based user interface for using the power of the Grid. Being web-based, the Portal is accessible from any computer connected to the internet, through all common browsers. Figure 3 illustrates the workflow.

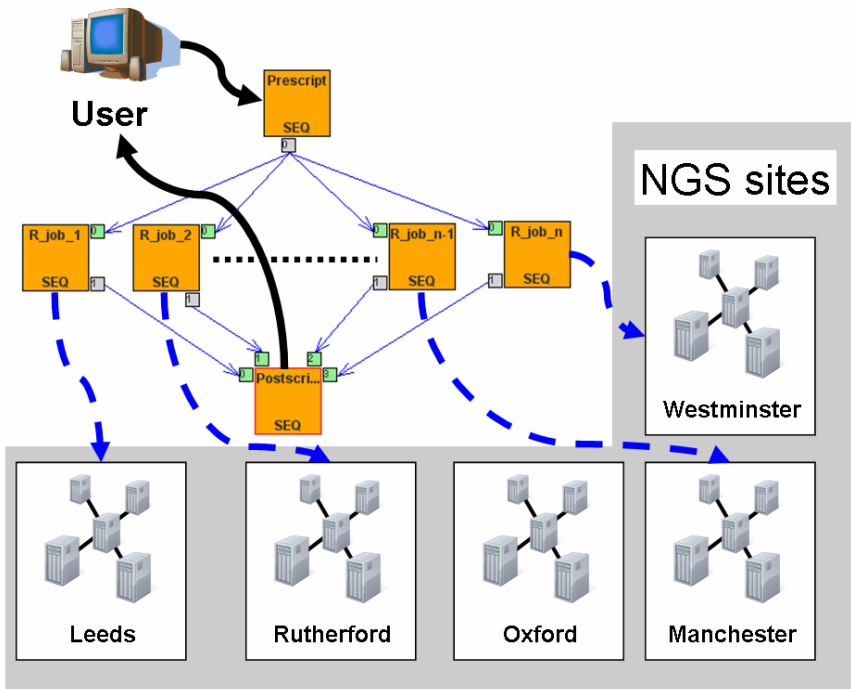


Fig. 3. The principles of the workflow based Grid implementation. The pre-script is the process where the data (i.e. cluster 2), R-script and the number of samples (i.e. 1000) are specified. Post-script collates the output from each sample. NGS refers to National Grid Service.

The pre-script is the process where the data (i.e. cluster 2), R-script and the number of samples (i.e. 1000) are specified by the user. The R-script extracts a random sample from the original data, estimates the parameters and computes the sum effects. Workflows prepared for these experiments were executed on National Grid Service

resources including; Rutherford Appleton Laboratories (Cambridge), and the Universities of Oxford, Leeds, Westminster & Manchester. The workflows were submitted using the workflow manager. The P-Grade Portal handles file distribution, job submission (sample 1 to sample 1000) and result collection. The post-script (see Figure 3) collates the output from each sample.

6 Application of the MTM in Grid Specific Features

The study comprised 167 NHS Acute and Foundation Trusts in England. In section 2.2, we clustered patients that experience similar length of stay in the community into three groups. The dominant disease in cluster 2 was the cancer patients, diseases of the circulatory system and diseases of the respiratory system were the main condition for cluster 3 and cluster 4 patients, respectively. From the 337,953, 3,270,325 and 1,133,398 patient readmissions in cluster 2, 3, and 4, respectively, a random sample (with replacement) of 15% from the original data is extracted, i.e. approximately 50000, 500000 and 170000 readmissions for cluster 2, 3, and 4, respectively. The average numbers of observations with 3 or more readmissions for a second order MTM from each of the randomly selected sample data are approximately 17000 and 96000 for cluster 2 and 3, respectively. A third order MTM (≥ 4 readmissions) contains approximately 41000 observations. Fitting a second order MTM with 96000 observations can take up to 30 minutes, hence, a dramatic reduction in computation time is observed by using the Grid.

One thousand samples were submitted to the specified Grid resource, and stacked up in the job-manager's queue, each to be executed on an available CPU as and when it becomes available. Potentially, at any one time all the CPUs hosted by that Grid resource could be executing the R-script, each taking one sample from the dataset, and analysing that sample of records. The P-Grade Portal handles the transfer of the outputs once each job has been completed. Once all the samples have been completed, the portal executes the post-script on the results of all the samples, and combines them into the desired comma delimited (csv) format. These results will then be easily downloadable from the P-Grade Portal as a simple download from a website.

