Health Care Management Science 7, 285–289, 2004

Using Coxian Phase-Type Distributions to Identify Patient Characteristics for Duration of Stay in Hospital

ADELE H. MARSHALL *
Department of Applied Mathematics and Theoretical Physics, David Bates Building, Queen's University of Belfast, Belfast, Northern Ireland, BT7 1NN, UK
E-mail: a.h.marshall@qub.ac.uk

SALLY I. McCLEAN
School of Computing and Information Engineering, Faculty of Informatics, University of Ulster, UK

Abstract. Coxian phase-type distributions are a special type of Markov model that describes duration until an event occurs in terms of a process consisting of a sequence of latent phases. This paper considers the use of Coxian phase-type distributions for modelling patient duration of stay for the elderly in hospital and investigates the potential for using the resulting distribution as a classifying variable to identify common characteristics between different groups of patients according to their (anticipated) length of stay in hospital. The identification of common characteristics for patient length of stay groups would offer hospital managers and clinicians possible insights into the overall management and bed allocation of the hospital wards.

Keywords: stochastic modelling, Coxian phase-type distributions, Markov models, survival analysis, geriatric medicine

1. Introduction

The current focus of many health care providers is the development of accurate health care models to assist with resource allocation. One of the most demanding of all the specialist areas is the care of the elderly. This is largely due to the dramatic increase in the proportion of elderly in the population which is expected to continue to rise for at least the next 50 years. The population projections for 2050 indicate a pronounced shift in the age distribution toward older age groups which is largely due to increasing longevity, higher birth rates following the post World War II baby boom, and the more recent decline in fertility rates [1]. Medical resources throughout the world are feeling the added strain of this increasing proportion of elderly in the population as the cost of health care of the elderly becomes an even more significant part of overall expenditure [2].

The effective care of elderly patients in hospitals may be enhanced by accurately modelling the length of stay of the patients in hospital and the associated costs involved. There has been substantial research carried out in the past concerning the modelling of the length of stay of patients in hospital. More recently, Coxian phase-type distributions have been successfully applied to modelling patient duration of stay in hospital. Previous research by Sorensen [3] has also highlighted the use of models to represent patient resource demands. Sorensen divides the hospital admissions into different groups depending on the length of time the patients have been in hospital. Each group consists of patients who require similar care levels and similar resource demands.

Neuts [4] described phase-type distributions as the time to absorption of a finite Markov chain in continuous time, where there is a single absorbing state and the stochastic process starts in a transient state. Phase-type distributions are considered highly versatile with many advantages. They may be used to fit a distribution to a series of statistical data according to the moments of that series or can be generalised to include almost all continuous distributions [5] such as the exponential, which will only have one phase, or the Erlang and mixed exponential distributions. In fact, the phase-type models were originally introduced as a natural probabilistic generalisation of Erlang distributions. The key difference is that the movement between all the transient stages and the absorbing phase can occur in the phase-type distribution whereas in the case of the Erlang transitions movement can only occur between sequential phases. In addition, the distributions have the ability to describe detailed information about the behaviour of the stochastic models while also allowing the *lack of memory* property to exist. Furthermore, in many situations the distributions replace cumbersome numerical integrations with more manageable matrix calculations whose numerical implementation is easy and inexpensive [6].

There are many examples in the literature where phase-type distributions are being used, not only in the applied probability domain but also as a tool for data analysis. Applications are wide ranging from calculating the expected load of mobile phone networks [7], to analysing the duration of stay of elderly patients in hospital [8]. Faddy and McClean [9] used this model to find a suitable distribution for the duration of stay of a group of male geriatric patients in hospital. They found that the phase-type distributions were ideal for measuring the lengths of stay of patients in hospital and showed how it was also possible to consider other variables that may influence duration. Faddy [5] also provides some other useful illustrations of phase-type distributions.

* Corresponding author.

This paper considers the use of Coxian phase-type distributions within the care of the elderly using a data set different to that used by Faddy and McClean [9]. The resulting distribution is further investigated for its potential as a classifying variable to identify common characteristics for groups of patients according to their (anticipated) length of stay in hospital. Apart from Sorensen's work, there has been very little research concerned with the grouping of such patient length of stay. The identification of common characteristics for patient length of stay groups would offer hospital managers and clinicians possible insights into the overall management and bed allocation of the hospital wards.

