Real-Time Hospital Bed Occupancy and Requirements Forecasting

Lalit Garg (lalit.garg@um.edu.mt)
University of Malta (Malta)

Adrian Mulvaney
University of Malta (Malta)

Vipul Gupta
Thapar University, (India)

Neville Calleja
University of Malta (Malta)

Abstract

To ensure better utilization and availability of the healthcare resources healthcare managers, planners and hospital staff need to develop policies. The hospital length of stay (LOS) of patients and therefore the resource requirements depend on many factors such as the covariates that represent the characteristics of the patients. Here we have used the discharge dataset of Mater Dei Hospital, Malta to model the LOS and admissions. Phase type survival tree is used to cluster patients into homogeneous groups with respect to the LOS and admissions.

Keywords: Length of Stay, Phase-Type Survival Trees, Clustering, Prognostications

Introduction

The life expectancy has increased rapidly thanks to the improvements in the health services during the last century and the standard of living. This resulted into higher demands to the healthcare service resources and making it challenging to continue providing the quality of care that is being provided at the moment. These days the healthcare systems are facing a major problem of lack of beds in hospitals and the lack of other hospital resources. To work with these problems the healthcare system needs to have an efficient way to forecast the resource requirements and to minimize the cost of care while maintaining the quality of care. This could be done by ensuring the optimal utilization and availability of the scarce healthcare resources (Garg et al., 2012). Modelling a healthcare system would help in better understanding it and aid to design the policies that can improve the quality of care and in the meantime ensuring the optimal utilization of the available resources.

To model the data survival trees are used as they are powerful methods that can cluster the data into clinically significant patient groups for forecasting, i.e. to determine the importance and effects of various input covariates (these include the patient's characteristics) and their effects on the output such as the length of stay (LOS) and the admissions (Gao et al., 2004; Davis and Anderson, 1989) Phase type survival trees (Garg et al., 2009, 2010) are types of survival trees where each node of the tree is modelled separately by a phase-type distribution (Garg et al., 2012). Phase-type distribution can be used to process the patients' flow through different stages of the healthcare system as a Markov stochastic process (Fackrell, 2009). Here we propose a phase-type survival tree method to cluster the patients into homogeneous groups with respect to their LOS and their admissions. Partitioning of the data is based on covariates representing patient characteristics such as gender, age at the time of admission, district of admission and source of admission for the LOS, while for the admissions the covariates are gender, age at the time of admission and district of admission. This paper will show how this approach can be used to identify and calculate the importance and effects of different input covariates (such as the patients' characteristics) and their interrelation with the patients' LOS in hospital and admission to hospital. An application of this approach is illustrated by using the dataset of all patients discharged during 2011 and 2012 from the Emergency department of Mater Dei Hospital, Malta. For the admissions we took into consideration all the admissions that occurred during the year 2011. The records of all the patients with an age less than one year were excluded as outliers because their admissions and discharge patterns were distinct than those of other patients. The data was all anonymized such that the patients' identity was removed.

Implementation

To model the patient flow, Phase-type distribution was used because they are useful class of distributions that have the memory less property (Garg et al., 2012). These are mostly used to model nonnegative longitudinal survival data (Garg et al., 2012). From all the phase-type distributions, the Coxian phase-type distributions have the advantage of being less likely to be over parameterized, and are less complex parameter estimation (Garg et al., 2012). The Coxian phase-type distributions give an intuitive description of patient flow through the healthcare system and have the ability to give a simple interpretation of fit that makes them an attractive choice for LOS and admissions data (Garg et al., 2012).

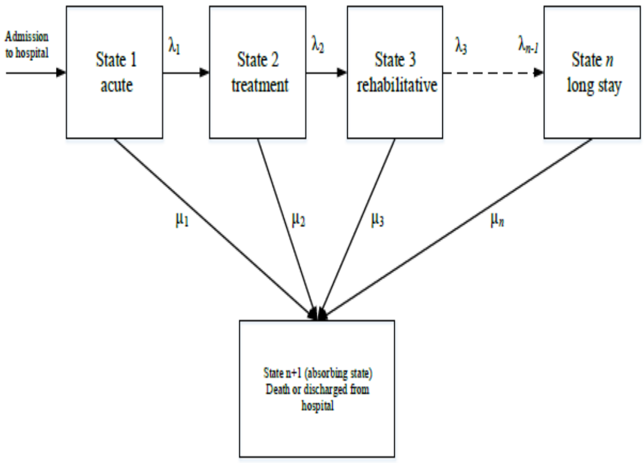

Figure 1 - Phase-type distribution as an n state continuous Markov process

Modelling LOS and Admissions using Coxian Phase-Type Distribution

Coxian Phase-type distribution was used to model the data. It is an n states continuous time Markov process with one absorbing state as in **F**igure1.