Using a standalone computer, it would take approximately 500 hours to estimate 1000 samples from the cluster 3 subgroup, however in the Grid implementation, the full 1000 samples were analysed in less than 24 hours.

6.1 Individual Hospital Performance Indices

The average sum effects over the 1000 random samples varied from a low of -0.62 to a high of 1.72 for cluster 2, -1.55 to 2.50 for cluster 3 and -1.29 to 1.37 for cluster 4 patient sub-groups, i.e. the minimum sum of hospital effects from the 1000 random sample for cluster 2 was -0.62 and the maximum was 1.72 . Tables 6 – 8 illustrate the national performance indices for the highest 5 and the lowest 5 ranked hospitals in England. From Table 6 four out of the five worst performing hospitals were in London for cluster 2 patients.

Hospital 194 is ranked as 167th with sum effects 1.722 , where the observed proportion of patients in the high risk group is 0.84 for cluster 2 subgroup. Consider hospital 27,

Table 6. Highest and lowest ranked hospitals for cluster 2 patients. First and second order transition effects are the mean (over the 1000 samples) of k th hospitals deviation from the population of first and second order transition effects. A negative transition effect implies a good performing hospital. Sum Effects = First order transition effects + Second order transition effects.

Hospital identifier	Sum effects	First order transition effect	Second order transition effect	Rank
146	-0.624	-0.294	-0.330	1
54	-0.537	-0.278	-0.259	2
376	-0.493	-0.249	-0.244	3
30	-0.482	-0.258	-0.224	4
45	-0.461	-0.213	-0.248	5
15	0.951	0.449	0.502	163
301	1.051	0.513	0.538	164
27	1.067	0.543	0.524	165
260	1.193	0.580	0.613	166
194	1.722	0.849	0.873	167

Table 7. Highest and lowest ranked hospitals for cluster 3 patients. First and second order transition effects are the mean (over the 1000 samples) of k th hospitals deviation from the population of first and second order transition effects. A negative transition effect implies a good performing hospital. Sum Effects = First order transition effects + Second order transition effects.

Hospital identifier	Sum effects	First order transition effect	Second order transition effect	Rank
341	-0.193	-0.076	-0.117	1
366	-0.181	-0.083	-0.098	2
13	-0.167	-0.082	-0.085	3
172	-0.159	-0.069	-0.090	4
64	-0.156	-0.059	-0.097	5
398	0.138	0.073	0.065	163
27	0.157	0.079	0.078	164
70	0.238	0.098	0.140	165
227	0.270	0.114	0.156	166
143	0.329	0.164	0.165	167

which was identified as having potential problems for all clustered patient sub-groups. This hospital, in the North West of England region, is a large acute teaching hospital with an observed proportion of 0.86, 0.40, and 0.39 of patients in the high risk group for the clustered sub-groups of patients, respectively.

Table 8. Highest and lowest ranked hospitals for cluster 4 patients. First, second and third order transition effects are the mean (over the 1000 samples) of k th hospitals deviation from the population of first, second and third order transition effects. A negative transition effect implies a good performing hospital. Sum effects = First order transition effects + Second order transition effects + Third order transition effects.

Hospital identifier	Sum effects	First order transition ef-	Second or- der transi-	Third or- der transi-	Rank
106	-0.260	-0.057	-0.094	-0.109	1
322	-0.232	-0.054	-0.073	-0.105	2
28	-0.205	-0.026	-0.073	-0.106	3
374	-0.181	-0.035	-0.080	-0.066	4
225	-0.179	-0.084	-0.042	-0.053	5
144	0.203	0.057	0.057	0.089	163
27	0.204	0.036	0.080	0.088	164
3	0.250	0.042	0.116	0.092	165
189	0.281	0.084	0.084	0.113	166
293	0.289	0.104	0.088	0.097	167

High risk group patients for clusters 2, 3, and 4 are defined to be those patients that are readmitted within 62, 36, and 35 days, respectively (see section 2.3). We notice that hospital 398 is ranked as 163rd for cluster 3 and 7th (not included in Table 8) for cluster 4, indicating that a hospital can perform exceptionally well for one group of patients and have some quality issues for another.