2. Coxian phase-type distributions

The generality of the phase-type distributions makes parameter estimation difficult. As a result, Coxian phase-type distributions were introduced to overcome such a problem by ensuring that the transient states (or phases) of the model are ordered. Coxian phase-type distributions [10] are employed to describe duration until an event occurs in terms of a process consisting of a sequence of latent phases. The process begins in the first phase and may either progress through the phases sequentially or enter into the absorbing state (the terminating event). Such phases may then be used to describe stages of a process which terminates at some point. For example, duration of stay in hospital can be thought of as a series of transitions through phases such as: acute illness, intervention, recovery or discharge. This may capture how a domain expert conceptualises the process. The phase-type distribution relates directly to survival analysis where the survivor function is the duration of time until a certain event takes place. The event could be leaving hospital due to transfer, discharge or death.

Cox and Miller [11] develop the theory of Markov chains such as those defined in this section. The Coxian phase-type distributions describe the probability $P(t)$ that the process is still active at time t, defined as follows. Let $\{X(t);\ t \geqslant 0\}$ be a (latent) Markov chain in continuous time with states $\{1, 2, \ldots, n, n+1\}$, $X(0) = 1$, and for $i = 1, 2, \ldots, n-1$, the probability that a patient will move from one phase to the next phase in the system, in the time interval δt may be written as

$$\text{prob}\big\{X(t+\delta t) = i+1 \mid X(t) = i\big\} = \lambda_i \delta t + o(\delta t), \quad (1)$$

and likewise for $i = 1, 2, \ldots, n$, the probability that a patient, during the time interval δt, will leave the system completely and enter the absorbing phase may be written as

$$\text{prob}\big\{X(t+\delta t) = n+1 \mid X(t) = i\big\} = \mu_i \delta t + o(\delta t) \quad (2)$$

with $\mu_1, \mu_2, \ldots, \mu_n$ representing the rates of movement or transitions of patients from any phase out of phases $Ph_1, \ldots, Ph_n$ to the absorbing phase Ph_{n+1}, and, $\lambda_1, \lambda_2, \ldots, \lambda_{n-1}$ the rates of movement from Ph_1 to Ph_2, Ph_2 to Ph_3 and Ph_{n-1} to Ph_n (figure 1).

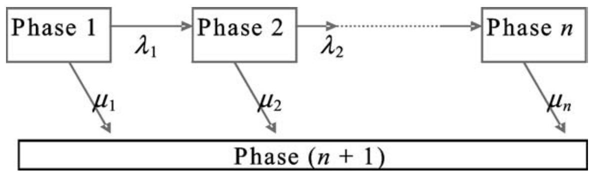

Figure 1. An illustration of coxian phase-type distributions.

The distributions can be represented in matrix notation where the probability density function of T is

$$f(t) = \boldsymbol{p}\exp\{\boldsymbol{Q}t\}\boldsymbol{q}, \quad (3)$$

$$\boldsymbol{p} = (1\,0\,0\ldots 0\,0), \quad (4)$$

$$\boldsymbol{q} = -\boldsymbol{Q}\mathbf{1} = (\mu_1\mu_2\ldots\mu_n)^{\mathrm{T}}, \quad (5)$$

and $\boldsymbol{Q}$ is the matrix of transition rates between states, as shown:

$$\boldsymbol{Q} = \begin{bmatrix} -(\lambda_1+\mu_1) & \lambda_1 & 0 & \ldots & 0 & 0 \\ 0 & -(\lambda_2+\mu_2) & \lambda_2 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -(\lambda_{n-1}+\mu_{n-1}) & \lambda_{n-1} \\ 0 & 0 & 0 & \ldots & 0 & -\mu_n \end{bmatrix}. \quad (6)$$

For example, when $i = 1$

$$\exp\{\boldsymbol{Q}t\} = \big[\mathrm{e}^{-\mu_1 t}\big], \quad \text{and} \quad f(t) = \mu_1\mathrm{e}^{-\mu_1 t}, \quad (7)$$

when $i = 2$

$$\exp\{\boldsymbol{Q}t\} = \begin{bmatrix} \mathrm{e}^{-(\lambda_1+\mu_1)t} & \big(\frac{\lambda_1}{\lambda_1+\mu_1-\mu_2}\big)\big(\mathrm{e}^{-\mu_2 t} - \mathrm{e}^{-(\lambda_1+\mu_1)t}\big) \\ 0 & \mathrm{e}^{-\mu_2 t} \end{bmatrix}, \quad (8)$$

and

$$f(t) = \frac{(\lambda_1+\mu_1)(\mu_1-\mu_2)}{\lambda_1+\mu_1-\mu_2}\mathrm{e}^{-(\lambda_1+\mu_1)t} + \frac{\mu_2\lambda_1}{\lambda_1+\mu_1-\mu_2}\mathrm{e}^{-\mu_2 t},$$

and so on, with the λ_i's and μ_i's defined in equations (1) and (2).

