A model-based approach to the analysis of patterns of length of stay in institutional long-term care.

Haifeng Xie
Thierry J. Chaussalet
Peter H. Millard

School of Informatics

A Model-Based Approach to the Analysis of Patterns of Length of Stay in Institutional Long-Term Care

Haifeng Xie, Thierry J. Chaussalet, and Peter H. Millard

***Abstract*—Understanding the pattern of length of stay in institutional long-term care has important practical implications in the management of long-term care. Furthermore, residents' attributes are believed to have significant effects on these patterns. In this paper, we present a model-based approach to extract, from a routinely gathered administrative social care dataset, high-level length-of-stay patterns of residents in long-term care. This approach extends previous work by the authors to incorporate residents' features. Two applications using data provided by a local authority in England are presented to demonstrate the potential use of this approach.**

***Index Terms*—Covariate, length-of-stay analysis, long-term care, Markov model, survival analysis.**

I. Introduction

ELDERLY people often experience a decline in physical and mental health, and activities of daily living; for example, feeding, toileting, and self care can become difficult. In cases when they can no longer be cared for at home, assistance is required from agencies providing long-term care (LTC). In England, LTC consists mostly of social service programs run by local authorities, and residential care (RC) and nursing care (NC) provided in institutional care homes. In general, RC consists of board and personal care for those who are frail but still able to manage their activities of daily living, while NC is for elderly people who are medically stable but have a greater degree of physical and mental disability and require input from a NHS (National Health Service) registered nurse.

Local authorities in England play an important role in the delivery of LTC. Under the NHS and Community Care Act 1990 [1] and the Care Standard Act 2000 [2], local authorities are responsible for the placement and finance of adults in publicly funded RC and NC that conforms to national standards. Furthermore, the government considers that the ability to discharge elderly patients to institutional LTC (ILTC) is essential to the planning and running of acute hospital care. The introduction of the Community Care (Delayed Discharges, etc.) Act 2003 in January 2004 means local authorities are facing financial penalties for failing to provide vacancies in institutional care homes for hospital discharges [3]. Therefore, local authorities have a keen interest in gaining a better understanding of the behavior of the LTC system, and of issues relating to the pattern of residents' length of stay (LOS) in LTC in particular.

To address the concerns of local authorities, we focus on publicly funded residents in ILTC since they constitute a large proportion of admissions to LTC in England [4], and hence represent a major consumption of a local authority's LTC resources. National longitudinal surveys in England have shown that most of the publicly funded residents who are admitted to ILTC are there on a permanent basis [4], [5]. For these residents, discharge from ILTC is predominantly by death; very rarely residents are discharged to the community (i.e., well enough to be maintained in their own homes); and discharge to a hospital usually means terminal care [5]. The surveys also found that about 20% of the residents admitted to RC would be transferred to NC sometime during their stay in ILTC; however, movements from NC to RC rarely occur. Furthermore, due to their obligations to use public funds for the purchase of RC and NC, most local authorities have organized means of determining suitable care placements for these residents. Therefore, these admissions usually reflect the physical conditions and needs of the residents.

For administrative purposes, local authorities routinely collect data about residents under their care such as date of admission, place of admission, date and place of transfer, and date of discharge if applicable. Simple analysis of these data, such as calculating the average LOS in care, are often performed in attempt to extract useful information to aid planning. However, due to the high degree of skewness that is common in this type of LOS data, the use of average LOS, a single number, to describe the overall LOS pattern gives incomplete and often misleading information [6]. Hence, in order to discover and understand the pattern of LOS for publicly funded residents in ILTC, a more sophisticated approach to the analysis of this type of administrative data is required.

Traditionally, the analysis of LOS data belongs to a branch of statistics called *Survival Analysis* [7], which typically uses LOS data as a vehicle to study the effect of different patient features on survival time. Whereas data mining approaches to the analysis of survival data; such as those based on decision-rules [8] and artificial neural networks [9], are usually concerned with predicting categorical survival outcomes for a given set of patient features (e.g., dead or alive at three years after treatment). For these methods, the underlying random process that generates the observed LOS data is usually modeled implicitly. Other methods, such as those based on flow modeling [6] and stochastic processes [10], [11], concentrate on modeling the underlying process explicitly and aim to capture high level patterns of LOS.

Manuscript received August 11, 2005; revised November 30, 2005.

The authors are with the Health and Social Care Modeling Group (HSCMG), Department of Information Systems, Cavendish School of Computer Science, University of Westminster, London W1W 6UW, U.K. (e-mail: xieh@wmin.ac.uk; chausst@wmin.ac.uk; phmillard@tiscali.co.uk).

Digital Object Identifier 10.1109/TITB.2005.863820

In this paper, we extend a modeling approach previously reported in [11] to study the pattern of LOS for publicly funded residents in ILTC, and to take into account different types of observation (e.g., censored and truncated) and the attributes of individual residents (e.g., age and gender). We present the findings of applying this modeling approach to data provided by an English local authority, and their practical implications are discussed.