7 Conclusion and Further Works

We have investigated several important issues concerning patient readmissions from a national perspective. Two key questions were addressed, namely, clustering clinical conditions that experience similar length of stay in the community before readmission, and the profiling of hospitals using a method that we have previously developed in Grid specific features. The process to profiling hospitals followed a number of steps, such as objectively defining readmission for each of the clustered patient sub-groups, and using the transition model, firstly to illustrate the effectiveness of using patient's past history of readmissions for predicting the risk of readmission, and secondly, selecting a parsimonious transition model for the multilevel analogue.

The performance measures were estimated using individual hospitals propensity of being in the high risk group for first, second and third (and so on) readmissions. We showed that these hospital effects can be linked to indices of hospital performance and how estimates of such indices can be derived in the course of fitting a MTM approach in Grid specific features (Gridification process). In this respect, the Grid provided a dramatic reduction in time and effort, as the output for the 1000 sample was provided in an Excel comma delimited (.csv) file that could easily be analysed for profiling hospitals.

The method of clustering similar length of stay in the community, the method of defining readmission, the developed Multilevel transition modelling framework for profiling purposes, and the Gridification process are all generic methods, which have been applied here to an English dataset and could easily be applied to any other datasets related to patient readmissions, whether or not related to differing countries or healthcare infrastructures.

The MTM and the implementation of this method using the Grid is a new technique, therefore a number of opportunities for further work could easily arise, such as further statistical developments of MTM, and using the Grid for more demanding experiments. For instance, patients nested within hospitals, where hospitals are nested within regions. Therefore, the transition effects will have to be estimated for each hospital and for each region, which would be a highly computationally intensive process, where the Grid can be a very useful tool.

References

1. Walter, R.: The revolving door of hospital re-admissions. *Caring* 17, 56–59 (1998)
2. Luthi, J.C., Burnand, B., McClellan, W.M., Pitts, S.R., Flanders, W.D.: Is readmission to hospital an indicator of poor process of care for patients with heart failure? *Quality and Safety of Health Care* 13, 46–51 (2004)
3. Healthcare Commission. Healthcare Commission 2005 performance ratings (April 25, 2007), <http://ratings2005.healthcarecommission.org.uk/home.asp>
4. Department of Health (June 08, 2008), <http://ratings2005.healthcarecommission.org.uk/home.asp>
5. Kossovsky, M.P., Perneger, V.T., Sarasin, P.F., Bolla, F., Borst, F., Gaspoz, J.M.: Comparison between planned and unplanned readmissions to a department of internal medicine. *Journal of Clinical Epidemiology* 52, 151–156 (1999)
6. Kossovsky, M.P., Perneger, V.T., Sarasin, P.F., Chopard, P., Sigaud, P., Gaspoz, J.M.: Unplanned readmissions of patients with congestive heart failure: Do they reflect in-hospital quality of care or patient characteristics? *American Journal of Medicine* 109, 386–390 (2000)
7. Lyratzopoulos, G., Havelly, D., Gemmell, I., Cook, A.G.: Factors influencing emergency medical readmission risk in a UK district general hospital: A prospective study. *BioMed Central Emergency Medicine* 5, 1–9 (2005)
8. Marcantonio, R.E., McKean, S., Goldfinger, M., Kleefield, S., Yurkofsky, M., Brennan, A.T.: Factors associated with unplanned hospital readmission among patients 65 years of age and older in a Medicare managed care plan. *American Journal of Medicine* 107, 13–17 (1999)
9. Rayner, C.H., Temple, R.M., Marshall, T., Clarke, D.: A comparison of hospital readmission rates between two general physicians with different outpatient review practices. *BioMed Central Health Services* 2, 1–8 (2002)
10. Ashton, M., Kuykendall, H., Johnson, L., Wray, O., Wu, L.: The association between the quality of inpatient care and early readmission. *Annals of Internal Medicine* 122, 415–421 (1995)
11. Ko, C.F., Yu, T.K.K., Ko, T.P.S.: A survey of hospital readmission in elderly patients. *Hong Kong Medical Journal* 2, 258–262 (1996)
12. Pearson, B., Skelly, R., Wileman, D., Masud, T.: Unplanned readmission to hospital: A comparison of the views of general practitioners and hospital staff. *British Geriatrics Society* 31, 141–143 (2002)
13. Gooding, J., Jette, A.: Hospital readmissions among the elderly. *Journal of American Geriatrics Society* 33, 595–601 (1985)
14. Kemper, A., Kennedy, E., Dechert, R., Saint, S.: Hospital readmissions for bronchitis. *Clinical Pediatrics* 44, 509–513 (2005)