The probability distribution function for the LOS of a patient before discharge is defined as:

$$f(t) = \boldsymbol{p}(\exp(\boldsymbol{Qt}))\boldsymbol{q} \qquad (1)$$

Having p as the row vector of size (n + 1) that is representing the initial probability distribution such that

$$\boldsymbol{p} = \{1, 0, 0, \dots, 0\} \qquad (2)$$

That is the first element is 1 and all other elements are 0. Q is the transition matrix that is the continuous time Markov process and has the size of (n + 1) x (n + 1) and is defined as

$$\boldsymbol{Q} = \begin{bmatrix} -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & -(\lambda_2 + \mu_2) & \lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -(\lambda_{n-1} + \mu_{n-1}) & \lambda_{n-1} \\ 0 & 0 & 0 & \cdots & 0 & -\mu_n \end{bmatrix} \qquad (3)$$

Having μ_i as the rate of transition (also known as the absorption) from the transient state i (where i = 1,2, …,n) to the absorbing state. λ_i is representing the transition rate from the transient state to i to the next transient state i + 1. And q is a column vector that has the size (n + 1) that represents the absorption rate from each state and is defined as:

$$\boldsymbol{q} = \{\mu_1, \mu_2, \mu_3, \dots, \mu_n, 0\}^T \qquad (4)$$

The likelihood function that is used for the distribution of patient i with LOS t_i is defined as:

$$Likelihood = \prod_{i=1}^{N} \boldsymbol{p}\big(exp(\boldsymbol{Q}t_i)\big)\boldsymbol{q} \qquad (5)$$

Where N represents the number of patients. From this we can define the log-likelihood as follows:

$$Log - likelihood = \sum_{i=1}^{N} \log\big(\boldsymbol{p}\big(exp(\boldsymbol{Q}t_i)\big)\boldsymbol{q}\big) \qquad (6)$$

To fit the data to the Coxian phase-type distribution, the number of free parameters required that have to be estimated are:

$$d = 2n - 1 \qquad (7)$$

Where n is the number of transient states.

Clustering

Clustering is used to maximize within-group homogeneity and intergroup heterogeneity. For this we have clustered the Emergency department data of Mater Dei Hospital Malta dataset into clusters, such that patients are clustered into clinically meaningful patient groups. The clusters are portioned based on the covariates that represent the patient characteristics such as age, gender, district and source of admission. For the LOS data we used one continuous covariate i.e., patient's age and three categorical covariates i.e., patient gender, patient district and the patient source of admissions. For the continuous covariate we used cut points that divided the patients into three almost equal subgroups. The patient gender categorical covariate has two different values females and males. For the categorical covariate of district the patients locations have been grouped into geographical districts of Malta that are South, Central, West, North, Gozo and Unknown. For the last categorical covariate the data was grouped into five different clusters that are Elderly Home, Home (Patient's usual home), Labour Ward, Police Custody and Other (Gov Hospital, Private, Mental and Abroad).

On the other hand for the admissions we used one continuous covariate i.e., age and two categorical covariates i.e., the district of the patient and gender.