II. Data

The Housing and Social Service Department of the London Borough of Merton (U.K.), like many local authorities in England, has been working hard toward the better management and delivery of their LTC system. For financial reporting purposes, Merton has electronic spreadsheet files containing records of annual payments made for publicly funded placements in RC and NC. Information such as a resident's unique identification number, gender, age at admission, the name of the care home and the corresponding starting date and, if appropriate, the ending date and the reason for discharge (e.g., by death, or transfer to another home with the same type of care or different type of care) are recorded.

For the purpose of this study, a cohort dataset containing information for publicly funded residents who were present in ILTC between 1 April 1997 and 1 April 2001 was created from an anonymized dataset provided by Merton. This dataset captured the type of care and the corresponding duration of the care that was received for each resident through their stay in the ILTC system. For example, a typical observation could be that a resident was admitted to RC, where he/she stayed for a couple of years before being transferred to NC, where he/she stayed for another year before being discharged by death.

The dataset contains 1244 records for 1071 publicly funded residents. On 1 April 1997, there were 282 residents (138 in RC and 144 in NC) in ILTC. During the four-year period, there were 889 new admissions (438 to RC and 451 to NC), and 73 RC residents were transferred to NC. On 1 April 2001, there were 436 residents in care (250 in RC and 186 in NC).

There are three different types of observation in the dataset. Residents present in the system on 1 April 1997 have *left-truncated* observations; i.e., we have no information on those who joined the system with them at the same time but discharged before the starting date of the dataset. Residents present on 1 April 2001, the ending date of the dataset, as well as those who are "lost" during observation (for example, move to another local authority) have *right-censored* observations; i.e., we only know that they have been in the system for a period of time but do not know when they will be discharged in the future. We have *complete* observations for those residents who joined and were discharged during the period. These different types of observation are illustrated in Fig. 1, in which case 1 is a complete observation; cases 2 and 3 are right-censored observations; case 4 is left-truncated; and cases 5 and 6 are both left-truncated and right-censored observations. These observations should be treated appropriately when analyzing the data [12]. This is especially true for left-truncated observations since the longest observation we have will only be as long as the data observation window, which will lead to underestimating the survival time of those who stay a very long time in the system.

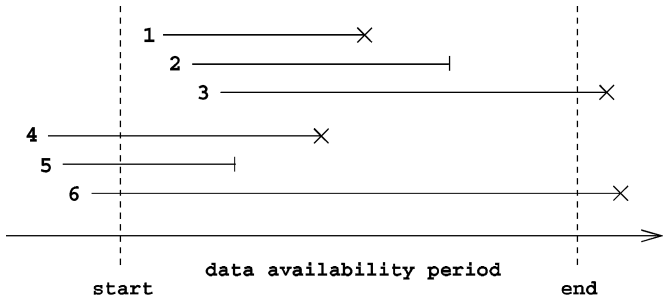


Fig. 1. Different types of observation in dataset. "×" represents discharge, and "|" represents observation is "lost".

III. Method

A. Modeling the Movement of Elderly People in ILTC

Research in the U.K. has shown that the mortality rate for publicly funded residents in NC is particularly high in the first few months, and then gradually levels out [5], [13]. This observation supports the notion of different phases in the stay of residents in care homes. In a previous study [11], we presented a continuous-time Markov model for the flow of elderly residents within and between RC and NC. The model, shown in Fig. 2, uses a combination of states (e.g., a short-stay state and a long-stay state) to capture the flow of residents through each type of care. For instance, a person admitted to RC might stay for a short period of time, then is either discharged (predominantly by death) or transferred to NC; or settle down and become a long-stay resident in RC. Eventually the person is either discharged or transferred to NC for further stay. The rare movements such as discharge home and transfer from NC to RC are not considered in this model. The structure of the model is relatively simple and conceptually intuitive to many practitioners working in the LTC system.

In this model, the actual state spaces (S_1, S_2, etc.) of the Markov model are aggregated to form super-states called *classes* [14]. There are three classes in this model, namely RC, NC, and Discharge; and we observe these classes instead of the actual states. For example, we observe that a person is in RC, but do not know whether the person is in the short-stay state (S_1) or in the long-stay state (S_2).

Given the structure of the Markov model, the generator matrix $\mathbf{Q}$ is written as

$$\mathbf{Q} = \left[\begin{array}{cc:c:c:c} q_{11} & q_{12} & q_{13} & 0 & q_{15} \\ 0 & q_{22} & q_{23} & 0 & q_{25} \\ \hdashline 0 & 0 & q_{33} & q_{34} & q_{35} \\ 0 & 0 & 0 & q_{44} & q_{45} \\ \hdashline 0 & 0 & 0 & 0 & 0 \end{array}\right] \tag{1}$$

where q_{ij} is the instantaneous transition rate between state i and state $j (i \neq j)$, and the diagonal elements $q_{ii} = -\sum_{j \neq i} q_{ij}$ are defined so that row sums are zero. This matrix is partitioned

