Development and Validation of Models to Predict Hospital Admission for Emergency Department Patients

Bin Xie[1,2,*]

[1]Department of Obstetrics & Gyneocology; [2]Department of Epidemiology & Biostatistics, University of Western Ontario, Ontario, Canada

Abstract: *Background*: Boarding, or patients waiting to be admitted to hospital, has been shown as a significant contributing factor at overcrowding in emergency departments (ED). Predicting hospital admission at triage has been proposed as having the potential to help alleviate ED overcrowding. The objective of this paper is to develop and validate a model to predict hospital admission at triage to help alleviate ED overcrowding.

Methods: Administrative records between April 1, 2010 and November 31, 2010 in an adult ED were used to derive and validate two prediction models, one based on Coxian phase type distribution (the PH model), the other based on logistic regression. Separate data sets were used for model development (data between April 1, 2010 and July 31, 2010) and validation (data between August 1, 2010 and November 31, 2010).

Results: There were a total of 14,542 ED visits and 2,602 (17.89%) hospital admissions in the derivation cohort. In both models, acuity levels, model of arrival, and main reason of the visit are strong predictors of hospital admission; number of patients at the ED, as well as gender, are also predictors, albeit with ORs closer to 1. Patient age and timing of visits are not strong predictors. The PH model has an AUC of 0.89 compared with AUC of 0.83 for logistic regression model; with a cut- off value of 0.50, the PH model correctly predicted 86.3% of visits, compared to 84.4% for the logistic regression model. Results of the validation cohort were similar: the PH model has an AUC of 0.88, compared to AUC of 0.83 for the logistic model.

Conclusions: PH and logistic models can be used to provide reasonably accurate prediction of hospital admission for ED patients, with the PH model offering more accurate predictions.

Keywords: Hospital admission, Emergency department, Wait times, Overcrowding, Coxian phase type distribution.

1. BACKGROUND

Lengthy wait times and overcrowding at emergency departments (ED) have been a serious problem in many communities [1-4]. Consequences for such lengthy wait times and overcrowding include decreased patient satisfaction [4, 5], increased patient morbidity and mortality [6-9], and increased costs [9, 10]. Many measures have been proposed or attempted to address this issue, with various degrees of success [1, 3, 9].

It has been shown that boarding, or the holding up of ED resources by patients waiting for hospital beds to be admitted, is a major contributor of ED overcrowding [11], as boarding reduces the overall throughput of the ED and negatively impact patient outcomes [12-14]. To minimize the negative impacts of boarding, it has been suggested that predicting hospital admission at the time of triage to enable advanced planning could help manage ED resources more effectively [15, 16].

A variety of approaches have been reported in the literature to predict hospital admission for ED patients, including statistical models based on logistic regression [15, 17, 18], artificial intelligence models such as bayesian network or artificial neural network models [12, 16, 19], or human prediction by staff [20-24]. The literature reported mixed performances of these approaches, [15, 17, 20-25], with some studies reported impressive performance using statistics such as area under curve (AUC) statistics for receiver operating characteristic (ROC) curve [15, 17], and other studies reported less impressive results [20-25].

Statistical models that increase the accuracy of admission prediction can therefore enhance the practical values of models currently in the literature. One potential way to increase accuracy is to utilize information on the process under which patients went through the EDs by using models based on Coxian phase-type distributions (hereafter referred to as PH distributions).

PH distributions are a special type of Markov chain model that describes duration until an event occurs in terms of a process consisting of a sequence of latent phases [9, 26]. These distributions have the ability to model probabilities of transition from one phase to another as well as probabilities of absorption from various phases, and have been used in various healthcare settings and have been shown to offer

*Address correspondence to this author at the Room E2-620B, LHSC-VH, 800 Commissioners Road East, London, Ontario, N6A 5W9, Canada; Tel: +1 519-685-8500-55174; Fax: +1 519-685-8176; E-mail: bxie5@uwo.ca

E-ISSN: 1929-6029/13

superior fit compared to alternative distributions such as lognormal or gamma distributions [9, 26-34]. Their ability to incorporate information of the underlying Markov process in ED, as well as their ability to model probabilities of admission from various phases, suggest that models based on Coxian phase-type distributions could be used to predict ED admission.

In this paper, the author presents the results of such a PH model to predict hospital admission. A logistic regression model is also presented as a performance benchmark. Methodologies for the phase-type parameterisation of covariates and the use of Bayes rule for classification were similar to those reported in the literature [30, 31].

2. METHODS

2.1. Data Source

Data in this study came from administrative ED records from an adult ED site in a middle-sized community with a population of half a million in southwest Ontario, Canada. The ED records include patient demographic (age, gender) and clinical (main reasons for the visit, acuity level) information, timing and mode of patient arrival, as well as timing and type of discharge. Operating conditions of the ED, including total number of patients at a given time, were obtained from the ED records using the timing information for arrival and discharge for all patients.

ED records between April 1, 2010 and July 31, 2010, were used for model derivation (the derivation cohort), and data between August 1, 2010 and November 31, 2010, were used for model validation (the validation cohort). This sequential division of derivation and validation cohort may introduce bias due to differences in timing of admission between the two cohorts; however given that our purpose is to see if model parameters derived from past data can be used to predict future visits (in different time), this approach was used instead of a random division approach. Eligible patients include all patients who presented in the ED with valid information for time of triage and time and type of discharge. Those who died while in the ED or those who left before being seen or left before completing treatments were excluded as the timing of their departure from the ED was unknown.

2.2. Measurement

Variables that have been widely used to predict hospital admission at ED in the literature, including patient age, gender, acuity level, model of arrival, and main reason for visit, are included in the prediction models. More specifically, patient age was used as a continuous variable as it has been shown that probability of admission increase with age in a monotonic and near linear fashion [25]. Acuity levels were determined at triage based on the Canadian Triage and Acuity Scale (CTAS) [35]. Mode of arrival was categorized into two groups: self arrive or arrival by ambulance. Main reason for visit was categorized into five categories using ICD 10 codes: “Mental and behavioral disorders” (ICD codes starting with F), “Pregnancy, childbirth and the puerperium” (ICD codes starting with O), “Diseases of the circulatory system” (ICD codes starting with I), “Injury, poisoning, or external causes of morbidity and mortality” (ICD codes starting with S, T, V, or Y), and other reasons.