Phase-Type Survival Tree

Survival trees are constructed by recursively partitioning the survival data into homogeneous groups using covariates, aiming to either improve within node homogeneity or inter-node heterogeneity (Garg et al., 2012). Survival trees are an efficient way to cluster survival data and to understand the relationship to the covariates. Survival trees are a special type of classification and regression trees (Garg et al., 2012). In phase-type distribution trees, each node of the tree is modelled separately by a phase-type distribution (Garg et al., 2012). The phase-type survival tree has to be developed by portioning the dataset recursively into homogeneous subsets on the basis of covariates, by using splitting and selection criteria either to aim to maximize within the node homogeneity or maximize between node separations (Garg et al., 2012). We will be using splitting and selection criterion to maximize within the node homogeneity; this would be by minimizing the weighted-average information criterion (WIC). WIC is a weighted average of the Bayesian information criterion (BIC) (Garg et al., 2012) and the Akaike information criterion (AIC) with a small sample size correction (Garg et al., 2012). WIC can be calculated as follows:

$$WIC(d) = -2(Loglikelihood) + d + \left(\frac{d\left(\left((\log(N)-1)\log(N)\right)\left(N-(d+1)\right)^2+2N\left(N+(d+1)\right)\right)}{\left(2N+\left(\log(N)\left(N-(d+1)\right)\right)\right)\left(N-(d+1)\right)}\right) \quad (8)$$

Coxian phase-type distribution is used to model each node of the tree separately. For a covariate ζ that has k values and the nodes are split into k partitions (where $\zeta 1, \zeta 2, \ldots, \zeta k$), then the total value of the WIC for the splits can be calculated as follows:

$$WIC_{total}(d_{total}) = \sum_{i=1}^{k} WIC_{\zeta i}(d_{\zeta i}) \qquad (9)$$

The gain is the improvement in WIC after the splitting of the node by covariate ζ and is defined as:

$$\begin{aligned} G_{\zeta} &= \left(WIC_p(d_p)\right) - \left(WIC_{total}(d_{total})\right) \\ &= \left(WIC_p(d_p)\right) - \left(\sum_{i=1}^{k} WIC_{\zeta i}(d_{\zeta i})\right) \qquad (10) \end{aligned}$$

In the equation above $WIC_p(d_p)$ is WIC of the node before splitting (Garg et al., 2012). Starting from the root node, at each node a split takes place and the split that minimizes WIC is selected for recursively portioning the node into child nodes to grow the tree (Garg et al., 2012). At a node where there is no split that provides a positive gain (that is improvement) in WIC, there will be no more splitting and the node is a terminal (leaf) node.

Figure 2 is the schematic representation of the final phase-type survival tree for the LOS data of the Emergency department of Mater Dei Hospital. The tree has a total of 19 terminal (leaf) nodes. **Table 1** lists the nodes of the LOS tree and the possible splits of these nodes. The covariates that are bold faced represent the nodes that were selected for splitting. **Table 1** also enlists the number of patients in each group and the mean LOS. This information can help in understanding the statistical difference in the length of stay among different patient groups. The total Gain in WIC is 12619.16.

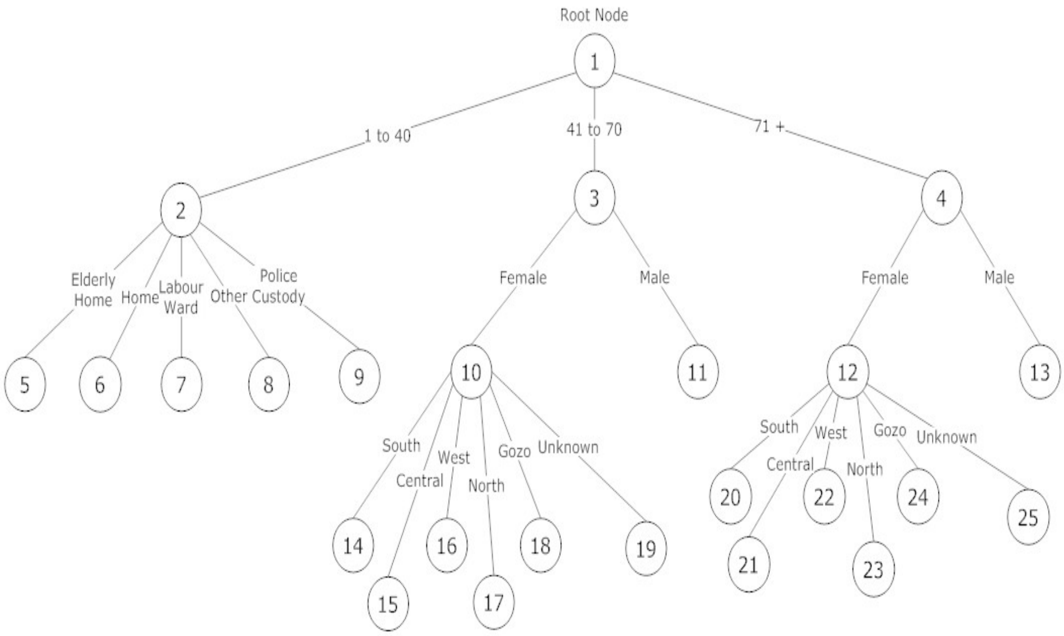

Figure 2 - LOS Phase Type Survival Tree

Table 1 - Phase-Type Survival Tree Construction Using WIC-Based Splitting Criteria for LOS