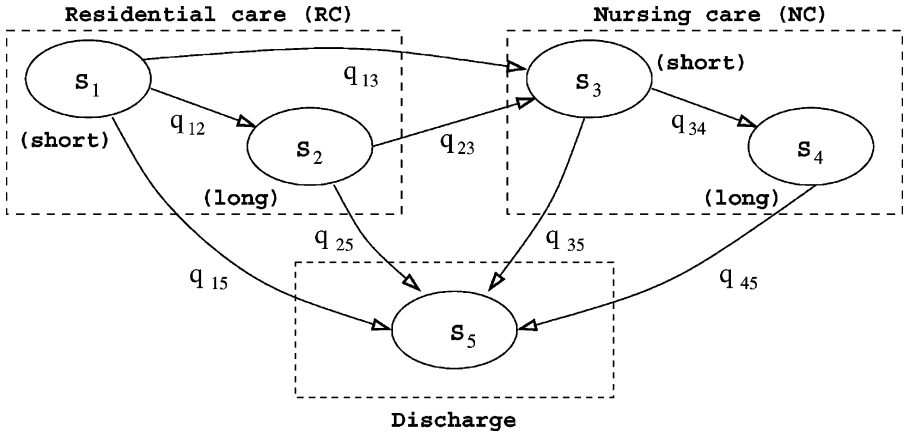


Fig. 2. Markov model for the movement of publicly funded residents in ILTC.

according to the class structure of the model as

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{RR} & \mathbf{Q}_{RN} & \mathbf{Q}_{RD} \\ \mathbf{0} & \mathbf{Q}_{NN} & \mathbf{Q}_{ND} \\ \mathbf{0} & \mathbf{0} & 0 \end{bmatrix} \tag{2}$$

where the submatrices correspond to the block matrices delimited by dotted lines in (1), and the subscripts R, N, and D represent the classes RC, NC, and Discharge, respectively.

Given the structure of the states within a class, the LOS in both RC and NC follow Coxian distributions [15] with probability density function of the form, say, for class A [16],

$$f_A(t) = -\boldsymbol{\phi}_A^{\mathrm{T}} \exp(\mathbf{Q}_{AA}t)\mathbf{Q}_{AA}\mathbf{1} \tag{3}$$

where $\boldsymbol{\phi}_A$ is a column vector of the probabilities of entering class A via each of its member states. For instance, for the model depicted in Fig. 2, $\boldsymbol{\phi}_R = (1, 0)^{\mathrm{T}}$ for RC. Throughout, we use $\boldsymbol{M}^{\mathrm{T}}$ to denote the transpose, and $\boldsymbol{M}^{-1}$ the inverse of a matrix $\boldsymbol{M}$.

The dataset provided by Merton captures, for each resident, their pathway and corresponding duration in each class visited. Formally, let the random variables $\boldsymbol{C}$ represent the sequence of classes a resident visited, and $\boldsymbol{T}$ the corresponding duration in each of the classes in $\boldsymbol{C}$. In this case, the domain of the elements in $\boldsymbol{C}$ is the set $\{R, N, D\}$. Therefore, an observation for a resident who was admitted to RC, where he stayed for a time t_1 before transferring to NC where he has been staying for t_2 so far, would be represented as $\boldsymbol{c} = (R, N)$ and $\boldsymbol{t} = (t_1, t_2)$. However, if that resident died after spending t_2 in NC, the observation would be represented as $\boldsymbol{c} = (R, N, D)$ and $\boldsymbol{t} = (t_1, t_2, 0)$. Throughout this paper, we use the convention that the duration in Discharge is zero, or in other words, a process stops upon reaching the absorbing state.

In general, the likelihood of observing $\boldsymbol{c}$ and $\boldsymbol{t}$ given the Markov model is

$$\begin{aligned} l(\boldsymbol{\theta}|\boldsymbol{c}, \boldsymbol{t}) = {} & \frac{\boldsymbol{\phi}_{c_1}^{\mathrm{T}} \exp(Q_{c_1c_1}t_1)\boldsymbol{Q}_{c_1c_2}}{\frac{\boldsymbol{\phi}_{c_1}^{\mathrm{T}} \exp(\boldsymbol{Q}_{c_1c_1}t_1^*)\boldsymbol{Q}_{c_1c_1}^{-1}\boldsymbol{Q}_{c_1c_2}}{\boldsymbol{\phi}_{c_1}^{\mathrm{T}}\boldsymbol{Q}_{c_1c_1}^{-1}\boldsymbol{Q}_{c_1c_2}}} \\ & \times \left\{ \prod_{j=2}^{r-1} \exp(\mathbf{Q}_{c_jc_j}t_j)\mathbf{Q}_{c_jc_{j+1}} \right\} \\ & \times \exp(\mathbf{Q}_{c_rc_r}t_r)\mathbf{1} \end{aligned} \tag{4}$$