The survival probability that $X(t) = 1, 2, \ldots, n$ is given by

$$S(t) = \boldsymbol{p}\exp\{\boldsymbol{Q}t\}\mathbf{1}. \quad (9)$$

3. Modelling patient duration of stay

The management of the care of elderly patients in hospitals may be improved if there was a model to represent and predict the length of stay of the patients in hospital. For instance, if a hospital manager were able to estimate the duration of stay of a patient on admission to hospital, the ward could be more efficiently managed with better allocation of beds and resources.

The distribution of the duration of stay of elderly patients in hospital tends to be highly skewed in nature where there is usually a large peak in the distribution at the start which then gradually tails off as duration increases. Millard et al.

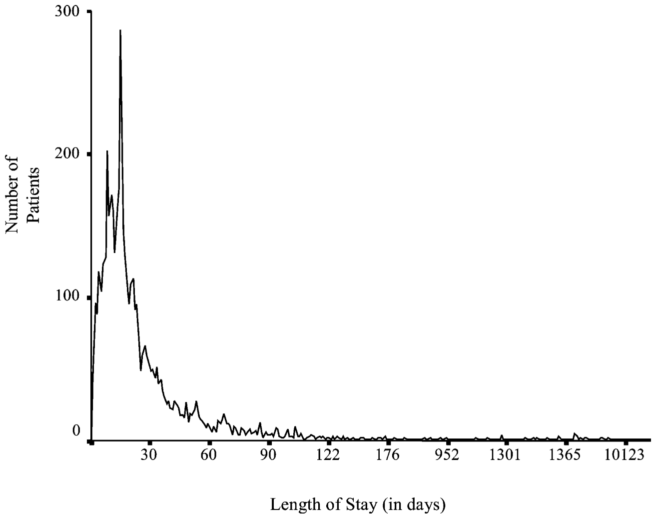

Figure 2. Distribution of duration of stay of geriatric patients in St. George's Hospital, London (1994–1997).

[12] have shown that a hospital's expenditure may be greatly influenced by those patients in the long tail of the distribution who stay in hospital for a long period of time. Sometimes this may not be regarded as a serious issue as the longer stay patients do not require the same amount of resources daily as those of shorter more acute stay patients and they are fewer in number. However, over time these elderly patients can consume significant quantities of medical resources. The modelling and prediction of patient destination and duration of stay in a geriatric department would therefore be beneficial to hospital managers to help estimate current patient requirements and predict future needs.

Past investigations of the elderly patients' length of stay led to the discovery that a two-term mixed exponential model produces a good representation of patient survival. Since then further research has endeavoured to improve the mixed exponential models with the incorporation of more complex compartmental systems, and more sophisticated stochastic models such as the Coxian phase-type distributions [1]. Alternative modelling techniques for representing elderly patient length of stay are fully discussed in [13].

The Clinics data set contains information collected during 1994–1997 to assist in the management of patients in a department of geriatric medicine. There are 4722 patient records comprising variables related to personal information such as age, gender, marital status, next of kin, lives alone; admission reasons such as stroke, fall, confusion; Barthel scores and, outcome details; destination on departure from hospital and duration of stay in hospital. The mean length of stay in hospital is 85 days while the median is 17 days indicating a skewed distribution; the distribution is illustrated in figure 2. This is confirmed by examining the shape of the distribution of length of stay which is highly skewed consisting of a large peak at the start of the distribution that gradually tails off as duration increases. As such this skewed nature makes management of patients more difficult. The development of models for the duration of stay of elderly patients in hospital would therefore assist hospital managers in the estimation of current patient requirements and the prediction of future needs. Marshall et al. [14] provide information on the preliminary analysis of the Clinics data set.

