An extended phase type survival tree for patient pathway prognostication

Lalit Garg, Sally McClean, Maria Barton
School of Computing and Information Engineering,
University of Ulster,
Coleraine, UK
{garg-l, si.mcclean, m.barton}@ulster.uk

Brian Meenan
School of Engineering,
University of Ulster,
Jordanstown, UK
bj.meenan@ulster.ac.uk

Ken Fullerton
College of Medicine and Health Sciences,
Queen's University,
Belfast, UK

***Abstract*— Survival tree based analysis is a powerful method of prognostication and determining clinically meaningful patient groups from a given dataset of patients' length of stay. In our previous work [1, 2] we proposed a phase type survival tree method for clustering patients into homogeneous groups with respect to their length of stay where partitioning is based on covariates representing patient characteristics such as gender, age at the time of admission, and primary diagnosis code. This paper extends this approach to examine the relationship between LOS in hospital and destination on discharge among these patient groups. An application of this approach is illustrated using 5 year retrospective data of patients admitted to Belfast City Hospital with a diagnosis of stroke (hemorrhagic stroke, cerebral infarction, transient ischaemic attack TIA, and stroke unspecified).**

Keywords- length of stay; phase type survival tree; patient pathways; capacity planning; stroke patients.

I. INTRODUCTION

Survival tree based analysis is a powerful method of partitioning survival data into clinically meaningful patient groups for prognostication i.e. for determining importance, effects of various input covariates (such as a patient's characteristics) and their effects on output measures such as patients' survival, their expected length of stay, discharge destination, treatment outcome, disease risk, or disease progress [3, 4]. Phase type survival tree [1] are special type of survival trees where each node of the tree is separately described by phase type distributions [5]. Phase type distributions can realistically model the process of a patient's journey through different stages of care as a Markov stochastic process [5]. In our previous work [1, 2], we proposed a phase type survival tree method for clustering patients into homogeneous groups with respect to their length of stay (LOS) where partitioning is based on covariates representing patient characteristics such as gender, age at the time of admission and primary diagnosis code. This paper first illustrates how this approach can be used to identify and quantify the significance and effects of various input covariates (such as a patient's characteristics) and their interrelation with a patient's length of stay in hospital. The paper then describes how such phase type survival trees can be extended to examine the relationship between LOS in hospital and destination on discharge among the patient groups identified by the tree. An application of this approach is illustrated using 5 year retrospective data [6, 7] for 1985 patients admitted between January 2003 and December 2007 to Belfast City Hospital with a diagnosis of stroke (hemorrhagic stroke, cerebral infarction, transient ischaemic attack TIA, and stroke unspecified). All patients were discharged between January 9th, 2003 and March 11th 2008. No information that identified individual patients was supplied. Patients were aged between 24 years and 101 years. The range of LOS is 0 days to 1425 days, mean LOS is 29.01 days, median LOS is 12 days, the mode LOS is 3 days, standard deviation is 52.84 days and coefficient of variation 182% [6, 7].

II. PHASE TYPE SURVIVAL TREE

Phase type distributions are among popular choices to fit spell length of stay data [5] as they are defined on the nonnegative real numbers (memoryless property) and provide an intuitive description of the patient pathways followed [5]. In [1] we illustrated how phase type survival trees can be constructed and used for clustering hospital length of stay data.

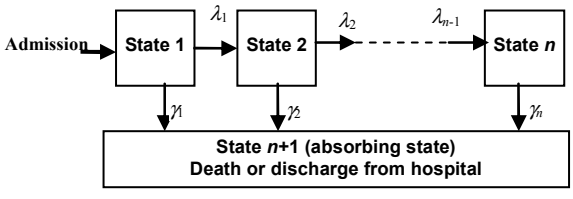

Figure 1. Stroke care system modeled as an *n* transient state Markov process with Coxian phase type distribution