Node	**Covariate**	**Covariate Value**	**Total Number of Patients**	**WIC**	**Mean LOS**	**Number of phases**	**Total WIC**	**Gain in WIC**
Level 1								
1 (Root Node)	All	Root Node	64439	351604.66	6.8411	6	351604.66	-
	Age	**1 to 40**	**20631**	**87222.35**	**4.1304**	**6**	**341295.6**	**10309.1**
		41 to 70	**22600**	**122877.8**	**6.7443**	**5**		
		71 +	**21208**	**131195.4**	**9.5813**	**5**		
	District	South	22237	121077.72	6.756	5	351775.15	-170.49
		Central	19480	107177.13	6.9864	4		
		West	8423	46460.1	7.0515	5		
		North	13542	72716.7	6.6032	4		
		Gozo	539	3227.25	8.3358	5		
		Unknown	218	1116.25	5.5	4		
	Source	Elderly Home	1925	11775.05	9.4732	6	351078.46	526.2
		Home	61356	332501.72	6.7339	6		
		Labour Ward	2	32.84	4.5	6		
		Other (Gov Hospital, Private, Mental and Abroad)	1060	6297.08	8.4632	6		
		Police Custody	96	471.77	4.7604	2		
	Gender	Female	32886	179393.48	6.8672	6	351637.51	-32.85
		Male	31553	172244.02	6.814	5		
Level 2								
2 (1 to 40)	All	1 to 40	20631	87222.35	4.1304	6	87222.35	-
	District	Age 1 to 40 from South	7323	31441.23	4.2134	5	87414.82	-192.47
		Age 1 to 40 from Central	5895	24526.66	4.0271	6		