Other variables, including the time and date of the visits and operating conditions of the ED, were also included in the models as these variables have been shown to affect the throughput of ED and could be associated with parameters of the PH model [36-38]. Number of patients in the ED at the time of triage is calculated by counting the number of patients who were already triaged but have yet to be discharged. Number of patients was log transformed before entering the models.

Binary variables were coded as 0/1 (0 for the reference value). Dummy variables were created for categorical variables with more than two categories, and the means of continuous variables were subtracted to assist in interpretation.

2.4. Statistical Modeling

Statistical associations between admission status and the covariates were assessed using t tests for continuous covariates and Chi-squared tests for categorical covariates. Probability of hospital admission is the outcome variable for the PH model. Details of the PH model, including model specification and parameter estimation, are included in the Appendix for interested readers. For the logistic model, outcome variable is hospital admission (Yes or No).

Estimation of model coefficients will be performed using data from the derivation cohort, and these two models will be validated using data from validation cohort. The performances of these two models will be measured using AUC statistic for the ROC curve and prediction accuracy (the probability of predicting correctly using a certain cut-off threshold value). All statistical analysis were performed using R 2.14.0 [39].

2.5. Ethics

The study was reviewed and approved by the research ethics board at the University of Western Ontario. The study protocol conforms to the ethical guidelines of the "World Medical Association Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects" adopted by the 18th WMA General Assembly, Helsinki, Finland, June 1964 and amended by the 59th WMA General Assembly, Seoul, South Korea, October 2008.Given the facts that the analysis used de-identified patient level data and that no intervention was applied to patient care, informed consent from the participants was not required.

3. RESULTS

3.1. Characteristics of Patients and their ED Visits

Table **1** presents characteristics of patients and their ED visits, as well as the differences between the visits that ended up with hospital admission and those that did not. There were 14,542 visits in the derivation cohort, 2,602 (17.89%) of which ended with hospital admission. Patients with higher acuity levels, whose main reason for visit was disease of circular systems, or who arrived by ambulance, were more likely to be admitted to hospital. All other covariates have weaker, albeit still statistically significant, associations with hospital admission. The validation cohort includes 15,105 visits, 2,726 (18.05%) of which ended with hospital admission. The characteristics of patients and their visits in the validation cohort is similar to those in the derivation cohort (results not shown in Table **1**).

Especially noteworthy is the difference in admission rates among visits of different acuity levels. For example, 85.92% of visits with “Resuscitation” level ended up with hospital admission, whereas only 5.59% of visits with “Less urgent” or “Non-urgent” ended up with admission. CTAS level alone can predict admission for these visits with high accuracy. On the other hand, visits with CTAS levels “Emergent” or “Urgent”, had a 27.57% probability of hospital admission.

Given the difficulty in predicting admission for these patients [25]as well as the fact that these visits counted for a majority of total visits (54.21%) and admissions (83.52%) in the derivation cohort, separate analysis was performed on these visits. Results of such analysis were presented in the Appendix.

3.2. ESTIMATION OF PARAMETERS OF THE PH DISTRIBUTION

A 4-phase PH distribution was deemed sufficient to fit the data, as a 5-phase PH distribution did not statistically significantly improve goodness of fit (p value: 0.235). Table **2** presents results of multivariate models that regress each of the parameter values (λ_i, μ_i) for the 4-phase PH model on all the covariates. Numbers presented in the table include estimated coefficient vectors a_i and c_i describe dependence of the parameters of the PH distribution on covariates, as well as an Intercept value for each parameter λ_i or μ_i, which represents the combined effect of coefficient vectors b_i (for λ_i) and d_i (for μ_i) (i.e., the parameter values when all covariates equal to 0). As detailed in section A.3.1 in the Appendix, a_i and b_i represent dependence of λ_i on the covariates, and c_i and b_i represent dependence of μ_i on covariates. The intercept values, or b_i (for λ_i) and b_i (for μ_i), can be interpreted as the parameter values (λ_i and μ_i) for a patient with all covariates equal to 0. The coefficient values for the covariates, or a_i (for λ_i) and c_i (for μ_i), represent change in parameter values (λ_i and μ_i) compared to the reference category (for categorical variables) or difference from the mean (for continuous variables). With these coefficients, λ_i and μ_i for any given covariate vector can be calculated using the formulation presented in section A.3.1.

Results in Table **2** provide information that can be used to formulate hypothesis for the meaning of the four phases in the PH distribution. For example, an interpretation of the four phases as waiting to be assessed, initial examination by ED staff, further diagnostic testing, and waiting for discharge (e.g., boarding, or waiting for testing results), appears to be consistent with results in Table **2**. For example, μ_i is extremely small, consistent with the fact that discharge before initial assessment was very rare. Other interpretations may also be plausible, but given that the focus of this study is the prediction of hospital admission, the meaning of these four phases will not be explored further.

3.3. Predicting Admission

Tables **3** and **4** present regression coefficients of PH model (e_i and f_i) and logistic regression model, respectively. For the PH model, Table **3** presents results of multivariate models that regress the logit of

Table 1: Admission vs. Characteristics of Study Participants and their ED Visits*

	Total (N=14542)	Admission (n=2602, 17.9%)	No admission (n=11940, 82.1%)	p value
	Number	Number (%)	Number (%)	
Age (mean [SD])	49.6 [23.4]	65.6 [19.7]	46.1 [22.7]	<0.01
Gender				0.031
Female	7590	1331 (17.5%)	6259 (85.5%)	
Male	6952	1271 (18.3%)	5681 (81.7%)	
Time of the day				<0.01
8AM - 4PM	6369	1189 (18.7%)	5180 (81.3%)	
4PM – Midnight	5697	1076 (18.9%)	4621 (81.1%)	
Midnight - 8AM	2476	337 (13.6%)	2139 (86.4%)	
Day of the week				<0.01
Weekdays	10564	1944 (18.4%)	8620 (81.6%)	
Weekends	3978	658 (16.5%)	3320 (83.5%)	
Main reason				<0.01
Diseases of the circulatory system	1045	524 (50.1%)	521 (49.9%)	
Mental and behavioural disorders	435	32 (7.4%)	403 (92.6%)	
Pregnancy, childbirth and the puerperium	72	6 (8.3%)	66 (91.7%)	
Injury, poisoning, or external causes of morbidity and mortality	3250	332 (10.2%)	2918 (89.8%)	
Other reasons	9740	1708 (17.5%)	8032 (82.5%)	
Triage level				<0.01
1-Resuscitation	71	61 (85.9%)	10 (14.1%)	
2-Emergency	1903	748 (39.3%)	1155 (60.7%)	
3-Urgent	5979	1425 (23.8%)	4554 (76.2%)	
4-Less Urgent	6146	366 (6.0%)	5780 (94.0%)	
5-Non Urgent	443	2 (0.5%)	441 (99.5%)	
Mode of arrival				<0.01
Self arrival	10273	977 (9.5%)	9296 (90.5%)	
Ambulance	4269	1625 (38.1%)	2644 (61.9%)	
Number of patients in the ED (Mean[SD])	25.7 [9.1]	26.7 [8.8]	25.5 [9.1]	<0.01