978-1-4244-4998-9/10/$25.00 2010 IEEE

A. Modelling LOS using Coxian-phase type distribution

We have used Coxian-phase type distributions to approximate each node of the survival tree. Coxian-phase type distributions model a patient's journey through different stages in the care system (i.e. a patient pathway) as an n state Markov process (See Fig. 1). These states are conceptual states representing the stages in hospital. A patient can be admitted to the care system only in the first state (state 1). Sequential transitions are possible from any state k (where k = 1, 2, …, n) to the next state k+1 with a transition rate λ_k. Also transition is possible from any state k to the absorbing state n+1 with a transition rate μ_k. The absorbing state represents the event discharge or death of the patient. The time spent in the hospital before discharge or death has the probability density function:

$$\mathrm{f}(t)=\mathbf{p}\exp(\mathbf{Q}t)\mathbf{q} \tag{1}$$

where the row vector

$$\mathbf{p}=(1\ \ 0\ \ 0\ \ \ldots\ \ 0\ \ 0))$$

The transition matrix **Q** is defined as

$$\mathbf{Q}=\begin{pmatrix} -(\lambda_1+\mu_1) & \lambda_1 & 0 & \cdots & 0 \\ 0 & -(\lambda_2+\mu_2) & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & \lambda_{n-1} \\ 0 & 0 & 0 & \cdots & -\mu_n \end{pmatrix} \tag{2}$$

and the column vector **q** represents absorption probabilities and is defined as $\mathbf{q}=(\mu_1\ \ \mu_2\ \ \ldots\ \ \mu_{n-2}\ \ \mu_n)^{\mathrm{T}}$.

The log likelihood function is defined as follows [8]:

$$Log\ likelihood=\sum_{i=1}^{N}\left(\log\left(\mathbf{p}\exp\{\mathbf{Q}t_i\}\mathbf{q}\right)\right). \tag{3}$$

where N is the total number of patients in the care system and t_i is the spell length of stay of a patient i (i = 1,2,3,… N). This n state Coxian phase type fit of spell length of stay data has 2n-1 free parameters (degrees of freedom) to be estimated.

We fit Coxian phase type distribution to each group starting with one state (exponential) and progressively increasing the number of states until an optimal number of states was determined. We used a freely available downloadable package EMpht [9, 10], which implements maximum likelihood parameter estimation using the expectation-maximization (EM) algorithm.

B. Survival tree Construction

A survival tree can be constructed by recursively splitting nodes into daughter nodes by one of the covariates. A split which maximizes with in node homogeneity by providing maximum significant improvement in the function (-2*Log likelihood) is selected to grow the tree. If at a node, there is no split providing significant improvement in the function (-2*Log likelihood), the node is designated as a terminal node. The value of the chi-square statistic with 0.05 significance level $\chi^2_{(df)}\,(p<0.05)$, is used to determine the significance of the improvement in the function (-2*Log likelihood).

We used three covariates gender, age at the time of admission and type of stroke diagnosed. For the continuous covariate age we used cut-point that divide patients into groups i.e., the covariate 'age' has value 'old' for those aged 70 or over and it has value 'young' for those aged below 70 years. According to the primary diagnosis code (ICD-10 [11]), patients can have any of the four values (hemorrhagic stroke, cerebral infarction, transient ischaemic attack TIA, and other strokes) for the covariate 'stroke diagnosed'.

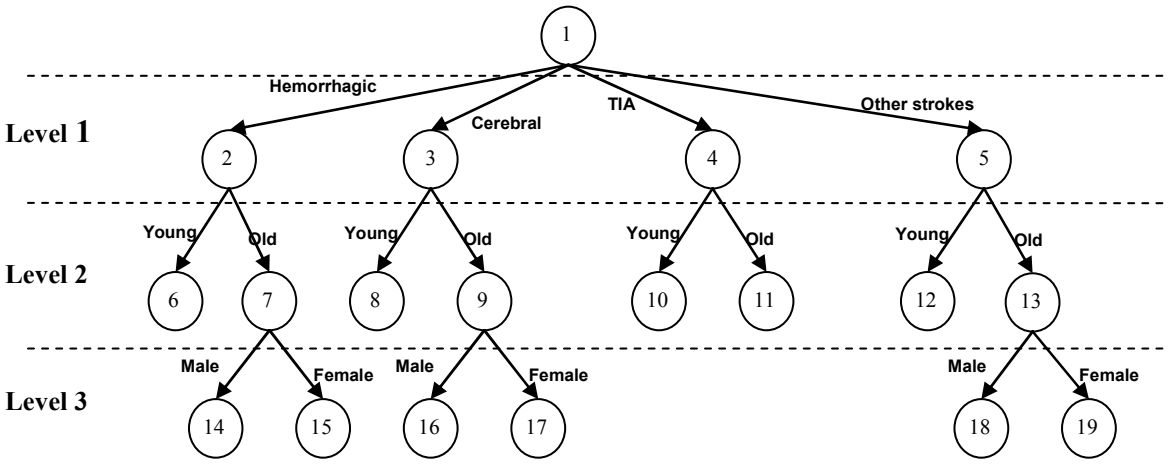

Figure 2. Phase type survival tree for length of stay data on stroke patients from the Belfast City Hospital

