

# Identification of the risk factors associated with ICU mortality

## Abstract

The objective of this study is to identify the risk factors that influence the surgical and medical Intensive Care Unit (ICU) mortality. We considered data that was collected at Bay State Medical Center in Springfield, Massachusetts.<sup>1</sup> We developed statistical models that identify the risk factors associated with ICU mortality. In order to identify the risk factors without subjective bias, we explored multiple variable selection methods. We explored several methods including what we call manually picked best model, forward selection, backward elimination, and the least absolute shrinkage and selection operator (LASSO). We applied 5-fold cross validation on the final model of manually picked best model, forward selection and backward elimination and applied both validation set approach and 5-fold cross validation on LASSO to create confusion matrices and calculate the error rate of each method. Finally, we recommended the model for predicting ICU mortality with lowest misclassification error rate.

Volume 6 Issue 1 - 2017

**Nasser Abdullah K Alghamdi, Munni Begum**

Department of Mathematical Sciences, Ball State University, USA

**Correspondence:** Munni Begum, Department of Mathematical Sciences, Ball State University, Muncie, IN 47306, USA, Tel 1-765-285-8673, Email mbegum@bsu.edu

**Received:** May 21, 2016 | **Published:** June 15, 2017

## Introduction

The care and treatment of critically ill patients in special rooms with life-saving technology is the major component of modern medical science. The diagnosis and treatment of the patients in critical conditions is highly dependent on invasive diagnostic as well as therapeutic procedures. However, the main disruption of host defense mechanisms comes from the life support systems.

According to Morandi, Jackson, and Wesley,<sup>2</sup> there are ICU-acquired infections that is responsible for the high mortality rate of the ICU patients. The researchers' study offers a useful information concerning the topic. The aim of their study was to determine the epidemiology as well as the risk factors for nosocomial infections and the mortality rate in the ICU. Due to the variations of the study methods, the infection rates from different ICUs are difficult to compare.

Girard, Pandharipande, & Ely<sup>3</sup> show that delirium, a fluctuating disturbance of cognition, is a sign of acute brain dysfunction in the patients with critical illnesses in the ICU. The patients with critical illnesses are more likely to have delirium. Also, Morandi, Jackson, & Wesley Ely<sup>2</sup> show that delirium can be related to the cognitive impairment that persists for an extended period after discharge. Techniques of treating delirium at the ICU has been the subject of investigations in the recent past.

Age is also suggested as one of the predictors of mortality in the ICU. The number of elderly patients who are being admitted to the ICU has been increased, not only in the USA but also internationally (Belayachi et al.,<sup>4</sup>) There are few studies that have been conducted to link old age with the ICU mortality rate. The current research includes age as one of the risk factors for ICU mortality.

## Background

The ICU mortality rate in any hospital is the highest compared to other units. The United States of America has an approximate 4 million ICU admissions annually and the mortality rate of 500,000 deaths every year. The medical errors occur in any unit of the hospital, but it more likely to occur in the ICU since the ICU patients undergo complex interventions.

A study on the mortality risk factors and validation of severity scoring systems in the critically ill patients with acute renal failure was conducted to identify the determinants for improving patient care. Renal failure has a high prevalence in the ICU and associated with high mortality rates. Identification of the mortality risk factors helps to address intervention to these risk factors and improves patient care (Lima, Zanetta, Castro & Yu.<sup>5</sup>)

Iwuafor, et al.,<sup>6</sup> conducted a study sought to determine the prevalence, risk factors, clinical outcome, and the microbiological profile of the hospital-acquired infections in the ICU of a Nigerian hospital. Infections commonly affect critically ill patients and have a high association with mortality. The study identified blood stream infections and the urinary tract infections as a significant risk factors associated with the ICU mortality.

## Data and variable description

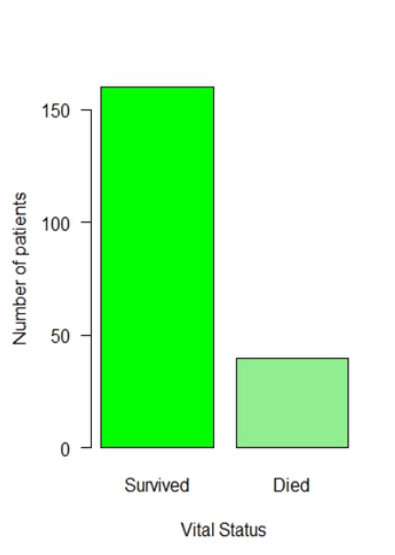
We considered data collected at Bay State Medical Center in Springfield, Massachusetts that can be downloaded from University of Massachusetts, Amherst website.<sup>1</sup> The dataset consists of 200 observations with 20 variables. The response variable (STA), vital status is categorical. The other categorical predictor variables are: Gender, Race, SER (Service at ICU Admission), CAN (Cancer Part

of Present Problem), CRN (History of Chronic Renal Failure), INF (Infection Probable at ICU Admission), CPR (CPR Prior to ICU Admission), PRE (Previous Admission to an ICU within 6 Months), TYP (type of admission), FRA (Long Bone, Multiple, Neck, Single Area, or Hip Fracture), PO2 (PO2 from Initial Blood Gases), PH (PH from Initial Blood Gases), PCO (PCO2 from initial Blood Gases), BIC (Bicarbonate from Initial Blood Gases), CRE (Creatinine from Initial Blood Gases) and LOC (Level of Consciousness at ICU Admission). The continuous predictor variables are Age, SYS (Systolic Blood Pressure at ICU Admission) and HRA (Heart Rate at ICU Admission).