*Some participants had multiple ED visits during the study time period; these multiple visits were counted as separate visits. P values were from t tests for continuous variables or chi-squared tests for categorical variables.

admission from each of the phase, $\text{logit}(p_a(i))$ on all the covariates. Numbers presented in the table include estimated coefficient vectors e_i describing dependence of the logit of admission on covariates, as well as an Intercept value for each phase, which represents the combined effect of coefficient vectors f_i. With these coefficients, the probability of admission for any given covariate vector can be calculated using the formulation presented in section A.3.1.

The probability of hospital admission for a patient with certain covariate profile can be estimated by summarizing the probabilities from each phase corresponding to the covariate profile. Results in Table **3** reveal important differences in the probabilities of admission from different phases and the impacts of different covariates on these probabilities. For example, the intercept value for phase 1 is several orders of magnitude smaller than those of other phases, suggesting that probability of admission from phase 1 is extremely small. For another example, compared to the reference category (Injury, poisoning, or external causes of morbidity and mortality), patients with

Table 2: Coefficients for Regressing Parameters of 4-Phase PH Distribution on Covariates*

Covariates	Phase 1		Phase 2		phase 3		Phase 4
	λ_1 (SE)	μ_1 (SE)	λ_2 (SE)	μ_2 (SE)	λ_3 (SE)	μ_3 (SE)	μ_4 (SE)
Intercept	-5.08 (0.52)	-20.32 (1.34)	-4.30 (0.56)	-5.85 (0.67)	-4.28 (0.45)	-6.92 (0.80)	-4.16 (1.43)
Age	0.08 (0.05)	0.11 (0.15)	-0.12 (0.03)	-0.05 (0.02)	-0.09 (0.04)	-0.06 (0.05)	0.02 (0.06)
Gender (Reference: Female)							
Male	0.06 (0.15)	1.76 (2.50)	0.11 (0.25)	0.17 (0.35)	0.67 (0.36)	0.52 (0.76)	-0.05 (0.35)
Time of the day (Reference: 8AM - 4PM)							
4PM – Midnight	-0.86 (0.12)	-1.24 (2.68)	0.39 (0.16)	-1.83 (0.35)	0.26 (0.15)	-0.30 (0.12)	0.29 (0.14)
Midnight - 8AM	1.33 (0.21)	-2.17 (0.23)	0.82 (0.12)	-1.51 (0.23)	1.53 (0.23)	-0.48 (0.35)	-0.70 (0.43)
Day of the week (Reference: Weekdays)							
Weekends	0.14 (0.05)	5.31 (0.25)	0.29 (0.08)	-5.24 (0.38)	0.32 (0.06)	-2.20 (0.50)	0.11 (0.05)
Main reason (Reference: Injury, poisoning, or external causes of morbidity and mortality)							
Diseases of the circulatory system	0.92 (0.12)	5.36 (1.05)	-1.59 (0.28)	-1.95 (0.57)	1.01 (0.43)	1.26 (0.35)	1.54 (0.92)
Mental and behavioural disorders	0.45 (0.14)	-3.50 (1.14)	0.53 (0.24)	2.20 (1.28)	0.04 (0.05)	-1.84 (0.83)	-1.22 (0.92)
Pregnancy, childbirth and the puerperium	-0.42 (0.35)	-0.92 (0.56)	-0.32 (0.18)	-1.35 (0.80)	-0.16 (0.15)	1.53 (1.28)	1.05 (0.99)
Other reasons	-0.61 (0.15)	-5.81 (0.65)	-0.53 (0.14)	-0.78 (0.20)	-0.58 (0.18)	1.15 (0.35)	0.32 (0.16)
Triage level (Reference: 3-Urgent)							
1-Resuscitation	2.59 (0.56)	8.38 (2.12)	1.02 (0.80)	-4.68 (2.20)	1.09 (0.34)	-4.82 (2.00)	1.88 (1.50)
2-Emergency	1.62 (0.45)	-1.65 (0.65)	1.99 (0.20)	-2.30 (1.13)	-1.14 (0.50)	2.27 (1.01)	0.89 (0.39)
4-Less Urgent	-0.86 (0.12)	-1.99 (0.45)	0.24 (0.05)	0.46 (0.10)	0.18 (0.09)	-0.75 (0.11)	-0.27 (0.12)
5-Non Urgent	-1.55 (0.43)	5.21 (1.34)	-0.08 (0.03)	1.78 (0.32)	-0.08 (0.06)	-2.26 (0.25)	-1.87 (0.52)
Mode of arrival (Reference: Self arrival)							
Ambulance	0.23 (0.05)	1.84 (0.20)	-0.14 (0.05)	-5.14 (0.83)	-0.44 (0.18)	3.55 (0.50)	0.26 (0.16)
log(# of patients in the ED)	-0.25 (0.03)	0.08 (0.06)	-0.30 (0.05)	0.16 (0.08)	-0.28 (0.10)	0.12 (0.09)	0.07 (0.09)

*****Binary variables were coded as 0/1, with the reference category coded as 0. For categorical variables with more than two categories, dummy variables were created for each category other than the reference category.

diseases of the circulatory system have higher probability of admission from phase 2 than other phases. These results are consistent with the interpretation of the meaning of the 4 phases discussed in section 3.2 above.

These results show that, as would be expected, acuity level, mode of arrival, and main reason for the visit are strong predictors of ED admission. For example, compared to patients with acuity level 3, patients with acuity level 1 are more than 10 times more likely to be admitted; compared to those who arrive by themselves, those who arrive by ambulance are more than 3 times more likely to be admitted. Gender and number of patients in the ED are also predictors of admission, albeit with ORs closer to 1. Other variables, including age, and timing of the visit, are not strong predictors of admission, although the ORs reached statistical significance given the large sample size.