Fig. 2 is the schematic representation of the final phase type survival tree for the length of stay data on stroke patients from the Belfast City Hospital. The resulting tree has 11 leaf nodes. Table 1 lists nodes of the tree and possible splits of these

nodes. Bold faced covariates were selected for splitting the parent node. Table 1 also enlists number of patients in each patient group (size of the group), mean LOS and standard deviation for each patient group. This information can help in understanding the statistical difference in the length of stay among different patient groups. The total improvement in the function (-2*Log likelihood) is 524.17216 at the cost of 50 additional free parameters (p<0.000001).

TABLE I. PHASE TYPE SURVIVAL TREE CONSTRUCTION (NODES AND POSSIBLE SPLITS)

Node	Covariate	Covariate value	Number of patients	Mean LoS	Standard deviation (LoS)	Loglikelihood (L_{max})	Number of phases	Degrees of freedom (df_{max})	Total Loglikelihood	Improvement in -2*Loglikelihood	Significance (p)
All	**Complete dataset**	**Root node**	**1985**	**29.0106**	**52.8382**	**-8407.800697**	**3**	**5**	**-8407.800697**	-	-
Level 1											
1 (Root node)	Gender	Male	933	26.5938	44.0575	-3859.812524	2	8	-8399.659463	16.282468	0.000993
		Female	1052	31.154	59.4698	-4539.846939	3				
	Age	Young	624	19.2564	39.1523	-2316.973959	2	8	-8341.221197	133.159	<0.000001
		Old	1361	33.4827	57.4932	-6024.247238	3				
	Diagnosis	**Hemorrhagic**	**154**	**33.6039**	**56.4456**	**-659.050186**	**3**	**18**	**-8241.709118**	**332.183158**	**<0.000001**
		Cerebral	**655**	**36.6611**	**47.6753**	**-2973.894118**	**4**				
		TIA	**425**	**9.31294**	**19.9516**	**-1298.626224**	**2**				
		Other	**751**	**32.5433**	**65.0453**	**-3310.13859**	**2**				
Level 2											
2 Hemorrhagic	Gender	Male	80	28.2	52.09832	-317.632016	4	12	-645.798691	26.50299	0.000410
		Female	74	39.4459	60.254	-328.166675	3				
	Age	**Young**	**50**	**24.56**	**55.117**	**-173.398747**	**4**	**14**	**-642.224621**	**33.65113**	**0.000103**
		Old	**104**	**37.9519**	**56.561**	**-468.825874**	**4**				
3 Cerebral	Gender	Male	302	33.70860	49.8833	-1334.897996	4	12	-2970.092036	7.604164	0.179447
		Female	353	39.18697	45.5501	-1635.19404	3				
	Age	**Young**	**194**	**24.0670**	**42.4506**	**-785.362917**	**3**	**8**	**-2959.269766**	**29.248704**	**<0.000001**
		Old	**461**	**41.961**	**48.787**	**-2173.906849**	**2**				
4 TIA	Gender	Male	207	8.7005	22.6817	-607.954717	2	8	-1294.228181	8.796086	0.117483
		Female	218	9.8945	16.9366	-686.273464	3				
	Age	**Young**	**176**	**5.83523**	**11.1641**	**-455.863901**	**2**	**6**	**-1283.235502**	**30.781444**	**0.000001**
		Old	**249**	**11.7711**	**24.0154**	**-827.371601**	**2**				
5 Other strokes	Gender	Male	344	30.7413	43.4091	-1490.577033	4	10	-3299.172701	21.931778	0.002611
		Female	407	34.0663	78.7981	-1808.118294	2				
	Age	**Young**	**204**	**24.9608**	**43.76126**	**-818.134738**	**4**	**10**	**-3285.020568**	**50.236044**	**<0.000001**
		Old	**547**	**35.3711**	**71.1697**	**-2466.88583**	**2**				
Level 3											
6 Hemorrhagic Young	Gender	Male	29	30.5172	69.1114	-108.832894	2	6	-179.005659	-11.213824	-
		Female	21	16.3333	22.8126	-70.172765	2				
7 Hemorrhagic Old	Gender	**Male**	**51**	**26.8823**	**39.2027**	**-211.392242**	**4**	**10**	**-464.673559**	**8.30463**	**0.040119**
		Female	**53**	**48.6038**	**67.5821**	**-253.281317**	**2**				
8 Cerebral Young	Gender	Male	104	24.6731	49.2715	-420.88798	2	10	-781.939584	6.846666	0.232301
		Female	90	23.36667	32.9415	-361.051604	4				
9 Cerebral Old	**Gender**	**Male**	**198**	**38.4545**	**49.6696**	**-903.584192**	**4**	**10**	**-2162.924781**	**21.964136**	**0.002577**
		Female	**263**	**44.6008**	**47.9429**	**-1259.340589**	**2**				
10 TIA Young	Gender	Male	88	5.7386	11.3263	-224.745885	2	6	-455.625623	0.476556	0.924023
		Female	88	5.9318	10.9988	-230.879738	2				
11 TIA Old	Gender	Male	119	10.8908	28.0847	-377.732704	2	10	-822.301534	10.140134	0.180778
		Female	130	12.5769	19.5270	-444.56883	4				
12 Other strokes Young	Gender	Male	119	30.1092	52.7719	-493.332527	3	10	-816.125375	4.018726	0.259452
		Female	85	17.7529	24.6624	-322.792848	3				
13 Other strokes Old	**Gender**	**Male**	**225**	**31.0756**	**37.52**	**-987.525677**	**4**	**10**	**-2457.984373**	**17.802914**	**0.012896**
		Female	**322**	**38.3727**	**87.1713**	**-1470.458696**	**2**				