where $\boldsymbol{\theta}$ denotes the set of parameters, here the transition rates q_{ij}, of the Markov model; r is the number of classes in $\boldsymbol{c}$; and t_1^* denotes the time spent in class c_1 (the first element of $\boldsymbol{c}$) prior to 1 April 1997, the starting date of the dataset. Thus $t_1^* = 0$ if a resident were admitted after 1 April 1997. Here, the division of vectors is defined element-wise [17]. The first term in (4) handles the left-truncated part $(t_1^* > 0)$ of an observation, and is the probability density of staying in class c_1 for t_1 and jumping to c_2, conditional on the probability that the resident has been in c_1 for t_1^* already, and will eventually jump to c_2. The product term in the second line of (4) calculates the density of visiting all intermediate classes of an observation. The final term in (4) handles the right-censored part $(t_r > 0)$ of an observation. One special case is that a resident remains in the same type of care throughout the data availability period. In this case, the likelihood of such an observation is

$$l(\boldsymbol{\theta}|\boldsymbol{c}, \boldsymbol{t}) = \frac{\boldsymbol{\phi}_{c_1}^{\mathrm{T}} \exp(\boldsymbol{Q}_{c_1c_1}t_1)}{\boldsymbol{\phi}_{c_1}^{\mathrm{T}} \exp(\boldsymbol{Q}_{c_1c_1}t_1^*)}\mathbf{1}. \tag{5}$$

Therefore, for a dataset containing observations of n residents, the log-likelihood of observing these n pairs of $\boldsymbol{c}$ and $\boldsymbol{t}$ given the Markov model is

$$L(\boldsymbol{\theta}) = \log \left\{ \prod_{i=1}^{n} l(\boldsymbol{\theta}|\boldsymbol{c}_i, \boldsymbol{t}_i) \right\} = \sum_{i=1}^{n} \log\{l(\boldsymbol{\theta}|\boldsymbol{c}_i, \boldsymbol{t}_i)\}. \tag{6}$$

We fit the Markov model to observed data by choosing a set of values for $\boldsymbol{\theta}$ which maximize the log-likelihood function (6).

We follow the two stage procedure for fitting the Markov model to observational data, which was developed in [11]. Briefly, the first stage determines the structure of the model (the number of states in each class) by fitting Coxian distributions with increasing number of phases (or states) to observed LOS data in each class separately. The number of states is chosen based on the Bayesian information criterion (BIC) [18], which is a measure representing a compromise between model complexity and goodness-of-fit. Once the structure of the Markov model is specified, we proceed to fit the Markov model to the overall LOS data. All model fitting is conducted by maximizing the log-likelihood function using general purpose optimizers such as those available from MATLAB [19].

B. Incorporating Residents' Features

The Markov model described in the previous section assumes that all transition rates are constant for all residents. This can be restrictive, considering the vast diversity that exists among residents. Therefore, it is desirable to extend the model to take into account residents' features.

We assume there exist measurements on p features (also called attributes or covariates) $X_1, \ldots, X_p$ for each resident upon admission to each type of care. Then for the kth resident, with observation $\boldsymbol{x}_k = (1, x_{k1}, x_{k2}, \ldots, x_{kp})$, the transition rate between states i and $j (i \neq j)$, where transition from i to j is possible, can be written as

$$q_{ij,k} = \exp\left(\boldsymbol{\beta}_{ij}^{\mathrm{T}} \boldsymbol{x}_k\right) \tag{7}$$

where $\boldsymbol{\beta}_{ij}$ is a vector of coefficients. The exponential function ensures the constraint that all transition rates are positive. A feature that has significant effect on a transition rate (hence on LOS) will have its corresponding coefficient significantly different from zero. Under this formulation, the coefficients $\boldsymbol{\beta}_{ij}$ become the model parameters and are estimated by numerically maximizing the log-likelihood function (6).

This approach, which consists of incorporating residents' features into the Markov model via a log-linear function (7), is relatively straightforward, and is similar to that taken by [10]. However, the number of parameters to be estimated grows rapidly as the number of features increases.

IV. Results

In this section, we apply the modeling approach outlined in Section III to the dataset described in Section II to discover trends and differences in patterns of LOS for publicly funded residents in ILTC in Merton. We first apply the methodology to study the main pattern of LOS. Then, we demonstrate, with two examples, that the incorporation of residents' features can reveal further insights into LOS patterns.

A. LOS Pattern for Publicly Funded Residents in ILTC

The Markov model was fitted to the Merton dataset using the two stage model fitting procedure described previously. As found in [11], the first stage fitting suggested one state in RC and two states in NC, which means $q_{12} = q_{23} = q_{25} = 0$ in Fig. 2. The second stage fitting converged quickly and successfully. The estimated parameters are summarized in Table I. Comparing these results with those in [11], where left-truncated observations were not taken into account, the estimated values change only slightly; as expected, the incorporation of left-truncated observation decreases the estimated mortality rates q_{15} in RC and q_{35} and q_{45} in NC.