Comparison between the ORs in Table **3** and Table **4** reveals a number of interesting differences between the two models. For example, as noted above, the ORs for the same covariate may vary, sometimes quite significantly, across different phases in the PH model, suggesting that the same covariate had different impacts on the odds of admission from different phases. These ORs also differ notably from the OR of the same covariate in the logistic regression model, suggesting that the impact of a covariate on the odds of admission overall can be quite different from the odds of admission from some of the phases. Statistical significance of these differences can be assessed by

Table 3: Regression Coefficients Using Probabilities of Admission from Phases 1-4 as Outcome Variables*

	Phase 1		Phase 2		phase 3		phase 4	
Covariates	**OR [95% CI]**	**p**	**OR [95% CI]**	**p**	**OR [95% CI]**	**p**	**OR [95% CI]**	**p**
Intercept	3.24e-9 [0.00-2.30e-5]	<0.01	0.01 [0.0001-0.02]	<0.01	0.020 [0.001-0.05]	<0.01	0.015 [0.002-0.04]	<0.01
Age	0.98 [0.97-0.99]	<0.01	1.01 [1.01-1.02]	<0.01	1.06 [1.04-1.08]	<0.01	1.02 [1.01-1.03]	<0.01
Gender (Reference: Female)								
Male	1.18 [1.05-1.30]	<0.01	1.23 [1.10-1.35]	<0.01	1.21 [1.10-1.32]	<0.01	1.19 [1.09-1.31]	<0.01
Time of the day (Reference: 8AM - 4PM)								
4PM – Midnight	1.03 [0.91-1.04]	0.508	1.05 [0.95-1.06]	0.361	1.00 [0.93-1.04]	0.179	0.98 [0.89-1.08]	0.263
Midnight - 8AM	0.98 [0.65-0.91]	<0.01	0.72 [0.65-0.91]	<0.01	0.83 [0.65-0.91]	<0.01	0.65 [0.65-0.91]	<0.01
Day of the week (Reference: Weekdays)								
Weekends	0.85 [0.73-0.99]	0.011	0.93 [0.83-1.04]	0.312	0.96 [0.88-1.05]	0.462	0.81 [0.69-0.95]	<0.01
Main reason (Reference: Injury, poisoning, or external causes of morbidity and mortality)								
Diseases of the circulatory system	1.93 [1.12-2.82]	<0.01	3.95 [3.17-4.72]	<0.01	3.50 [2.82-4.20]	<0.01	3.01 [2.15-3.92]	<0.01
Mental and behavioural disorders	0.38 [0.21-0.56]	<0.01	0.45 [0.28-0.63]	<0.01	0.40 [0.24-0.62]	<0.01	0.39 [0.24-0.60]	<0.01
Pregnancy, childbirth and the puerperium	0.89 [0.50-1.32]	0.346	1.21 [0.85-1.55]	0.208	1.99 [1.23-2.73]	0.028	2.10 [1.53-2.70]	0.031
Other reasons	1.35 [1.15-1.56]	<0.01	1.48 [1.24-1.75]	<0.01	1.40 [1.19-1.62]	<0.01	1.46 [1.26-1.67]	<0.01
Triage level (Reference: 3-Urgent)								
1-Resuscitation	18.01 [8.53-27.40]	<0.01	10.50 [5.82-15.36]	<0.01	20.03 [13.65-26.39]	<0.01	9.85 [4.18-15.52]	<0.01
2-Emergency	1.21 [1.01-1.42]	0.039	1.68 [1.40-1.98]	0.011	1.46 [1.31-1.60]	0.009	1.53 [1.30-1.75]	0.012
4-Less Urgent	0.29 [0.20-0.47]	<0.01	0.20 [0.11-0.29]	<0.01	0.42 [0.31-0.43]	<0.01	0.48 [0.36-0.61]	0.011
5-Non Urgent	0.02 [0.01-0.05]	<0.01	0.02 [0.01-0.06]	<0.01	0.03 [0.01-0.09]	<0.01	0.05 [0.01-0.10]	<0.01
Mode of arrival (Reference: Self arrival)								
Ambulance	4.05 [3.16-4.96]	<0.01	2.62 [2.07-3.19]	<0.01	3.93 [3.08-4.89]	<0.01	2.92 [2.26-3.59]	<0.01
log(# of patients in the ED)	1.20 [1.05-1.35]	<0.01	1.23 [1.08-1.39]	<0.01	1.32 [1.15-1.50]	<0.01	1.30 [1.11-1.50]	<0.01

*Binary variables were coded as 0/1, with the reference category coded as 0. For categorical variables with more than two categories, dummy variables were created for each category other than the reference category.

comparing the 95% confidence intervals (CIs) of these ORs; non-overlapping CIs suggest statistically significant differences. For another example, some covariates are not statistically significant predictors of admission in the logistic regression model, but are statistical significant predictors of admission from some of the transitory phases in the PH model, suggesting that the impacts of these covariates may be limited to certain phases. Examples include Day of week, which is not statistically significant predictor in the logistic regression model, but is statistically significant in the phases 1 and 4 of the PH model.

3.4. Performance Measures

Figure **1** presents the ROC curves and the AUC statistics of the two models for the deviation and

Table 4: Results of the Logistic Regression Model*

	OR [95% CI]	p
Intercept	0.013 [0.008-0.023]	<0.01
Age	1.02 [1.02-1.03]	<0.01
Gender		
Female	Reference	
Male	1.20 [1.08-1.32]	<0.01
Time of the day		
8AM - 4PM	Reference	
4PM – Midnight	1.02 [0.91-1.04]	0.747
Midnight - 8AM	0.77 [0.65-0.91]	<0.01
Day of the week		
Weekdays	Reference	
Weekends	0.91 [0.82-1.02]	0.120
Main reason		
Diseases of the circulatory system	3.48 [2.87-4.22]	<0.01
Mental and behavioural disorders	0.40 [0.26-0.59]	<0.01
Pregnancy, childbirth and the puerperium	1.68 [0.63-3.70]	0.236
Injury, poisoning, or external causes of morbidity and mortality	Reference	
Other reasons	1.44 [1.25-1.66]	<0.01
Triage level		
1-Resuscitation	13.63 [6.97-29.40]	<0.01
2-Emergency	1.50 [1.33-1.70]	<0.01
3-Urgent	Reference	
4-Less Urgent	0.32 [0.29-0.37]	<0.01
5-Non Urgent	0.03 [0.01-0.10]	<0.01
Mode of arrival		
Self arrival	Reference	
Ambulance	3.14 [2.83-3.49]	<0.01
Number of patients in the ED	1.29 [1.11-1.49]	<0.01