This paper utilises the Clinics data set to demonstrate how the patient duration of stay can be modelled using a suitable Coxian phase-type distribution. Following this discussion, the paper examines a further use of the model whereby the patients within the length of stay groups are examined to identify common characteristics. The identification of such common characteristics between patient length of stay groups would assist clinicians understanding of patient duration of stay in hospital.

4. Methodology

The length of stay of elderly patients in hospital can be modelled using Coxian phase-type distributions. The procedure adopted is sequential in nature whereby increasing numbers of n phases are tried, starting with $n = 1$ (corresponding to the exponential distribution), until there is very little improvement to the fit from adding an additional phase. The approach is implemented by using the following likelihood function:

$$L = \sum_{i=1}^{n} \log\big(\boldsymbol{p} \exp\{\boldsymbol{Q} t_i\} \boldsymbol{q}\big) \tag{10}$$

for the phase-type distribution in a series of likelihood ratio tests. The Nelder–Mead algorithm [15] is implemented using MATLAB software [16] to perform the likelihood ratio tests which determine the most suitable number of phases in the distribution. The Nelder–Mead algorithm [15] is a non-gradient approach which uses a simplex formed by a set of $(n + 1)$ mutually equidistant points in n dimensional space. The method compares the values of the function at the $(n + 1)$ vertices using the simplex which it then guides towards the optimum point during the iterative process. The three basic operations used to direct the simplex are reflection, expansion and contraction. The approach is considered a very robust, powerful technique provided the number of variables is not very large.

The following formula is derived in order to represent the length of stay in terms of k phases. Let π_i be the probability that an individual departs the system from Ph_i. This can be calculated by taking the probability density formula for each phase or state. Then

$$\pi_1 = \int_0^\infty \mu_1 \mathrm{e}^{-(\lambda_1+\mu_1)t}\,\mathrm{d}t = \frac{\mu_1}{\lambda_1 + \mu_1}. \tag{11}$$

Similarly,

$$\begin{aligned} \pi_2 &= \int_0^\infty \mu_2 \mathrm{e}^{-(\lambda_2+\mu_2)t}\,\mathrm{d}t \int_0^\infty \lambda_1 \mathrm{e}^{-(\lambda_1+\mu_1)t}\,\mathrm{d}t \\ &= \left(\frac{\lambda_1}{\lambda_1 + \mu_1}\right)\left(\frac{\mu_2}{\lambda_2 + \mu_2}\right), \\ &\vdots \end{aligned} \tag{12}$$

$$\pi_k = \left(\frac{\lambda_1}{\lambda_1 + \mu_1}\right)\left(\frac{\lambda_2}{\lambda_2 + \mu_2}\right)\cdots\left(\frac{\lambda_{k-1}}{\lambda_{k-1} + \mu_{k-1}}\right). \quad (13)$$

Patients may then be divided into groups according to their length of stay where the data is grouped in the ratio $\pi_1 : \pi_2 : \cdots : \pi_k$. In general the kth length of stay group S_k can be determined by the following equation:

$$S_k = \left\{x^{(j)}\colon m\sum_{i=1}^{k-1}\pi_i < j \leqslant m\sum_{i=1}^{k}\pi_i\right\}, \quad \text{for } k = 1, \ldots, n, \quad (14)$$

where $x^{(1)}, \ldots, x^{(m)}$ represents the ordered length of stay data for each patient and m represents the number of patients in the data set. The patients' details within each length of stay group may then be examined to determine if they have any common characteristics.

5. Results

The following table (table 1) of results was produced by using the previously described Nelder–Mead algorithm and likelihood ratio tests. It is apparent from inspection of the likelihood values, that there is no (significant) improvement in fit by adding the fourth phase to the distribution. Therefore a three-phase Coxian distribution is considered to most suitably represent length of stay of geriatric patients in hospital in the Clinics data set. The estimates for the parameters of the three phase distribution are $\hat{\mu}_1 = 0.0394$, $\hat{\mu}_2 = 0.0011$, $\hat{\mu}_3 = 0.0002$, $\hat{\lambda}_1 = 0.0011$, $\hat{\lambda}_2 = 1.56 \times 10^{-5}$.