		Age 1 to 40 from West	2652	11251.58	4.1478	4		
		Age 1 to 40 from North	4457	18669.77	4.0449	5		
		Age 1 to 40 from Gozo	215	1108.2	5.3953	5		
		Age 1 to 40 from Unknown	89	417.38	4.8427	3		
	Gender	Age 1 to 40, Female	12220	51064.07	3.9629	4	87270.43	-48.09
		Age 1 to 40, Male	8411	36206.37	4.3737	4		
	Source of Admission	**1 to 40, Elderly Home**	**12**	**76.35**	**8**	**1**	**86477.36**	**744.99**
		1 to 40, Home	**20342**	**84944.65**	**4.1008**	**6**		
		1 to 40, Labour Ward	**1**	**1.39**	**2**	**1**		
		1 to 40, Other (Gov Hospital, Private, Mental and Abroad)	**208**	**1116.25**	**6.6058**	**5**		
		1 to 40, Police Custody	**68**	**338.73**	**4.7647**	**2**		
3 (41 to 70)	All	41 to 70	22600	122877.78	6.7443	5	122877.78	-
	Age, District	Age 41 to 70 from South	8081	44010.37	6.7598	4	122510.88	366.9
		Age 41 to 70 from Central	6589	36087.12	6.8774	5		
		Age 41 to 70 from West	2799	15135.04	6.0831	5		
		Age 41 to 70 from North	4841	25558.59	9.4944	6		
		Age 41 to 70 from Gozo	205	1267	9.0537	3		
		Age 41 to 70 from Unknown	85	452.76	5.6824	3		
	Age, Gender	**Age 41 to 70, Female**	**9088**	**49410.24**	**6.817**	**4**	**122091**	**786.78**
		Age 41 to 70, Male	**13512**	**72680.77**	**6.6954**	**4**		
	Age, Source of Admission	41 to 70, Elderly Home	158	1064.51	12.5063	3	122859.01	18.77
		41 to 70, Home	21954	118923.91	6.6666	5		
		41 to 70, Labour Ward	1	3.89	7	1		
		41 to 70, Other (Gov Hospital, Private, Mental and Abroad)	459	2725.98	8.5991	4		
		41 to 70, Police Custody	28	140.71	4.75	2		
4 (71 +)	All	71 +	21208	131195.43	9.5813	5	131195.43	-
	Age, District	Age 71 + from South	6833	42291.14	9.4764	5	131335.03	-139.59
		Age 71 + from Central	6996	43403.75	9.5826	4		
		Age 71 + from West	2972	18533.21	9.8765	4		
		Age 71 + from North	4244	26013	9.4944	4		
		Age 71 + from Gozo	119	841.4	12.4117	1		
		Age 71 + from Unknown	44	252.52	6.4773	2		
	Age, Gender	**Age 71 +, Female**	**11578**	**72543.24**	**9.9719**	**5**	**130740.7**	**454.71**
		Age 71 +, Male	**9630**	**58197.48**	**9.1118**	**5**		
	Age, Source of Admission	71 +, Elderly Home	1755	10652.36	9.2103	4	131153.79	41.64
		71 +, Home	19060	118037	9.6216	5		
		71 +, Other (Gov Hospital, Private, Mental and Abroad)	393	2464.43	9.2875	3		
Level 3								
10 (Age 41 to 70, Female)	All	Age 41 to 70, Female	9088	49410.24	6.817	4	49410.24	-
	District	**41 to 70, South, F**	**3164**	**17051**	**6.8587**	**6**	49148.34	261.9
		41 to 70, Central, F	**2782**	**15094.21**	**6.8724**	**5**		
		41 to 70, West, F	**1123**	**6118.53**	**6.9154**	**5**		
		41 to 70, North, F	**1933**	**10357.31**	**6.5525**	**3**		
		41 to 70, Gozo, F	**55**	**366.03**	**9.9454**	**1**		
		41 to 70, Unknown, F	**31**	**161.25**	**4.9678**	**3**		
	Source of Admission	41 to 70, Elderly Home, F	81	561.03	12.4445	3	49396.46	13.78
		41 to 70, Home, F	8835	47791.27	6.7268	4		
		41 to 70, Labour Ward, F	1	3.89	7	1		
		41 to 70, Other (Gov Hospital, Private, Mental and Abroad), F	170	1038.88	8.8529	4		
		41 to 70, Police Custody, F	1	1.39	2	1		
12 (Age 71 +,	All	Age 71 +, Female	11578	72543.24	9.9719	5	72543.24	-
	District	**71 +, South, F**	**3663**	**22859.81**	**9.8444**	**6**	**72219.66**	**323.58**

Female)		**71 +, Central, F**	**3880**	**24104.55**	**9.8023**	**6**		
		71 +, West, F	**1736**	**11040.79**	**10.4919**	**4**		
		71 +, North, F	**2242**	**13837.2**	**10.0589**	**6**		
		71 +, Gozo, F	**40**	**287.23**	**12.825**	**1**		
		71 +, Unknown, F	**17**	**90.08**	**4.8235**	**1**		
	Source of Admission	71 +, Elderly Home, F	1257	7655.41	9.4121	4	72532.89	10.35
		71 +, Home, F	10093	63415.04	10.0396	6		
		71 +, Other (Gov Hospital, Private, Mental and Abroad), F	228	1462.44	10.057	4		

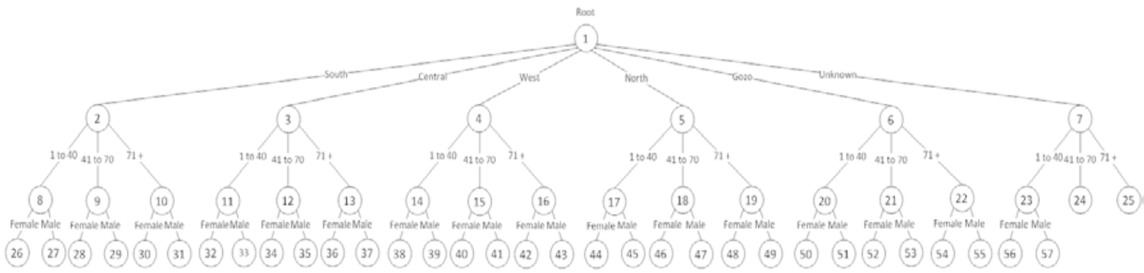

Figure 3 - Admissions Phase Type Survival Tree

Table 2- Phase-Type Survival Tree Construction Using WIC-Based Splitting Criteria for Admissions