*Binary variables were coded as 0/1, with the reference category coded as 0. For categorical variables with more than two categories, dummy variables were created for each category other than the reference category.

validation cohorts with all CTAS levels. It can be seen that for both models, the results of the validation cohort are almost identical to those of the deviation cohort, suggesting that the models can be used to predict admission for future visits. It can also be seen that the PH models resulted in higher AUC than logistic regression models (AUC 0.89 vs. 0.83, p=0.009), suggesting that it offers more accurate prediction. The PH model also had higher accuracy (i.e., the frequency of total cases correctly predicted using model results) than the logistic model using the cut-off value that maximizes accuracy (results not shown in the graph). For the PH model, with a cut-off value of 0.43 (corresponding sensitivity 0.87: specificity: 0.68), the model was able to correctly predict the admissions status of 87.2% of visits, compared to 84.4% for the logistic regression model, using a cut-off value of 0.46 (corresponding sensitivity: 0.83; specificity: 0.68).

4. DISCUSSION

Given the role boarding plays in ED overcrowding, predicting hospital admission for ED patients at time of triage can help alleviate ED overcrowding [15]. Results from this paper show that both the PH model and the logistic model can offer reasonably accurate

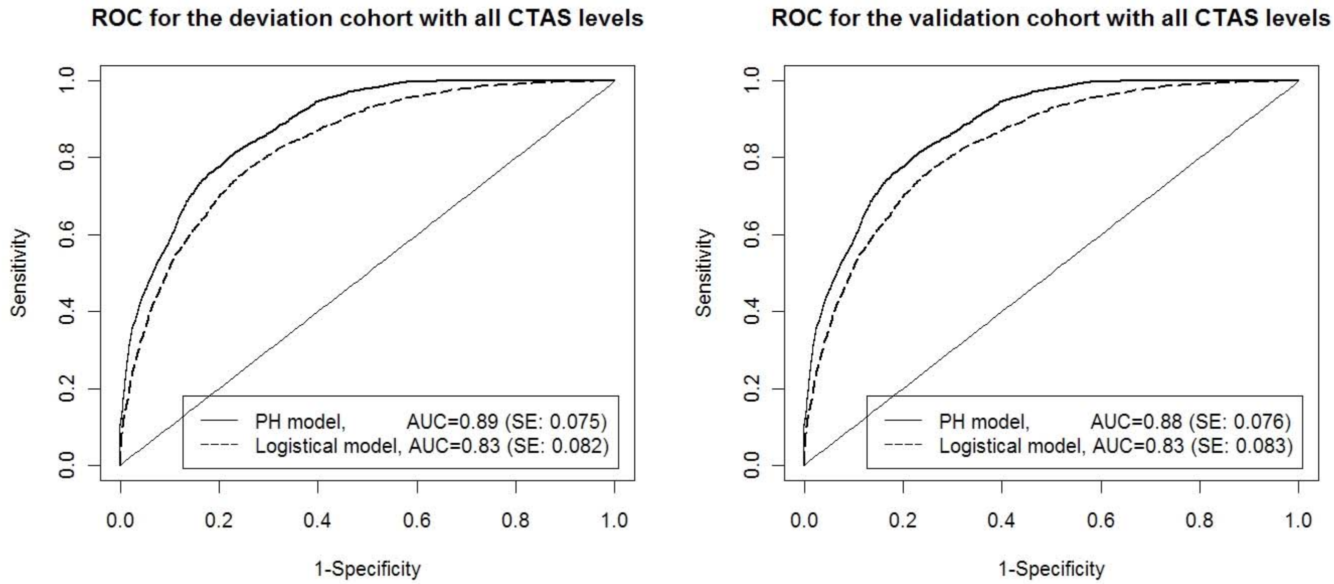

Figure 1: Performance comparison for the two models using all visits.

predictions. Of these two models, the PH model offers more accurate predictions and should be the preferred model and can be used where results of other approaches are deemed unsatisfactory [20-24]. PH models also provide additional information that can potentially be used for planning purpose. For example, PH models not only provide prediction of hospital admission, it also provides information on the possible timing of such admission. This information could be useful in further facilitating advanced planning and is unavailable from logistic regression models.

Besides offering reasonably accurate predictions of ED admission, this study also provide some interesting findings that need to be explored further for their implications to help policy makers and ED providers. For example, results show that the number of patients in the ED at the time of triage is associated with higher likelihood of hospital admission, after controlling for other covariates. The reason for this positive association is not apparent; one possible reason could be that when the ED is more crowded and when hospital beds are available, patients are more likely to be admitted to available hospital beds to release resources in the ED to accommodate the increased load. If this is the case, a more detailed investigation should be carried out to assess the effectiveness of this strategy, as it could be counterproductive in the long run, when decreased availability of hospital beds led to increased likelihood of boarding.

These performances notwithstanding, it is important to point out that using statistical models to predict probability of admission at the individual patient level has some inherent difficulties, as these models are supposed to predict the proportion of success among a subgroup of subjects with similar covariate values and would be less valid in predicting individual success [40]. This difficulty is especially apparent in the error rate of predicting visits with intermediate acuity levels (19.5% error rate using 0.50 as cut off value), suggesting that caution should be used when using such information for resource planning purpose. It is possible, for example, that even this improved accuracy of PH model is unacceptable. For example, in some institutions, a named bed would be reserved for a patient predicted to need admission [20]. In such arrangement, extreme caution should be used to avoid reserving beds for patients who end up not needing them and worsening the wait for patients predicted to not needing admission but end up needing it [20].

Besides this inherent difficulty to make predictions in the individual level, there are a number of other limitations specific to this study that need to be acknowledged. To begin with, the results are based on data from a single institution, and it is unclear to what extent the conclusions of this study can be applied to institutions with different characteristics. Another limitation is the fact that PH models involve the estimation of many more parameters than the logistic regression method, thus raising the risk of overfitting. To overcome this potential problem, a validation cohort was used to ensure that results from the PH models can yield valid predictions for future patients.