15. Moloney, E., Bennett, K., Silke, B.: Patient and disease profile of emergency medical re-admissions to an Irish teaching hospital. *Postgraduate Medical Journal* 80, 470–474 (2004)
16. Reed, R., Pearlman, R., Buchner, D.: Risk factors for early unplanned hospital readmission in the elderly. *Journal of General Internal Medicine* 6, 223–228 (1991)
17. Datar, A., Sood, N.: Impact of postpartum hospital-stay legislation on newborn length of stay, readmission, and mortality in California. *Pediatrics* 118, 63–72 (2006)
18. Jimenez, A.P., Garcia, J.A., Gomez, J.A., Hidalgo, L.R., Lorenzo, L.N., Perea, M.L.E., Fernandez, C.N.J.: Readmission rate as an indicator of hospital performance: the case of Spain. *International Journal of Technology Assessment in Health Care* 20, 385–391 (2004)
19. Gray, L.: Readmission of elderly patients to hospital: still ill-defined and poorly understood – a response. *International Journal for Quality in Health Care* 13, 181–182 (2001)
20. Burack, J., Impellizzeri, P., Homel, P., Cunningham, N.: Public report of surgical mortality: A survey of New York State cardiothoracic surgeons. *Annals of Thoracic Surgery* 68, 1195–1200 (1999)
21. Green, J., Wintfeld, N.: Report cards on cardiac surgeons-assessing New York State’s approach. *New England Journal of Medicine* 332, 1229–1233 (1995)
22. Krumholz, H.M., Normand, S.L., Spertus, J.A., Shahian, D.M., Bradley, E.H.: Measuring performance for heart attacks and heart failure: the case for outcomes measurement. *Health Affairs* 26, 75–85 (2007)
23. Shahian, D.M., Torchiana, D.F., Normand, S.L.: Implementation of a cardiac surgery report card: lessons from the Massachusetts experience. *Annals of Thoracic Surgery* 80, 1146–1150 (2005)
24. Normand, S.L.T., Glickman, M.E., Gatsonis, C.A.: Statistical methods for profiling providers of medical care: issues and applications. *Journal of the American Statistical Association* 92, 803–814 (1997)
25. Leyland, A.H., Boddy, F.A.: League tables and acute myocardial infarction. *Lancet* 351, 555–558 (1998)
26. Demir, E., Chausalet, T.J., Xie, H., Millard, P.H.: Emergency Readmission Criterion: A Technique for Determining Emergency Readmission Time Window. *IEEE Transactions on Information Technology in Biomedicine* 12, 644–649 (2008)
27. Sundararajan, V., Henderson, T., Perry, C., Muggivan, A., Quan, H., Ghali, W.A.: New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *Journal of Clinical Epidemiology* 57, 1288–1294 (2004)
28. Office of the Deputy Prime Minister: The English indices of deprivation. Report (2004)
29. Demir, E.: Modelling Readmission: Defining and Developing a Framework for Profiling Hospitals. Unpublished PhD thesis (2008b)
30. Virtue, A., Chausalet, T., Millard, P., Whittlestone, P., Kelly, J.: A system for patient management based discrete-event simulation and hierarchical clustering. In: Proceedings of 19th IEEE Symposium on Computer-Based Medical Systems, pp. 800–804 (2006)
31. Chatfield, C., Collins, A.J.: Introduction to Multivariate Analysis. Chapman and Hall, Boca Raton (1980)
32. Demir, E., Chausalet, T.J., Xie, H., Millard, P.H.: Modelling risk of readmission with phase-type distribution and transition models. *IMA Journal of Management Mathematics* (in press) (2008c)
33. Hosmer, D.W., Lemeshow, S.: Applied Logistic Regression, 2nd edn. Wiley, Chichester (2000)
34. R Development Core Team. R: A language and environment for statistical computing, R Foundation for Statistical Computing. Vienna, Austria (2005)
35. Berman, F., Fox, G., Hey, A.: Grid computing: making the global infrastructure a reality. Wiley, Chichester (2003)
36. Kacsuk, P., Farkas, Z., Sipos, G., Hermann, G., Kiss, T.: Supporting workflow-level ps applications by the P-GRADE Grid portal. In: Proc. of the CoreGRID Symposium, pp. 253–263 (2007)