The results reveal some interesting findings on patterns of LOS for publicly funded residents in ILTC in Merton. A single state in RC indicates that residents depart from RC in a homogeneous fashion; i.e., there is a constant rate of departure from RC. The average LOS in RC is estimated to be $1/(q_{13} + q_{15}) = 973$ days (about 2.7 years). Upon leaving RC, about 78% (estimated by $q_{15}/(q_{13} + q_{15})$) of the residents will be discharged (permanently) and 22% will be transferred to NC. Two distinct states are suggested in NC: a short-stay state with an average LOS of 63 days and a long-stay state with an average LOS of 842 days (about 2.3 years). The rate of discharge from the short-stay state is about five times that from the long-stay state. This finding agrees with empirical observations that initial mortality is higher for the first few months following admission to NC [5], [13]. About 62% (estimated by $q_{34}/(q_{34} + q_{35})$) of the publicly funded residents entering NC will become long-stay residents. Furthermore, the average LOS in the long-stay state in NC is found to be similar to that of RC. This suggests that once NC residents pass through the short-stay state, their LOS pattern is similar to those in RC. Hence, the short-stay state may represent some sort of "settle-in" period for residents admitted to NC.

TABLE I
Estimated Parameters for the Markov Model With Left-Truncated Observations. The Numbers in Parentheses Are the Corresponding Estimates Without Considering Left-Truncated Observations Reported in [11]

Parameter	Estimate		Standard error	
q_{13}	0.000231	(0.000228)	0.000027	(0.000034)
q_{15}	0.000796	(0.000855)	0.000050	(0.000065)
q_{34}	0.009922	(0.010874)	0.002181	(0.002961)
q_{35}	0.006063	(0.006138)	0.000707	(0.000793)
q_{45}	0.001187	(0.001275)	0.000086	(0.000135)

We can graphically examine the pattern of LOS by plotting the survivor curve. In this case, a survivor curve shows the probability that a resident will remain in care after a period of time. Fig. 3 shows the survivor curves for LOS in RC and NC estimated by a Kaplan–Meier-type estimator [20] (jagged lines) with 95% confidence intervals (dotted lines), and those estimated by the fitted Markov model (smooth lines). The Kaplan–Meier estimator is a nonparametric maximum likelihood estimator of survivor probability [7]. Therefore, without making any parametric assumption about the underlying distribution of LOS and in the presence of censored data, the Kaplan–Meier estimated survivor curve (the jagged lines) can be regarded as representing the observed data. The close agreement between these survivor curves suggests that the fitted model captures the overall behavior of the LOS pattern. The initial sharp drop of the survivor curve for NC residents is captured by the use of a short-stay state in addition to a long-stay state in NC in the Markov model. This corresponds to the empirical observation that mortality is higher in the few months following admission to NC. It is also evident from the plots that after the initial drop, the survivor curves in RC and NC are almost parallel, suggesting that the LOS pattern for long-stay residents in NC is similar to those in RC.

B. Difference in LOS Pattern Among Residents in NC

The Markov property implies that the LOS pattern for residents admitted to NC directly is the same as for those transferred from RC. However, there might be reasons to believe that a person who spent a long time in RC might also stay for a long time in NC. In other words, there might be strong correlation

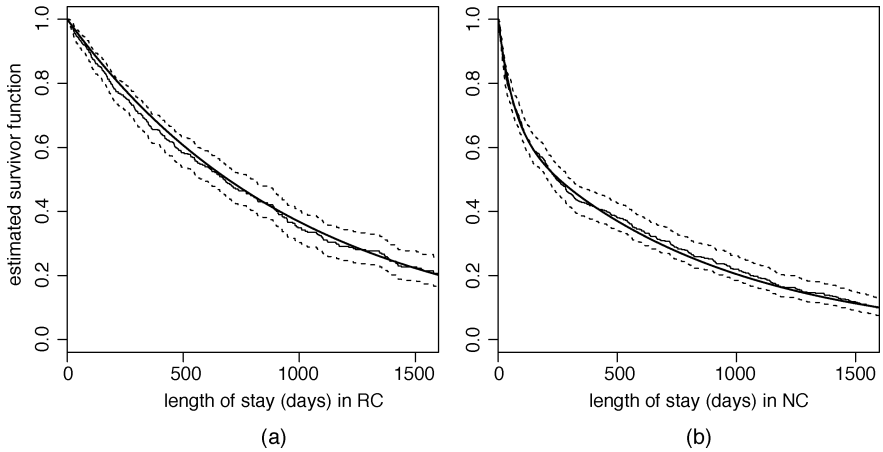


Fig. 3. Kaplan–Meier type estimated (jagged line) with 95% confidence interval (dotted lines), and Markov model fitted (smooth line) survivor curves of LOS, for (a) RC, and (b) NC for the Merton dataset.

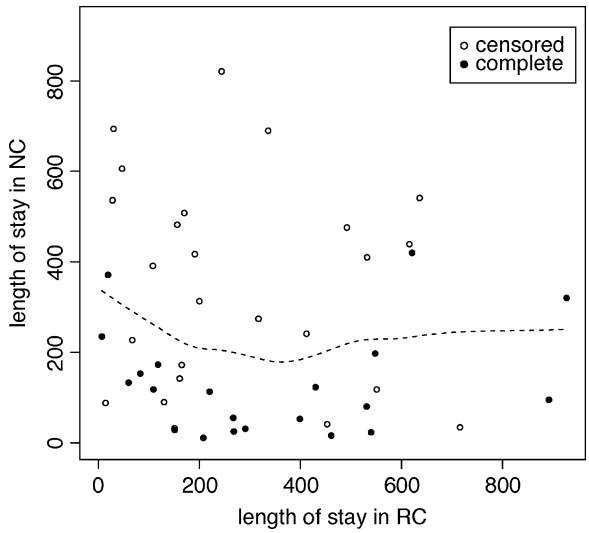


Fig. 4. Scatter plot of LOS in RC and NC among NC residents who were transferred from RC in the Merton dataset. Grouping is based on whether a resident was alive (right-censored observation) or had been discharged (complete observation) on 1 April 2001. The dotted lines correspond to the locally fitted smoothing curves.