III. PROGNOSTICATION USING PHASETYPE SURVIVAL TREE

Fig. 2 shows that phase type survival tree analysis determined 11 clinically meaningful patient groups (prognostic groups) from the survival data on stroke patients from the Belfast City Hospital. Each group follows a distinct patient pathway within the system. We can examine the relationship between age, gender, diagnosis and LOS by further analysis of the results in Table 1.

At level 1, it shows that most significant split is by the covariate 'stroke diagnosed' ($\chi^2_{(df=18)}$ statistic 332.183158, p<0.000001) i.e., there was most significant difference among different stroke diagnosis groups. So patients with a diagnosis of TIA (transient ischemic attack) were most likely to have a shorter length of stay (mean LOS 9.31294, with standard deviation 19.9516, patient pathway is described by only 2 states) while patients with a diagnosis of cerebral infarction

were least likely to have shorter length of stay (mean LOS 36.6611, with standard deviation 47.6753, patient pathway is described by 4 states). The second best splitter at the level 1 is the covariate 'age' ($\chi^2_{(df=8)}$ statistic 133.159, p<0.000001). Young patients were most likely to have a shorter length of stay (mean LOS 19.2564, with standard deviation 39.1523, patient pathway is described by only 2 states) while old patients were less likely to have shorter length of stay (mean LOS 33.4827, with standard deviation 57.4932, patient pathway is described by 3 states). The other covariate 'gender' also provided significant split ($\chi^2_{(df=8)}$ statistic 16.282468, p=0.000993) however it was least significant among the three covariates.

At level 2, for all nodes, the covariate 'age' provided the most significant splits while the covariate 'gender' did not provide significant splits for the group of patients with diagnosis cerebral infarction and for the group of patients with diagnosis TIA. This can also be verified by the mean length of stay for each splits (see Table 1). For example among patients with TIA, young patients were most likely to have a shorter length of stay (mean LOS 5.83523, with standard deviation 11.1641) while old patients were less likely to have relatively shorter length of stay (mean LOS 11.7711, with standard deviation 24.0154). Similarly among patients with hemorrhagic stroke, young patients were most likely to have a shorter length of stay (mean LOS 24.56, with standard deviation 55.117) while old patients were less likely to have relatively shorter length of stay (mean LOS 37.9519, with standard deviation 56.561).