5. CONCLUSIONS

Both models can predict hospital admission for ED patients with reasonable accuracy using information available at triage, with PH model offering more accurate predictions. PH models also can provide additional information, including timing of hospital admission, that is unavailable from logistic models.

APPENDIX

A.1. Mathematical Properties of a n-Phase Coxian Phase Type Distribution

The transition probabilities from one state to its next state for a n-phase PH distribution can be written as:

$$P\{X(t+\delta t)=i+1 \mid X(t)=i\}=\lambda_i\delta t+o(\delta t), \text{ for } i=1,2\ldots,\mathrm{n}-1$$

The probabilities of absorption can be written as:

$$P\{X(t+\delta t)=n+1 \mid X(t)=i\}=\mu_i\delta t+o(\delta t), \text{ for } i=1,2\ldots,\mathrm{n}$$

The probability density function (pdf) of the time spent before absorption is:

$$f(t)=\mathbf{P}\exp\{\mathbf{Q}t\}\mathbf{q},$$

where

$$\mathbf{P}=(1 \quad 0 \quad 0 \quad \ldots \quad 0 \quad 0),$$

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

$$\mathbf{q}=(\mu_1 \quad \mu_1 \quad \ldots \quad \mu_n)^{\mathrm{T}}.$$

A.2. PH Regression Model Specification

A *n*-phase PH distribution represents a Markov chain with *n* transitory phases and an absorption phase, as shown in Figure **1**. Such a PH distribution has 2*n*-1 parameters, $\lambda_i, i=1,\ldots,n-1$ and $\mu_i, i=1,\ldots,n$. A patient's journey in this Markov chain is determined by λ_i and μ_i: the time spent on each transitory phase *i* follows an exponential distribution with mean $\frac{1}{\lambda_i+\mu_i}$, the probability of transitioning to next phase is $\frac{\lambda_i}{\lambda_i+\mu_i}$, and the probability of absorption is $\frac{\mu_i}{\lambda_i+\mu_i}$.

A.3. Parameter Estimation of the PH Model

A.3.1. Outline

Estimation of the probability of hospital admission will be done in a three-step process. In step 1, parameter values of a PH distribution that fit the length of stay (LOS) data at the ED will be estimated using maximum likelihood methods. The number of phases, *n*, will be determined by gradually increasing the number of phases (starting with 1 phase) until the difference in goodness of fit between the *n*-phase PH distribution and a *n*+1 phase distribution is no longer statistically significant. Parameters λ_i and μ_i will be estimated using the formulation: $\lambda_i=e^{aiX+bi}, i=1,\ldots,n-1$ and $\mu_i=e^{ciX+di}, i=1,\ldots,n$ where X represent the covariate vector, a_i, b_i, c_i and d_i are regression coefficient vectors.

In step 2, probabilities of admission through the *n* transitory phases will be estimated using information on patient discharge (timing and probability of hospital admission) as well as parameters of this *n*-phase PH distribution, using the formulation $\mathrm{logit}(p_a(i))=e^{eiX+fi}$ where $p_a(i), i=1,2,\ldots,n$ represents the probability of admission from phase *i*, X represent the covariate vector, and e_i and f_i are the coefficient vectors. Please note that probability of admission from phase *i* is different from the probability of absorption from phase *i*, as the latter include probability of discharge home as well as probability of admission. In step 3, a probability of admission for each patient will be estimated by adding up the probabilities of admission from these *n* phases.

A.3.2. Estimating Coefficients a_i, b_i, c_i and d_i

Suppose $\lambda_i = e^{aiX+bi}, i=1,\ldots,n-1$, and $\mu_i = e^{ciX+di}, i=1,\ldots,n$. Replace λ_i and μ_i in the definition of p, Q and q with e^{aiX+bi} and e^{ciX+di} the likelihood function for the data is:

$\mathrm{L} = \prod_{k=1}^{N} \mathbf{P}\exp\{\mathbf{Q}t_k\}\mathbf{q}$. Coefficients a_i, b_i, c_i and d_i can be estimated by maximizing this likelihood function.

A.3.3. Estimating coefficients e_i and f_i

Suppose that a n-phase PH distribution has been fit to describe the ED process as described in A.2 above. Denote $p_a(i), i=1,\ldots,n$, as the probability of hospital admission at phase *i*, and assume that $p_a(i)$ depends on covariates in the form of $\mathrm{logit}(p_a(i))=e^{eiX+fi}$. The conditional probability of hospital admission given LOS of t, $p_a(i|t)$ can be expressed as $p_a(i|t)=\sum_{i=1}^{n} p(i|t)p_a(i)$, where $p(i|t)$ is the conditional probability of a discharge that ended at time *t* happened at phase *i*, and

$$\mathrm{p(i|t)} = \frac{\mathrm{f(t|i)}\prod_{\mathrm{k=1}}^{\mathrm{i}}\frac{\lambda_\mathrm{k}}{\lambda_\mathrm{k}+\mu_\mathrm{k}}}{\sum_{\mathrm{j=1}}^{\mathrm{n}}\mathrm{f(t|j)}\prod_{\mathrm{k=1}}^{\mathrm{j}}\frac{\lambda_\mathrm{k}}{\lambda_\mathrm{k}+\mu_\mathrm{k}}}$$

, where f(t|i) is the pdf of the time of absorption given that absorption occurs from phase *i*. $f(t|i) = \mathbf{p}\exp\{\mathbf{Q}t\}\mathbf{q}$ where

$$\mathbf{P} = (1 \quad 0 \quad 0 \quad \ldots \quad 0 \quad 0),$$

$$\mathbf{Q} = \begin{pmatrix} -(\lambda_1+\mu_1) & \lambda_1+\mu_1 & 0 & \cdots & 0 \\ 0 & -(\lambda_2+\mu_2) & \lambda_2+\mu_2 & \vdots & \vdots & 0 \\ \vdots & \vdots & \vdots & & & \vdots \\ & & & & & 0 \\ 0 & 0 & 0 & \cdots & & -(\lambda_i+\mu_i) \end{pmatrix}$$

$$\mathbf{q} = (0 \quad 0 \quad \ldots \quad (\lambda_i+\mu_i))^{\mathrm{T}}.$$

So the conditional likelihood function is:

$L = \prod_{k-1}^{N}\sum_{i=1}^{n} p(i|t_k)p_a(i)$. Coefficients e_i and f_i can be estimated by maximizing this conditional likelihood function.