Node	Covariate	Covariate Value	Total Admissions	WIC	Mean	Number of Phases	Average WIC	Total WIC	Gain in WIC
Level 1									
1 (Root Node)	All	Root Node	32277	3171.43	89.43	22	3171.43	3171.43	-
	Age	1 to 40	10386	2561.57	29.45	10	853.86	2576.47	594.96
		41 to 70	11244	2590.39	31.81	10	863.46		
		71 +	10647	2577.45	30.17	10	859.15		
	Gender	Female	16510	2793.52	44.2	10	1396.76	2811.39	360.04
		Male	15767	2829.26	46.23	10	1414.63		
	District	**South**	**11211**	**2581.18**	**31.72**	**10**	**430.2**	**1756.39**	**1415.04**
		Central	**9690**	**2491.79**	**27.55**	**10**	**415.3**		
		West	**4270**	**2051.09**	**12.7**	**10**	**341.85**		
		North	**6774**	**2289.19**	**19.56**	**10**	**381.53**		
		Gozo	**289**	**895.58**	**1.79**	**6**	**149.26**		
		Unknown	**43**	**229.51**	**1.12**	**10**	**38.25**		
Level 2									
2 (South)	**Age**	**1 to 40**	**3781**	**2028.31**	**11.36**	**8**	**112.68**	**334.23**	**95.97**
		41 to 70	**4015**	**2023.7**	**12**	**9**	**112.43**		
		71 +	**3415**	**1964.13**	**10.36**	**8**	**109.12**		
	Gender	Female	5669	2189.02	16.53	10	182.42	364.74	65.45
		Male	5542	2187.92	16.18	10	182.33		
3 (Central)	**Age**	**1 to 40**	**2952**	**1899.26**	**9.09**	**7**	**105.51**	**324.94**	**90.36**
		41 to 70	**3267**	**1950.83**	**9.95**	**7**	**108.38**		
		71 +	**3471**	**1998.81**	**10.51**	**8**	**111.05**		
	Gender	Female	5020	2169.14	14.75	10	180.76	358.14	57.16
		Male	4670	2128.53	13.79	9	177.38		
4 (West)	**Age**	**1 to 40**	**1326**	**1582.64**	**4.63**	**4**	**87.92**	**264.86**	**76.99**
		41 to 70	**1405**	**1562.65**	**4.85**	**5**	**86.81**		
		71 +	**1539**	**1622.14**	**5.22**	**6**	**90.12**		
	Gender	Female	2293	1787.68	7.28	6	148.97	292.59	49.26
		Male	1977	1723.42	6.42	6	143.62		
5 (North)	**Age**	**1 to 40**	**2186**	**1746.03**	**6.99**	**6**	**97**	**297.06**	**84.47**
		41 to 70	**2428**	**1828.69**	**7.65**	**6**	**101.59**		
		71 +	**2160**	**1772.36**	**6.92**	**6**	**98.46**		
	Gender	Female	3388	1963.06	10.28	8	163.59	327.13	54.4
		Male	3386	1962.54	10.28	8	163.54		