At level 3, for all groups of young patients with any type of stroke diagnosis (node 6, node 8, node 10 and node 12), the covariate gender did not provide prognostically significant splits (all such splits have p>0.05). For example, among young patients with unspecified stroke, the split into groups of male and female patients has a $\chi^2_{(df=3)}$ statistic 4.018726, p=0.259452. Similarly, among young patients with TIA, the split into groups of male and female patients has $\chi^2_{(df=3)}$ statistic 0.476556, p=0.924023, while at level 3, for groups of old patients with stroke diagnosis hemorrhagic stroke, cerebral infarction and stroke unspecified (node 7, node 9 and node 13) the covariate gender provided prognostically significant splits ($\chi^2_{(df=3)}$ statistic 8.30463, p=0.040119 for old patients with hemorrhagic stroke, $\chi^2_{(df=7)}$ statistic 21.964136, p=0.002577 for old patients with cerebral infarction and $\chi^2_{(df=7)}$ statistic 17.802914, p=0.012896 for old patients with stroke unspecified). For the group of old patients with TIA (node 11) the covariate gender split is not prognostically significant ($\chi^2_{(df=7)}$ statistic 10.140134, p=0.180778).

It illustrates that the phase type survival tree based analysis can be used to identify independent predictors of LOS and to estimate the length of stay of a patient based his/her characteristics (age, gender, diagnosis) available at the time of admission. It provides better understanding of the patient flow, heterogeneity of patient pathways and length-of-stay characteristics in addition to clustering survival data into clinically meaningful patient groups. In the next section we illustrate how this method can be extended to examine the relationship between outcome measures such as LOS in hospital and destination on discharge and their interrelationship with patient characteristics.

IV. THE EXTENDED PHASE TYPE SURVIVAL TREE

The phase type survival tree method can be extended to examine the effect of discharge destination on patient's length of stay distribution and to determine how each group of patients (determined using phase type survival tree method) can be further partitioned into subgroups with more homogeneous patient pathways. The covariate 'discharge destination' can have any of the three values death, private nursing home or other destination such as patient's normal residence.

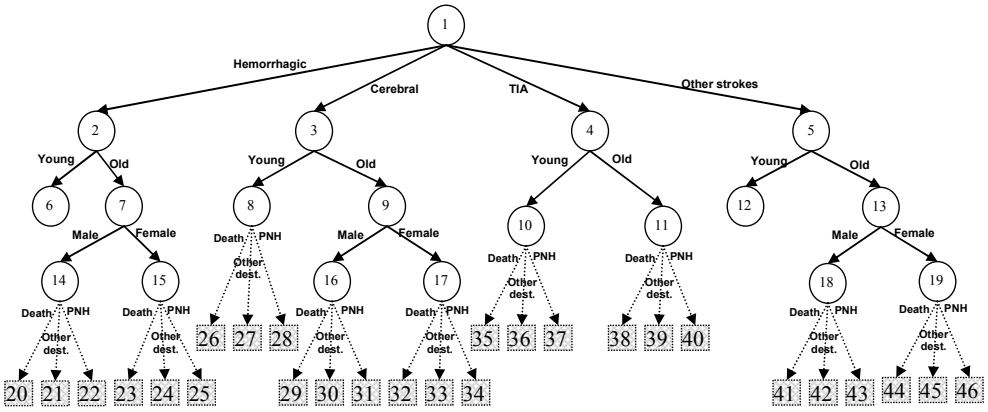

Figure 3. Extended phase type survival tree for length of stay data on stroke patients from the Belfast City Hospital

Each leaf node (or terminal node) of the survival tree of Fig. 2 is further partitioned into daughter nodes by the covariate 'discharge destination'. We grow the tree if the split maximizes node homogeneity by minimizing the BIC (Bayesian information criteria [12, 13]). If at a node, the split does not provide the lower BIC, the node is kept as a terminal node. Here we used selection criteria minimizing the BIC in place of significant improvement in the function

(-2*Log likelihood) as the population size (N) of some patient groups is very small. A Bayesian information criterion (BIC) does not only penalize the likelihood for the complexity (number of free parameters) of the model [14], it is order consistent [15] and choose the most parsimonious model even in case of a small population size [14, 15]. The Bayesian information criterion is defined to be [12, 13]:

$$\text{BIC} = -2*\text{Log likelihood} + df*\log(N) \quad (4)$$

Fig. 3 is the schematic representation of the extended phase type survival tree for the length of stay data on stroke patients from the Belfast City Hospital. The resulting tree now has 29 leaf nodes. Table 2 lists original leaf nodes of the tree and possible splits of these nodes by the covariate discharge destination.