A.4. Separate Analysis for CTAS Levels 2 and 3 Visits

Figure **A1** presents the ROC curves and the AUC statistics of the two models for the deviation cohorts with visits of CTAS levels 2 and 3, as well as the accuracy of prediction (i.e., the probability of prediction being true) using different cut-off values. Results of the validation cohort were almost identical (not shown in the graph). The improvement of prediction accuracy of the PH model over the logistic regression model for these visits is especially noticeable. For example, for a cut-off value of 0.50 (i.e., visits with estimated admission probability above 0.50 were deemed admitted, whereas those below 0.50 were deemed not admitted), the PH model predicted 80.5% visits correctly, whereas the logistic regression model predicted 76.1% correctly.

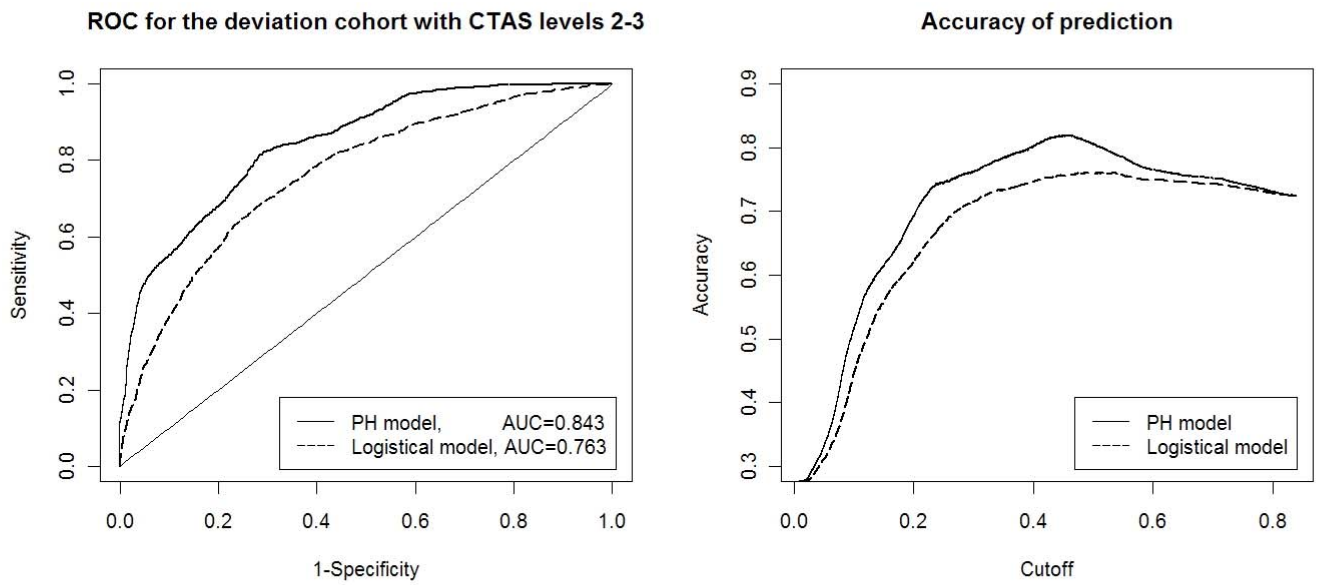

Figure A1: Performance comparison for the two models using visits with CTAS levels 2 and 3.

REFERENCES

[1] Hoot NR, Aronsky D. Systematic review of emergency department crowding: Causes, effects, and solutions. Ann Emerg Med 2008; 52(2): 126-36. http://dx.doi.org/10.1016/j.annemergmed.2008.03.014

[2] Begley CE, Chang Y, Wood RC, *et al*. Emergency department diversion and trauma mortality: Evidence from houston, texas. J Trauma 2004; 57(6): 1260-5. http://dx.doi.org/10.1097/01.TA.0000135163.60257.A6

[3] Moskop JC, Sklar DP, Geiderman JM, *et al*. Emergency department crowding, part 2--barriers to reform and strategies to overcome them. Ann Emerg Med 2009; 53(5): 612-7. http://dx.doi.org/10.1016/j.annemergmed.2008.09.024

[4] National ambulatory care reporting system. [Internet]. Toronto: CIHI; 2011 [updated Oct 2011; cited October 31, 2011]. Available from: http: //www.cihi.ca/CIHI-ext-portal/internet/en/document/types+of+care/hospital+care/emergency+care/services_nacrs.

[5] Derlet RW, Richards JR. Overcrowding in the nation's emergency departments: Complex causes and disturbing effects. Ann Emerg Med 2000; 35(1): 63-8. http://dx.doi.org/10.1016/S0196-0644(00)70105-3

[6] Sprivulis PC, Da Silva JA, Jacobs IG, *et al*. The association between hospital overcrowding and mortality among patients admitted *via* western australian emergency departments. Med J Aust 2006; 184(5): 208-12.

[7] Richardson DB. Increase in patient mortality at 10 days associated with emergency department overcrowding. Med J Aust 2006; 184(5): 213-6.

[8] Bernstein SL, Aronsky D, Duseja R, *et al*. The effect of emergency department crowding on clinically oriented outcomes. Acad Emerg Med 2009; 16(1): 1-10. http://dx.doi.org/10.1111/j.1553-2712.2008.00295.x

[9] Aaronson LS, Mural CM, Pfoutz SK. Seeking information: Where do pregnant women go? Health Educ Q 1988; 15(3): 335-45. http://dx.doi.org/10.1177/109019818801500307

[10] Bayley MD, Schwartz JS, Shofer FS, *et al*. The financial burden of emergency department congestion and hospital crowding for chest pain patients awaiting admission. Ann Emerg Med 2005; 45(2): 110-7. http://dx.doi.org/10.1016/j.annemergmed.2004.09.010

[11] Yonkers KA, Wisner KL, Stewart DE, *et al*. The management of depression during pregnancy: A report from the american psychiatric association and the american college of obstetricians and gynecologists. Gen Hosp Psychiatry 2009; 31(5): 403-13. http://dx.doi.org/10.1016/j.genhosppsych.2009.04.003

[12] Leegon J, Jones I, Lanaghan K, *et al*. Predicting hospital admission in a pediatric emergency department using an artificial neural network. AMIA Annu Symp Proc 2006; 1004.