6 (Gozo)	**Age**	**1 to 40**	**114**	**510.04**	**1.31**	**10**	**28.34**	**74.08**	**75.19**
		41 to 70	**116**	**525.91**	**1.32**	**9**	**29.22**		
		71 +	**59**	**297.46**	**1.16**	**10**	**16.53**		
	Gender	Female	121	532.86	1.33	10	44.4	99.67	49.59
		Male	168	663.2	1.46	8	55.27		
7 (Unkown)	**Age**	**1 to 40**	**27**	**129.21**	**1.07**	**10**	**7.18**	**9.38**	**28.87**
		41 to 70	**13**	**28.04**	**1.04**	**10**	**1.56**		
		71 +	**3**	**11.67**	**1.01**	**8**	**0.65**		
	Gender	Female	19	68.14	1.05	10	5.68	15.5	22.75
		Male	24	117.85	1.07	10	9.82		
Level 3									
8 (South, 1 to 40)	**Gender**	**Female**	**2263**	**1817.71**	**7.2**	**5**	**50.49**	**94.97**	**17.71**
		Male	**1518**	**1601.38**	**5.16**	**5**	**44.48**		
9 (South, 41 to 70)	**Gender**	**Female**	**1602**	**1617.75**	**5.39**	**5**	**44.94**	**94.31**	**18.11**
		Male	**2413**	**1777.52**	**7.61**	**7**	**49.38**		
10 (South, 71 +)	**Gender**	**Female**	**1804**	**1680.7**	**5.94**	**5**	**46.69**	**91.78**	**17.34**
		Male	**1611**	**1623.45**	**5.41**	**5**	**45.1**		
11 (Central, 1 to 40)	**Gender**	**Female**	**1761**	**1719.87**	**5.82**	**5**	**47.77**	**89.34**	**16.18**
		Male	**1191**	**1496.32**	**4.26**	**5**	**41.56**		
12 (Central, 41 to 70)	**Gender**	**Female**	**1325**	**1565.73**	**4.63**	**5**	**43.49**	**91.16**	**17.21**
		Male	**1942**	**1716.2**	**6.32**	**6**	**47.67**		
13 (Central, 71 +)	**Gender**	**Female**	**1934**	**1725.28**	**6.3**	**5**	**47.92**	**92.36**	**18.68**
		Male	**1537**	**1599.83**	**5.21**	**5**	**44.44**		
14 (West, 1 to 40)	**Gender**	**Female**	**820**	**1357.36**	**3.25**	**4**	**37.7**	**69.44**	**18.49**
		Male	**506**	**1142.3**	**2.39**	**4**	**31.73**		
15 (West, 41 to 70)	**Gender**	**Female**	**565**	**1200.36**	**2.55**	**4**	**33.34**	**70.41**	**16.41**
		Male	**840**	**1334.26**	**3.3**	**4**	**37.06**		
16 (West, 71 +)	**Gender**	**Female**	**908**	**1387.43**	**3.49**	**4**	**38.54**	**71.95**	**18.17**
		Male	**631**	**1202.62**	**2.73**	**4**	**33.41**		
17 (North, 1 to 40)	**Gender**	**Female**	**1304**	**1563.15**	**4.57**	**4**	**43.42**	**81.14**	**15.86**
		Male	**882**	**1357.83**	**3.42**	**4**	**37.72**		
18 (North, 41 to 70)	**Gender**	**Female**	**959**	**1411.44**	**3.63**	**4**	**39.21**	**84.06**	**17.54**
		Male	**1469**	**1614.66**	**5.02**	**5**	**44.85**		
19 (North, 71 +)	**Gender**	**Female**	**1125**	**1488.1**	**4.08**	**4**	**41.34**	**81.41**	**17.05**
		Male	**1035**	**1442.69**	**3.84**	**4**	**40.07**		
20 (Gozo, 1 to 40)	**Gender**	**Female**	**64**	**323.82**	**1.18**	**10**	**8.99**	**16.17**	**12.16**
		Male	**50**	**258.44**	**1.14**	**10**	**7.18**		
21 (Gozo, 41 to 70)	**Gender**	**Female**	**64**	**323.82**	**1.18**	**10**	**8.99**	**20.15**	**9.06**
		Male	**82**	**401.76**	**1.23**	**10**	**11.16**		
22 (Gozo, 71 +)	**Gender**	**Female**	**24**	**100.2**	**1.07**	**10**	**2.78**	**7.27**	**9.26**
		Male	**35**	**161.34**	**1.1**	**10**	**4.48**		
23 (Unknown, 1 to 40)	**Gender**	**Female**	**13**	**22.86**	**1.04**	**10**	**0.64**	**1.89**	**5.29**
		Male	**14**	**45.21**	**1.04**	**10**	**1.26**		

Figure 3 is the survival tree representation for the admission data of the Emergency department of Mater Dei Hospital. The tree has a total of 34 leaf nodes. Table 2 lists the nodes of admissions tree and the possible splits. Table 2 enlists the number of admissions per group and the mean number of admissions. This information can help in understanding the statistical difference in the rate of admissions among different patient groups. The total Gain in WIC is 2111.41.