TABLE II. TREE EXTENSION (NODES AND POSSIBLE SPLITS BY THE COVARIATE DISCHARGE DESTINATION)

Node	Destination	Number of patients (N)	Mean LoS	Standard deviation (LoS)	Loglikelihood (L_{max})	Number of phases	BIC_{mi}	Degrees of freedom (df_{max})	Total BIC	Improvement in BIC
6 Hemorrhagic Young	***All***	***50***	***24.56***	***55.11702459***	***-173.398747***	***4***	***374.1816553***	***7***	-	-
	Death	17	16.41176471	40.62462137	-45.164673	2	98.82898612	6	374.6474129	-0.4657576
	Other	33	28.75757576	60.8346145	-132.664452	2	275.8184268			
14 Hemorrhagic Old Male	***All***	***51***	***26.88235294***	***39.2026645***	***-213.977873***	***2***	***439.751223***	***3***	-	-
	Death	**21**	**10.9047619**	**14.72576221**	**-66.370745**	**2**	**141.8750574**	**5**	**420.4698446**	**19.2813784**
	Other	**27**	**31.88888889**	**38.84759713**	**-120.480944**	**1**	**244.2577249**			
	PNH	**3**	**93.66666667**	**67.4800382**	**-16.619225**	**1**	**34.3370623**			
15 Hemorrhagic Female	***All***	***53***	***48.60377358***	***67.58205065***	***-253.281317***	***2***	***518.4735099***	***3***	-	-
	Death	**27**	**24.14814815**	**35.46903052**	**-107.557149**	**2**	**225.0018087**	**7**	**504.5802341**	**13.8932758**
	Other	**24**	**74.16666667**	**85.36376411**	**-127.351546**	**1**	**257.8811459**			
	PNH	**2**	**74.16666667**	**85.36376411**	**-9.808919**	**2**	**21.69727956**			
8 Cerebral Young	***All***	***194***	***24.06701031***	***42.4505887***	***-786.459329***	***2***	***1588.722233***	***3***	-	-
	Death	**14**	**21.28571429**	**35.22464351**	**-52.239817**	**2**	**112.3968061**	**7**	**1587.106894**	**1.615339**
	Other	**174**	**22.2816092**	**40.50083021**	**-696.255928**	**2**	**1407.989022**			
	PNH	**6**	**82.33333333**	**63.21568019**	**-32.464653**	**1**	**66.72106549**			
16 Cerebral Old Male	***All***	***198***	***38.45454545***	***49.66956931***	***-911.680741***	***2***	***1839.226283***	***3***	-	-
	Death	**44**	**32.34090909**	**35.27872342**	**-196.958634**	**1**	**397.7014577**	**5**	**1828.948448**	**10.277835**
	Other	**139**	**33.92086331**	**45.72754231**	**-623.032153**	**2**	**1260.867728**			
	PNH	**15**	**98.4**	**75.12815717**	**-83.835606**	**1**	**170.3792622**			
17 Cerebral Old Female	***All***	***263***	***44.60076046***	***47.94289537***	***-1261.80847***	***1***	***2529.189094***	***1***	-	-
	Death	**68**	**37.80882353**	**52.24300047**	**-310.475206**	**2**	**633.6089352**	**5**	**2518.013299**	**11.175795**
	Other	**157**	**39.60509554**	**40.42367314**	**-734.59636**	**1**	**1474.248966**			
	PNH	**38**	**77.39473684**	**54.87907255**	**-203.258906**	**1**	**410.1553982**			
10 TIA Young	***All***	***176***	***5.835227273***	***11.16412212***	***-455.863901***	***2***	***927.2392541***	***3***	-	-
	Death	**2**	**57.5**	**12.5**	**-9.379398**	**2**	**20.83823756**	**9**	**907.0748145**	**20.1644396**
	Other	**173**	**4.803468208**	**7.772531268**	**-430.380193**	**2**	**876.2202609**			
	PNH	**1**	**81**	**0**	**-5.008158**	**2**	**10.016316**			
11 TIA Old	***All***	***249***	***11.77108434***	***24.01538595***	***-827.371601***	***2***	***1671.295561***	***3***	-	-
	Death	**11**	**33.27272727**	**30.30778751**	**-49.552113**	**1**	**101.5021213**	**9**	**1657.672845**	**13.622716**
	Other	**231**	**10.58008658**	**23.31844934**	**-741.999622**	**2**	**1500.326497**			
	PNH	**7**	**17.28571429**	**18.0136003**	**-26.949158**	**1**	**55.84422617**			
12 Other strokes Young	***All***	***204***	***24.96078431***	***43.76125549***	***-824.575588***	***2***	***1665.105536***	***3***	-	-
	Death	22	20.27272727	27.83748587	-88.204068	1	179.4991785	5	1667.239802	-2.134266
	Other	179	25.12849162	45.47772766	-720.844129	2	1457.250416			
	PNH	3	49.33333333	19.48218559	-14.695798	1	30.4902083			
18 Other strokes Old Male	***All***	***225***	***31.07555556***	***37.51999263***	***-998.19481***	***1***	***2001.80572***	***1***	-	-
	Death	**53**	**37.79245283**	**46.98388262**	**-245.501788**	**1**	**494.973868**	**3**	**1985.666354**	**16.139366**
	Other	**160**	**25.36875**	**28.49026629**	**-677.362872**	**1**	**1359.800918**			
	PNH	**12**	**77.5**	**55.08402672**	**-64.203331**	**1**	**130.8915687**			
19 Other strokes Old Female	***All***	***322***	***38.37267081***	***87.17130418***	***-1470.4587***	***2***	***2958.241047***	***3***	-	-
	Death	**89**	**44.21348315**	**151.970168**	**-395.257131**	**2**	**803.9801713**	**5**	**2928.422606**	**29.818441**
	Other	**198**	**30.60606061**	**36.63657668**	**-875.397183**	**1**	**1756.082633**			
	PNH	**35**	**67.45714286**	**48.1149474**	**-182.402227**	**1**	**368.3598021**			