The data may then be divided into the ratio $\pi_1 : \pi_2 : \cdots : \pi_k$ where π_i are estimated from equation (13) thus providing the following three patient length of stay groups as in (14):

$$\begin{aligned} S_1 &= \left\{x^{(j)}\colon j \leqslant m\pi_1\right\}, \\ S_2 &= \left\{x^{(j)}\colon m\pi_1 < j \leqslant m(\pi_1 + \pi_2)\right\}, \\ S_3 &= \left\{x^{(j)}\colon j > m(\pi_1 + \pi_2)\right\}, \end{aligned} \quad (15)$$

where $x^{(1)}, \ldots, x^{(m)}$ represents the ordered length of stay data for each patient and $m = 4722$, the number of patients in the data set, and, $\pi_1 = 0.9630$, $\pi_2 = 0.0364$ and $\pi_3 = 0.0005$. Upon fitting the Coxian phase-type distributions, the patients may be regarded as belonging to one of three 'length of stay' groups: 0–130 days, 131–1450 days, >1450 days. By performing statistical analyses on the patient details for each group, comparisons can be made and common characteristics identified.

Table 1
Results of fitting the coxian phase-type distributions.

Log-likelihood	Estimation of parameters
$n = 1$, $L = 24260$	$\hat{\mu}_1 = 0.0160$
$n = 1$, $L = 21100$	$\hat{\mu}_1 = 0.0393$, $\hat{\mu}_2 = 0.0009$, $\hat{\lambda}_1 = 0.0015$
$n = 3$,[a] $L = 21097$	$\hat{\mu}_1 = 0.0394$, $\hat{\mu}_2 = 0.0011$, $\hat{\mu}_3 = 0.0002$, $\hat{\lambda}_1 = 0.0011$, $\hat{\lambda}_2 = 1.56 \times 10^{-5}$
$n = 4$, $L = 21097$	$\hat{\mu}_1 = 0.0394$, $\hat{\mu}_2 = 0.0011$, $\hat{\mu}_3 = 0.0002$, $\hat{\mu}_4 = 2.51 \times 10^{-6}$ $\hat{\lambda}_1 = 0.0015$, $\hat{\lambda}_2 = 1.57 \times 10^{-5}$, $\hat{\lambda}_3 = 8.51 \times 10^{-11}$

[a]Indicates the most suitable number of phases for the Clinics data set.

The first group of patients (length of stay group 1: 0–130 days) consists of the majority of the elderly patients and as a result includes a diverse range of patient attributes that have few common characteristics. However, if the second group of patients (length of stay group 2: 131–1450 days) are considered and compared with the first, there are distinguishable differences between the two patient groups in particular, in association with the admission reasons, admission month, Barthel (dependency) score, and outcome (destination) on departure from hospital. Although there are few patients in the third group of patients (length of stay group 3: >1450 days) close inspection of the group indicates common characteristics for instance, all the patients are female, aged between 57 and 64 (where the age range for the full Clinics data set is between 42 and 105). In fact, it would be possible to consider these patients as outliers who although few in number consume a considerable amount of resources due to the very long lengths of stay in hospital. More focus can then be directed to the other two groups of patient length of stay.

The results of the fitted Coxian phase-type distribution relate directly to previous work carried out on compartmental models [17]. The first phase in the length of distribution (0–130 days) can be considered the shorter stay patients mainly comprising of those patients who are discharged home or in such an acute state on arrival to hospital that they die after a short period of time. The second phase (131–1450 days) would be considered those patients who have a much longer duration of stay in hospital and will mainly comprise of the patients who eventually transfer, e.g., to nursing home care. The third and final phase (>1450 days) is regarded as the patients with a very long, extreme length of stay in hospital. Quite often it is these few cases that will consume a large proportion of the resources.

The Coxian phases have highlighted a strong relationship between the patient outcome showing three stages of patient behaviour reflected by whether the patient transfers to another ward, is discharged back into the community or dies during their stay in hospital. This agrees with previous work carried out using Conditional phase-type distributions [18] and other review papers which have highlighted the appeal of using the Coxian phase-type distribution by making comparisons between standard techniques for modelling length of stay of patients in hospital and the Coxian phase-type distribution [19].