between LOS in RC and the subsequent LOS in NC. To test this hypothesis, we use our model to investigate the possible differences in LOS pattern among NC residents in the Merton dataset; that is between those admitted to NC directly and those transferred from RC.

We treated LOS in RC (being zero for those who were admitted to NC directly) as a resident's feature. Two different ways of incorporating this feature into the model were considered, namely, Model 1: LOS in RC as a continuous variable, and Model 2: previous presence in RC as a binary variable, where "1" represents having stayed in RC prior to admission to NC. Therefore, including the base model (Model 0) that does not take into account previous LOS in RC, there are three models under consideration.

Comparisons between these (nested) models, based on the likelihood ratio test statistics, show that the incorporation of the feature provided no significant improvement over the base model, judging at 5% significance level (Model 0 versus Model 1: $X^2 = 5.36, p = 0.1473$; Model 0 versus Model 2: $X^2 = 4.60, p = 0.2035$). Therefore, the fitted result did not suggest a significant difference in LOS pattern between NC residents transferred from RC and those admitted directly. A scatter plot of the LOS data for NC residents who were transferred (see Fig. 4) shows no clear relationship between LOS in RC and subsequent LOS in NC. This finding is consistent with the result of a formal hypothesis test commonly used in the analysis of survival data; see, e.g., [7], (log-rank test, $W_L = 1.875, p = 0.1709$).

TABLE II
ESTIMATED PARAMETERS OF THE MODEL TO STUDY THE EFFECT OF GENDER ON PATTERN OF LOS IN ILTC

Parameter	Estimate	Standard error	95% Confidence interval
a_{13}	-8.622	0.189	(-8.992, -8.251)
a_{34}	-4.210	0.397	(-4.988, -3.433)
a_{15}	-7.120	0.079	(-7.275, -6.966)
a_{35}	-5.353	0.153	(-5.652, -5.054)
a_{45}	-6.711	0.110	(-6.927, -6.495)
b_{13}	0.827	0.302	(0.235, 1.419)
b_{34}	-0.991	0.428	(-1.831, -0.152)

C. Gender Difference in LOS Pattern

A recent national survey in care homes in England showed that the survival prospect was particularly poor for publicly funded male residents admitted to NC, especially in the first few months following admission [5]. We use our model to study whether there is a significant difference in LOS pattern between male and female residents in ILTC, and to quantify this difference if it exists.

We parameterized the model to incorporate gender as a feature. More specifically, the transition rate between state i and state j (where transition from state i to state j is possible) for the kth person is written as $q_{ij,k} = \exp(a_{ij} + b_{ij}x_k)$, where $x_k = 1$ if the person is male and $x_k = 0$ otherwise.

The model was fitted to the Merton dataset with the feature (gender) entering each transition rate in a sequential manner.

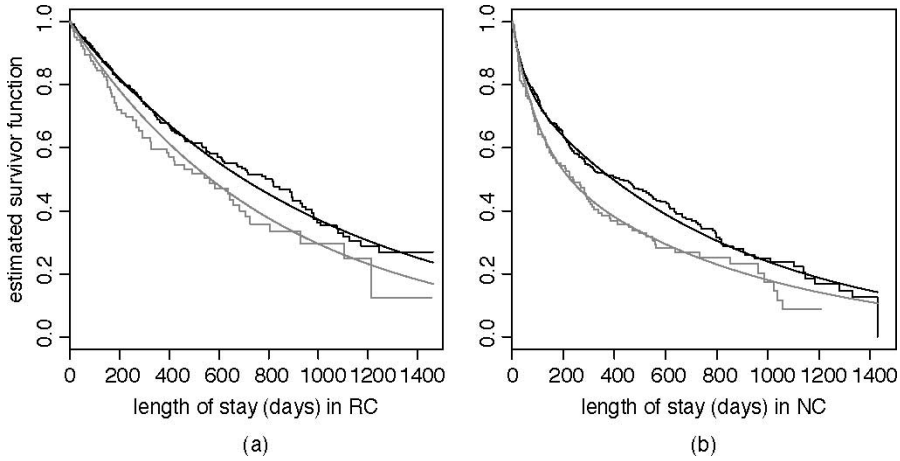


Fig. 5. Kaplan-Meier estimated (jagged line) and Markov model fitted (smooth line) survivor curves of LOS for female (black color) and male (grey color) residents in (a) RC and (b) NC.

The competing models were tested by the likelihood ratio test statistic. The final model showed that gender had a significant influence on transition rates q_{13} and q_{34} (see Table II).