Bold faced splits were selected for splitting the parent node. Parent nodes are represented by bold italic faced row with destination all. Similar to Table 1, Table 2 also presents the number of patients in each patient group (size of the group), mean LOS and standard deviation for each patient group. This information can help in understanding the statistical difference in the length of stay among different patient groups. The total improvement in the BIC is 135.99 (the total BIC of the extended tree is 16377.24, the total BIC of the survival tree of Fig. 2 is 16513.23 and the BIC_{min} of the root node is 16853.57).

After growing the tree with the covariate 'discharge destination', we can cluster the length of stay data into 29 clinically meaningful patient groups each represents a distinct patient pathway within the system. By further analysis of the results in Table 2, we can examine the relationship between

LOS and discharge destination and its interrelation with age, gender and diagnosis. We can see that in all except two patient groups (i.e., leaf nodes in Fig. 2), the discharge destination has prognostic significance, i.e., patients with different discharge destinations follow different patient pathways, while, there is homogeneity among patient pathways followed by the group of young patients with Hemorrhagic stroke. Similarly young patients with unspecified stroke followed homogeneous patient pathways. Also in all but one patient groups (i.e., leaf nodes in Fig. 2), those patients who are eventually discharged to a private nursing home are most likely to have longer length of stay. The only exception is the group of patients with TIA.

It illustrates that the extended phase type survival tree method can effectively be used to examine the relationship between LOS and destination at discharge and their interrelation with patient characteristics such as age, gender and diagnosis. It provides understanding of the heterogeneity of patient pathways and length-of-stay characteristics in addition to clustering survival data into groups of patient following homogeneous patient pathways. Although the information about the discharge destination is not available at the time of admission, we can assign the probability to each discharge destination using cohort analysis. Using the resource planning model of [16], this information can be used for estimating bed requirements and cost of care separately for each patient group following homogeneous patient pathways and thus better estimations of resource requirements and cost of care for the whole care unit as it considers the effects of individual cluster (or cohort) of patients, their interactions in the whole care unit and the effect of demographic changes in the patient population.