6. Conclusion and further work

This paper discusses Coxian phase-type distributions and their use in modelling patient duration of stay in hospital. In particular the length of stay of elderly patients were analysed and found to be most suitably represented by a three-phase

Coxian distribution. The model was used to identify common characteristics of patients in the same length of stay group. Such analysis could be expanded to further investigate any potential relationship in patient variables and the length of stay category to which they belong. It is also hoped that further work will involve the application of this method on an extensive data set incorporating additional clinical variables. Additionally, there is the possibility to expand the current model to incorporate other information such as the influence that the doctor's decision may have on the overall length of stay of the patients in hospital. However, if the medical information is accurate and appropriately incorporated in the model, it is more likely that the most influential variables on patient duration of stay will be the patient's medical condition.

The modelling technique may be used to represent a group of patients according to their length of stay and the characteristics that they have in common. Such information could then be utilised to give an indication to hospital staff the likely length of stay category and destination of future patients. This would be most useful, if the model were extended to include more patient information, to assess the overall activity in the hospital ward by providing an estimate of the number of bed days required for the current group of patient and the resulting availability of beds for new admissions.

Future work may also involve the development a Coxian phase-type model that will take account of the costs attached to the various phases or groups of patient. The resulting model would assist hospital managers with bed allocation whereby various scenarios may be investigated to identify the most cost effective case-mix.

References

[1] Population projections for the United Kingdom, Government Actuary's Department, http://www.gad.gov.uk/

[2] K. Kinsella and V.A. Velkoff, An Ageing World US Census Bureau, Series P95/01-1, US Government Printing Office, Washington, DC (2001).

[3] J. Sorensen, Multi-phased bed modelling, Health Services Management Research 9 (1996) 61–67.

[4] M.F. Neuts, *Structured Stochastic Matrices of M/G/1 Type and Their Application* (Marcel Dekker, New York, 1989).

[5] M.J. Faddy, Examples of fitting structured phase-type distributions, Applied Stochastic Models and Data Analysis 10 (1994) 247–255.

[6] M.F. Neuts, *Matrix-Geometric Solutions in Stochastic Models – An Algorithmic Approach* (John Hopkins University Press, 1981).

[7] P. Fazekas, S. Imre and M. Telek, Modeling and analysis of broadband cellular networks with multimedia connections, Telecommunications Systems 19 (3–4), (2002) 263–288.

[8] M.J. Faddy, On inferring the number of phases in a coxian phase-type distribution, Communications of the Statistician – Stochastic Models 14(1–2) (1998) 407–417.

[9] M.J. Faddy and S.I. McClean, Analysing data on lengths of stay of hospital patients using phase-type distributions, Applied Stochastic Models and Data Analysis (2000).

[10] D.R. Cox, A use of complex probabilities in the theory of stochastic processes, Proceedings of the Camb. Phil. Soc. 51 (1955) 313–319.

[11] D.R. Cox and H.D. Miller, *The Theory of Stochastic Processes* (Methuen, London, 1965).

[12] P.H. Millard, G. Christodoulou and S.I. McClean, Survival in long-term care – discussion document on the interactions and costs between health and social care, submitted to the Royal Commission on Long Term Care, Department of Geriatric Medicine, St George's Hospital Medical School, London (1998) 1–30.

[13] A.H. Marshall and S.I. McClean, Conditional phase-type distributions for modelling patient length of stay in hospital, International Transactions in Operational Research 10 (2003) 565–576.

[14] A.H. Marshall, S.I. McClean, C.M. Shapcott, I.R Hastie and P.H. Millard, Developing a bayesian belief network for the management of geriatric hospital care, Health Care Management Science Journal 4 (2001) 23–28.

[15] J.A. Nelder and R.A Mead, Simplex method for function minimization, Computer Journal 7 (1965) 308–313.

[16] MATLAB® Reference Guide, The MathsWorks Inc., Natick, MA (1992).

[17] G.W. Harrison, Compartmental Models of Patient Occupancy Patterns, in: *Modelling Hospital Resource Use: A Different Approach to the Planning and Control of Health Care Systems*, eds. P.H. Millard and S.I. McClean (Royal Society of Medicine Press, 1994).

[18] A.H. Marshall, S.I. McClean, C.M. Shapcott and P.H. Millard, Modelling patient duration of stay to facilitate resource management of geriatric hospitals, Health Care Management Science Journal 5 (2002) 313–319.

[19] C. Vasilakis and A.H. Marshall, Modelling nationwide hospital length of stay: Opening the black box, Journal of the Operational Research Society Practice Notes (2004).