A significant b_{13} suggests that there is significant difference in LOS pattern between male and female RC residents. The average LOS for male residents in RC is estimated to be 819 days or 2.2 years (estimated by $1/(e^{a_{13}+b_{13}}+e^{a_{15}})$), which is almost 20% shorter than that of female residents (1011 days or 2.8 years). Upon leaving RC, 18% of female residents and 34% of male residents will be transferred to NC. In other words, on average, male residents have a shorter stay in RC, and are 16% more likely to make the transition from RC to NC than female residents.

The fitted result also suggests significant difference (due to a significant b_{34}) in short-stay patterns between genders in NC. On average, male residents stay almost twice as long as female residents in the short-stay state (98 days for male, and 51 days for female). Having a longer average "settle-in" period means male NC residents do not transit into the long-stay state as quickly as female residents, which leaves them exposed for a longer period of time to 22% higher risk of discharge. In other words, male NC residents are more likely to be discharged earlier than female residents. These findings quantify the empirical observations that male residents admitted to NC take more time to settle down [5] and have poorer short-term survival prospects than female residents [21].

Fig. 5 shows the survivor curves estimated by the Kaplan–Meier type estimator and those by the Markov model for both genders in RC and NC. The difference in LOS pattern between male and female residents following admission is apparent in both plots. In particular, the sharper drop in survivor curve for male NC residents is reflected by the significant parameter b_{34}. Furthermore, this difference diminishes as time passes, which suggests that the LOS pattern in NC tends to be similar for both genders once residents settle in their new care environment.

Standard statistical tests detected significant difference (judging at 5% significant level) in LOS due to gender (log-rank test, for RC: $W_L = 4.431, p = 0.0353$; for NC: $W_L = 5.846, p = 0.0156$). However, these tests typically provide a global measure of whether significant differences exist, but not on where these differences occur. Therefore, by incorporating gender into the Markov model, our model complements standard techniques and provides additional insights into the LOS pattern for publicly funded residents in ILTC.

V. Conclusion

In this paper, we have developed a modeling approach to analyze trends and differences in LOS patterns of publicly funded residents in ILTC using routinely collected administrative data. The flow of publicly funded residents within and between RC and NC, as well as discharge from RC and NC, is modeled using a continuous-time Markov model in which a resident's stay in both RC and NC is modeled as consisting of a short-stay and a long-stay phase. Due to the nature of the data available for this study, appropriate processing of different types of observations are needed when analyzing the data. In particular, we have extended the model presented in [11] to take into account left-truncated observations as well as right-censored ones. In addition, the model has been extended to incorporate residents' features, and is able to provide additional insights into the behavior of the flow of residents in ILTC system.

The approach presented in this paper can be regarded as a model-based approach to the mining of an administrative social-care dataset. The structure of the model was inspired by domain knowledge of the system. Application of the approach to a dataset provided by the London Borough of Merton showed that such a conceptually simple model can capture high-level behavior of a complex ILTC system, and provide useful information for the planning and management of the system.

Acknowledgment

The authors thank P. Crowther, T. Temple, and the late T. Bucher from the Housing and Social Services Department of the London Borough of Merton for providing the data.

REFERENCES

[1] *The National Health Service and Community Care Act 1990 (c. 19).* London, U.K.: The U.K. Parliament, The Stationary Office Ltd., 1990

[2] *Care Standards Act 2000 (c. 14).* London, U.K.: The U.K. Parliament,The Stationary Office Ltd.

[3] D. R. Rowland and A. M. Pollock, "Choice and responsiveness for older people in the 'patient centred' NHS," *Brit. Med. J.*, vol. 328, pp. 4–5, 2004.

[4] A. Netten, A. Bebbington, R. Darton, and J. Forder, "Care homes for older people (volume 1): Facilities, residents and costs," Personal Social Services Research Unit (PSSRU), Univ. Kent, Canterbury, U.K., 2001.

[5] A. Bebbington, R. Darton, and A. Netten, "Care homes for older people (volume 2): Admissions, needs and outcomes," Personal Social Services Research Unit (PSSRU), Univ. Kent, Canterbury, U.K., 2001.

[6] G. W. Harrison and P. H. Millard, "Balancing acute and long-term care: The mathematics of throughput in departments of geriatric medicine," *Methods Inform. Med.*, vol. 30, pp. 221–228, 1991.

[7] D. R. Cox and D. Oakes, *Analysis of Survival Data*. London, U.K.: Chapman & Hall, 1984.

[8] A. Kusiak, B. Dixon, and S. Shah, "Predicting survival time for kidney dialysis patients: A data mining approach," *Comput. Biol. Med.*, vol. 35, pp. 311–327, 2005.

[9] F. Sato, Y. Shimada, F. M. Selaru, D. Shibata, M. Maeda, G. Watanabe, Y. Mori, S. A. Stass, M. Imamura, and S. J. Meltzer, "Prediction of survival in patients with esophageal carcinoma using artificial neural networks," *Cancer*, vol. 103, pp. 1596–1605, 2005.