V. CONCLUSION

Phase type survival tree based analysis can effectively be used for prognostication of survival data and for clustering survival data into groups of patients following homogeneous patient pathways. It is an effective method for determining the relationship between input covariates and outcome measures and their interrelations. It provides understanding of heterogeneity of patient pathways stratified by covariates representing patient characteristics such as age, gender, diagnosis and outcome measures such as destination at discharge. We can also use the model to estimate the length of stay of a patient based on his/her characteristics (age, gender, diagnosis) available at the time of admission. We can extend this approach by further growing the tree by partitioning the leaf nodes into subgroups with more homogeneous patient pathways based on covariates representing outcome measures such as discharge destination. Although the information about the discharge destination is not available at the time of admission, we can assign the probability to each discharge destination using cohort analysis. This information can be used for estimating bed requirements for each group of patients (following homogeneous patient pathways) and capacity planning for the whole care system. As future work we will use phase type survival tree based analysis for modelling cost of care, blocking queues, and effect of delayed discharge in a stroke unit of a hospital.

ACKNOWLEDGMENT

The authors acknowledge support for this work from the EPSRC (Grant References EP/E019900/1 and GR/S29874/01). Any views or opinions presented herein are those of the authors and do not necessarily rep-resent those of RIGHT or MATCH, their associates or their sponsors. Support has also been received from the HSC R&D Office.

REFERENCES

[1] Garg L., McClean S. I., Meenan B. J., Millard P. H. 2009. A phase-type survival tree model for clustering patients according to their hospital length of stay. The XIII International Conference on Applied Stochastic Models and Data Analysis (ASMDA 2009), June 30- July 3, 2009, Vilnius, pp. 477-481

[2] Garg L., McClean S. I., Meenan B. J., El-Darzi E., Millard P. H. 2009. Clustering patient length of stay using mixtures of Gaussian models and phase type distributions. 22nd IEEE Symposium on Computer-Based Medical Systems (CBMS 2009), Albuquerque, New Mexico, USA, August 3-4, 2009, pp. 1-7

[3] Gao, F., Manatunga A. K. and Chen S. Identification of prognostic factors with multivariate survival data. Computational Statistics & Data Analysis 2004, 45: 813-824

[4] Davis R. and Anderson J. Exponential Survival Trees. Statistics in Medicine, 1989, 8: 947-962

[5] Fackrell M. 2009. Modelling healthcare systems with phase-type distributions. Health Care Management Science, 12: 11–26

[6] Barton M., McClean S. I., Garg L., Fullerton K. 2009. Modelling Stroke Patient Pathways using Survival Analysis and Simulation Modelling. The XIII International Conference on Applied Stochastic Models and Data Analysis (ASMDA 2009) Vilnuis, Lithuania, June 30 - July 3, 2009, pp. 370-373

[7] McClean S. I., Barton M., Garg L., Fullerton K. 2009., Combining Analytical and Simulation approaches to model Patient Flows. submitted to ACM Transactions on Modeling and Computer Simulation, unpublished

[8] Marshall, A. H. and McClean S. I. 2004. Using Coxian Phase-Type Distributions to Identify Patient Characteristics for Duration of Stay in Hospital. Health Care Management Science 7: 285–289

[9] Asmussen S., Nerman O. and Olsson M. Fitting phase-type distributions via the EM algorithm. Scandinavian Journal of Statistics, 1996, 23: 419-441

[10] Olsson, M. 1996. Estimation of phase-type distributions from censored data. Scandi-navian Journal of Statistics 23: 443–460

[11] World Health Organisation. 1993. International Classification of Diseases, ninth revi-sion (ICD-9). WHO: Geneva

[12] Schwarz, G. 1978. Estimating The Dimension of a Model, The Annals of Statistics, 6 (2), 461 – 464.

[13] Rissanen, J. 1978. Modelling by Shortest Data Description, Automatica, 14, 467 – 471.

[14] Sen, L. K. and Shitan, M. 2002. The Performance of AICC as an Order Selection Criterion in ARMA Time Series Models. Pertanika J. Sci. & Technol. 10(1): 25-33

[15] Shittu, O.I. and Asemota, M.J. 2009. Comparison of Criteria for Estimating the Order of Autoregressive Process: A Monte Carlo Approach. European Journal of Scientific Research, 30(3): 409-416.

[16] Garg, L., McClean, S. I., Meenan, B. J., Millard, P. H. 2009., A nonhomogeneous discrete time Markov model for admission scheduling and resource planning in a care system. Healthcare Management Science. In press. doi: 10.1007/s10729-009-9120-0.