[13] Hodgins MJ, Moore N, Legere L. Who is sleeping in our beds? factors predicting the ED boarding of admitted patients for more than 2 hours. J Emerg Nurs 2011; 37(3): 225-30. http://dx.doi.org/10.1016/j.jen.2010.02.020

[14] American College of Emergency Physicians (ACEP). Boarding of admitted and intensive care patients in the emergency department. policy statement. Ann Emerg Med 2011; 58(1): 110. http://dx.doi.org/10.1016/j.annemergmed.2011.04.027

[15] Sun Y, Heng BH, Tay SY, *et al*. Predicting hospital admissions at emergency department triage using routine administrative data. Acad Emerg Med. 2011; 18(8): 844-50. http://dx.doi.org/10.1111/j.1553-2712.2011.01125.x

[16] Dexheimer JW, Leegon J, Aronsky D. Predicting hospital admission at triage in an emergency department. AMIA Annu Symp Proc 2007; 937.

[17] Meisel ZF, Pollack CV, Mechem CC, *et al*. Derivation and internal validation of a rule to predict hospital admission in prehospital patients. Prehosp Emerg Care 2008; 12(3): 314-9. http://dx.doi.org/10.1080/10903120802096647

[18] van der Wulp I, Schrijvers AJ, van Stel HF. Predicting admission and mortality with the emergency severity index and the manchester triage system: A retrospective observational study. Emerg Med J 2009; 26(7): 506-9. http://dx.doi.org/10.1136/emj.2008.063768

[19] Leegon J, Jones I, Lanaghan K, *et al*. Predicting hospital admission for emergency department patients using a bayesian network. AMIA Annu Symp Proc 2005; 1022.

[20] Beardsell I, Robinson S. Can emergency department nurses performing triage predict the need for admission? Emerg Med J 2011; 28(11): 959-62. http://dx.doi.org/10.1136/emj.2010.096362

[21] Kosowsky JM, Shindel S, Liu T, *et al*. Can emergency department triage nurses predict patients' dispositions? Am J Emerg Med 2001; 19(1): 10-4.
http://dx.doi.org/10.1053/ajem.2001.20033

[22] Holdgate A, Morris J, Fry M, Zecevic M. Accuracy of triage nurses in predicting patient disposition. Emerg Med Australas 2007; 19(4): 341-5.
http://dx.doi.org/10.1111/j.1742-6723.2007.00996.x

[23] Price TG, Hooker EA, Neubauer J. Prehospital provider prediction of emergency department disposition: Implications for selective diversion. Prehosp Emerg Care 2005; 9(3): 322-5.
http://dx.doi.org/10.1080/10903120590962012

[24] Brillman JC, Doezema D, Tandberg D, *et al*. Triage: Limitations in predicting need for emergent care and hospital admission. Ann Emerg Med 1996; 27(4): 493-500.
http://dx.doi.org/10.1016/S0196-0644(96)70240-8

[25] Ruger JP, Lewis LM, Richter CJ. Identifying high-risk patients for triage and resource allocation in the ED. Am J Emerg Med 2007; 25(7): 794-8.
http://dx.doi.org/10.1016/j.ajem.2007.01.014

[26] Marshall AH, McClean SI. Using coxian phase-type distributions to identify patient characteristics for duration of stay in hospital. Health Care Manag Sci 2004; 7(4): 285-9.
http://dx.doi.org/10.1007/s10729-004-7537-z

[27] Faddy M, Graves N, Pettitt A. Modeling length of stay in hospital and other right skewed data: Comparison of phase-type, gamma and log-normal distributions. Value Health 2009; 12(2): 309-14.
http://dx.doi.org/10.1111/j.1524-4733.2008.00421.x

[28] Faddy MJ, McClean SI. Using a multi-state model to enhance understanding of geriatric patient care. Aust Health Rev 2007; 31(1): 91-7.
http://dx.doi.org/10.1071/AH070091

[29] Faddy MJ, McClean SI. Markov chain modelling for geriatric patient care. Methods Inf Med 2005; 44(3): 369-73.

[30] Faddy MJ, McClean SI. Analysing data on lengths of stay of hospital patients using phase-type distributions. Appl Stochastic Models Bus Ind 1999; 15(4): 311-7.
http://dx.doi.org/10.1002/(SICI)1526-4025(199910/12)15:4<311::AID-ASMB395>3.0.CO;2-S

[31] McClean SI, Faddy MJ, Millard PH. Markov Model-Based Clustering for Efficient Patient Care. CBMS 2005; 467-472.

[32] Marshall AH, McClean SI, Shapcott CM, *et al*. Modelling patient duration of stay to facilitate resource management of geriatric hospitals. Health Care Manag Sci 2002; 5(4): 313-9.
http://dx.doi.org/10.1023/A:1020394525938

[33] Marshall AH, McClean SI, Millard PH. Addressing bed costs for the elderly: A new methodology for modelling patient outcomes and length of stay. Health Care Manag Sci 2004; 7(1): 27-33.
http://dx.doi.org/10.1023/B:HCMS.0000005395.77308.d1

[34] Xie H, Chaussalet TJ, Millard PH. A continuous time markov model for the length of stay of elderly people in institutional long-term care. J Royal Statist Soc 2005; 168: 51-61.
http://dx.doi.org/10.1111/j.1467-985X.2004.00335.x

[35] Murray M, Bullard M, Grafstein E, *et al*. Revisions to the canadian emergency department triage and acuity scale implementation guidelines. CJEM 2004; 6(6): 421-7.

[36] Hodgins MJ, Moore N, Legere L. Who is sleeping in our beds? factors predicting the ED boarding of admitted patients for more than 2 hours. J Emerg Nurs 2011; 37(3): 225-30.
http://dx.doi.org/10.1016/j.jen.2010.02.020

[37] Powell ES, Khare RK, Venkatesh AK, *et al*. The relationship between inpatient discharge timing and emergency department boarding. J Emerg Med 2012; 42(2): 186-96.
http://dx.doi.org/10.1016/j.jemermed.2010.06.028

[38] Arkun A, Briggs WM, Patel S, *et al*. Emergency department crowding: Factors influencing flow. West J Emerg Med 2010; 11(1): 10-5.

[39] R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing 2011.

[40] Cox DR. Analysis of binary data. 2nd ed. Boca Raton: Chapman & Hall 1989.

Received on 30-10-2012 Accepted on 30-11-2012 Published on 31-01-2013

http://dx.doi.org/10.6000/1929-6029.2013.02.01.07