Prognostication

Figure 2 and Figure 3 are showing the phase type survival trees analysis of the determined clinically meaningful patient groups from the survival data of the Emergency department of Mater Dei Hospital. In Fig. 2 we can examine the relationship between the age, gender, district, source of admission and the LOS by further analysis of the results in Table 1. At the first level we see that the highest split gain (in WIC) is by the covariate age (WIC Gain 10309.1) i.e. there was most significant difference between the age groups. So patients within the age group 1 to 40 were most likely to have a shorter LOS (mean LOS 4.1304) while patients in the age group of 71 + were least likely to have longer LOS (mean LOS 9.5813). The other significant splitter among the covariates at level 1 is the covariate source of admission (WIC gain 526.20). At Level 2 we find that for the age group 1 to 40, the covariate with the most significant split is the source of admission (WIC Gain 744.99) and the other two covariates had no significance for this age group. On the other hand for the other two age groups the most significant splitter is the Gender covariate (for the age group 41 to 70, WIC Gain 786.78, for the age group 71 +, WIC Gain 454.71). This can be verified by the mean LOS for each split in Table 1. At level 3 for the nodes under the age group 1 to 40 (nodes 5, 6, 7, 8 and 9) the covariates did not provide any significant splits. In the other age groups the male covariate did not provide any significance splits and the Female covariate had their covariates split by the District of Admission covariates (for the age group 41 to 70, WIC Gain 261.9 and for the age group 71+ WIC Gain 323.58).

In Figure 3 that is the phase type survival tree for the admissions at level 1 we find the district covariate that has the highest split gain (WIC Gain 1415.04) i.e. the most significant difference the admission rates is between the district groups. The next significant split at level was the age covariate (WIC Gain 594.96). At level 2, for all nodes the covariate age provides the most significant splits as it can be seen in Table 2. At level 3 we can find that almost all the nodes apart from node 24 and 25, all have a significant split by the gender covariate.

Conclusion

From the research work presented in this paper, we can conclude that the phase type survival tree analysis can be used effectively not only to prognosticate and cluster survival data into homogeneous patient groups based on their LOS and understand its relationship with covariates and their interrelations but also to understand admission patterns and their relationship with covariates. These models could also be used to forecast the bed occupancy (Garg et al. 2010, 2012) and resource requirements at any point in time in future. The LOS model can be used to estimate the LOS of a patient at admission and the admissions model can be used to estimate the number of admissions by the patient characteristics.

For future work we are extending our models by taking into consideration the other covariates like the admitting ward, diagnosis and the discharge locations to characterize their relationship with admission rate and LOS. Also these models can be improved by taking into consideration the waiting lists and discharge delay i.e., the delay in discharging patients from hospital.

References

Davis, R. and Anderson, J. (1989), “Exponential Survival Trees”, *Statistics in Medicine,* Vol. 8, pp. 947-962.

Fackrell,M.(2009), “Modelling healthcare systems with phase-type distributions”, *Health Care Management Science*, Vol. 12, pp. 11-26.

Gao, F., Manatunga A. K. and Chen S. (2004), “Identification of prognostic factors with multivariate survival data” ,*Computational Statistics & Data Analysis,* Vol.45, pp. 813-824.

Garg, L., McClean, S. I., Meenan, B. J., and 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.

Garg, L., McClean, S. I., Meenan, B. J., and Millard P. H. (2010), “A non-homogeneous discrete time Markov model for admission scheduling and resource planning in a cost or capacity constrained healthcare system”, *Health Care Management Science*. Vol. 13, No. 2, pp. 155–169.

Garg, L., McClean, S. I., Barton, M., Meenan, B. J., and Fullerton, K. (2010), “Forecasting hospital bed requirements and cost of care using phase type survival trees”, *The 2010 IEEE Conference on Intelligent Systems (IEEE IS’10)*, July 7-9, 2010, London, UK, pp. 185-190.

Garg, L., McClean, S. I., Barton, M., Meenan, B. J., and Fullerton, K. (2012), “Intelligent Patient Management and Resource Planning for Complex, Heterogeneous, and Stochastic Healthcare Systems” *Systems And Humans*, Vol. 42, No. 6.