[10] M. J. Faddy and S. I. McClean, "Analysing data on lengths of stay of hospital patients using phase-type distributions," *Appl. Stochastic Models Business Industry*, vol. 15, pp. 311–317, 1999.

[11] H. Xie, T. J. Chaussalet, and P. H. Millard, "A continuous-time Markov model for the length of stay of elderly people in institutional long-term care," *J. R. Stat. Soc. A*, vol. 168, pp. 51–61, 2005.

[12] S. I. McClean and J. O. Gribbin, "Estimating failure time distributions with left truncation and right censoring," *Appl. Stochastic Models Data Anal.*, vol. 3, pp. 13–25, 1987.

[13] I. C. Rothera, R. Jones, R. Harwood, A. J. Avery, and J. Waite, "Survival in a cohort of social services placements in nursing and residential homes: Factors associated with life expectancy and mortality," *Public Health*, vol. 116, pp. 160–165, 2002.

[14] D. R. Fredkin and J. A. Rice, "On aggregated Markov processes," *J. Appl. Probability*, vol. 23, pp. 208–214, 1986.

[15] D. R. Cox, "A use of complex probabilities in the theory of stochastic processes," *Proc. Cambridge Philos. Soc.*, vol. 51, pp. 313–319, 1955.

[16] M. J. Faddy, "Examples of fitting structured phase-type distributions," *Appl. Stochastic Models Data Anal.*, vol. 10, pp. 247–255, 1994.

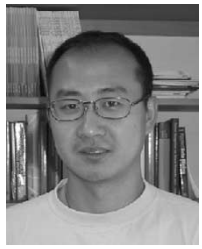
[17] M. P. Wand, "Vector differential calculus in statistics," *Amer. Statistician*, vol. 56, pp. 1–8, 2002.

[18] G. Schwarz, "Estimating the dimension of a model," *Ann. Statistics*, vol. 6, pp. 461–464, 1978.

[19] *MATLAB Reference Guide*, The Mathworks Inc., Natick, MA, 1992.

[20] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *J. Amer. Statistical Assoc.*, vol. 53, pp. 465–481, 1958.

[21] J. Cohen-Mansfield, M. S. Marx, S. Lipson, and P. Werner, "Predictors of mortality in nursing home residents," *J. Clinical Epidemiol.*, vol. 52, pp. 273–280, 1999.



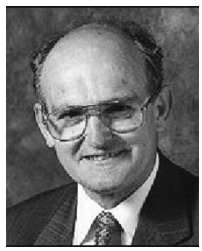
Haifeng Xie received the B.Sc. degree in applied physics from Shenzhen Univeristy, Shenzhen, China, in 1994, and the M.S. degree in decision science and the Ph.D. degree in operational research, from the University of Westminster, London, U.K., in 1999 and 2004, respectively. His Ph.D. research was concerned with the modeling of placement, survival, and costs of elderly people in long-term care.

He is currently a Research Fellow with the Health and Social Care Modeling Group (HSCMG) at the University of Westminster. His general research interests lie in the modeling of healthcare systems using stochastic processes, and statistical data mining on healthcare datasets.

Thierry J. Chaussalet received the first degree in mathematics from the University of Franche-Comte, Besançon, France, in 1986, and the M.Sc. and Ph.D. degrees in probability and stochastic processes from North Carolina State University, Raleigh, in 1990 and 1993, respectively.

He is currently Reader in Operational Research at the University of Westminster, London, U.K., where he has been lecturing for over ten years. His research interests are in the use of IT and modeling techniques in health care management, and the development of tools that can aid healthcare decision-making. He was a cofounder, with Prof. Peter Millard, of the Health and Social Care Modeling Group (HSCMG) at the University of Westminster. He has published widely in the area of healthcare modeling and decision support, and was Guest Editor of the special issue "Modeling the Process of Care" of *Health Care Management Science*, February 2001.

Dr. Chaussalet is currently Chair of the Operational Research Society (U.K.) Health and Social Services Special Interest Group, and Chair of the steering group of MASHnet, a healthcare modeling and simulation network, funded by the Engineering and Physical Sciences Research Council (U.K.).



Peter H. Millard received the M.B., B.S. (Hons. Surg.) degree from the University College Hospital, University of London, London, U.K., in 1960, and the M.D. and Ph.D. degrees from St. George's Hospital Medical School, University of London, in 1988 and 1993, respectively.

He is currently a Visiting Professor at the University of Westminster, London, U.K. In 1968, he was the first Consultant in geriatric medicine appointed to St. George's Hospital, London, U.K. Between 1993–1999, he was Eleanor Peel Professor of Geriatric Medicine at St. George's Hospital Medical School, where his research interest in modeling and new methods of measuring activity in hospitals began. Between 1993–2002, he was a Visiting Professor in the Department of Mathematics at the University of Ulster. There he cosupervised several successful Ph.D. candidates and published widely with Prof. S. McClean. In 1999, he was an Expert Advisor to the Council of Europe.

Dr. Millard's honors include President of the British Geriatrics Society, and President of the Geriatrics and Gerontology section of the Royal Society of Medicine.