Table 1 shows the total number of ICU patients according to vital status. It shows 80% of patients survived and 20% died. Figure 1 shows the graphical representation of these statistics.

**Table 1** Vital status of ICU patients

Status frequency			Status percentage		
Survived	Died	Total	Survived (%)	Died (%)	Total (%)
160	40	200	80	20	100



**Figure 1** Number of ICU patients according to Vital Status.

The table above shows that, 24 males died (12%) out of 124 total males and 16 females died (8%) out of 76 females at the ICU. From these results, it clearly demonstrates that males are more vulnerable to the ICU mortality as compared to females. The study also sought to determine whether ethnicity is associated with mortality. From the table above, 37 (18.5%) of 175 Whites died, 1 (0.5%) out of 15 Blacks died, and 2 (1%) out of 10 others died. The survival rate for the black patients and other races was less (7% and 4% respectively) compared to that of the white patients (69%). However, there wasn't sufficient data in the Black and Other category to make decisive comparison. We can see 46.5% of patients were medically treated at ICU compared to 53.5% of patients who were treated surgically. There was no incidence of cancer in 90% of ICU patients. Out of those, 18% died. In addition, 2% of patients who had cancer died. The above results demonstrate that the cancer is a low predictor of ICU mortality for the

critically ill patients admitted to the ICU. According to these results, the presence of cancer can lead to the survival of the patient in the ICU. We notice that only 9.5% of patients had chronic renal failure compared to 95.5% who did not have that failure. The results above show that the history of chronic renal disease is not a risk factor for mortality of the critically ill patients admitted to the ICU. For the patients with a history of chronic renal failure, the mortality was 4%, which is four times lower compared to those who did not have a history of chronic renal failure which was 16%. We notice that 12% of patients who had infection at ICU admission died compared to 8% without infection who died. Infection at ICU admission is a useful predictor of mortality in the intensive care unit with those infected having a higher likelihood of death compared to the uninfected ones. Also 6.5% of ICU patients had CPR prior to ICU admission. The percentage of patients with CPR prior to the ICU admission who died is low compared to the percentage of patients without CPR. It can be seen 15% of patients had previous admission to an ICU within 6 months. The previous admission to the intensive care unit is not a predictive factor for the mortality rate in the ICU. The percentage of people with previous admissions to the ICU who died is low compared to those who had not been admitted before. We can see that there are more deaths associated with emergency admission compared to the elective admission. From the above results, we can see that the mortality rate is 19% for the emergency admission as compared to the 1% of elective admission. The only 15 patients had fracture and 3 of them died. In contrast 37 patients out of 185 who did not have fracture died. The majority of patients at ICU had PO2 from initial blood gases greater than 60. The mortality rate for patients whose PO2 is greater than 60 is 17.5%. Patients whose initial blood gases' PH was higher than 7.25 showed higher mortality rate that was twice that of the patients whose initial blood gases' PH was below 7.25. In addition, the mortality rate for patients whose PCO2 from initial blood gases less than 45 is higher than those whose PCO2 is greater than 45. The mortality rate of patients whose Bicarbonate from initial blood gases was greater than 18 is seven times the mortality rate of patients whose Bicarbonate from initial blood gases was less than 18. It can be seen the ICU patients whose creatinine level from initial blood gases was greater than 2.0 have a higher mortality rate. The results above show that the mortality rate of ICU patients who had deep stupor or coma was high. Patients who had no coma survived with a probability of more than 85%.

The results in Table 3 demonstrate that age plays a significant role in the admission of patients to the ICU.

Table 3 and Figure 2 show that patients' age is between 16 – 92 years old and the majority of ICU patients are between 40 – 80 years old. In addition, the boxplot shows the mean age of died patients is 70.

Figure 3 shows that the majority of patients were at risk who had systolic blood pressure between 120 to 140. The mean of systolic blood pressure of the patients at ICU is 132.3 mmHg. The boxplot shows a patient who had 256 mmHg died.

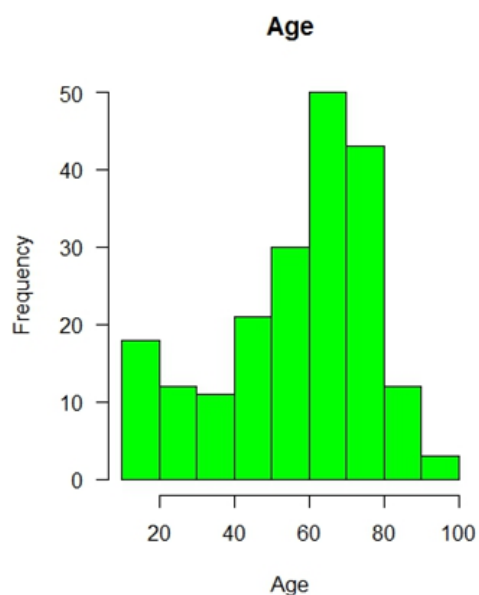
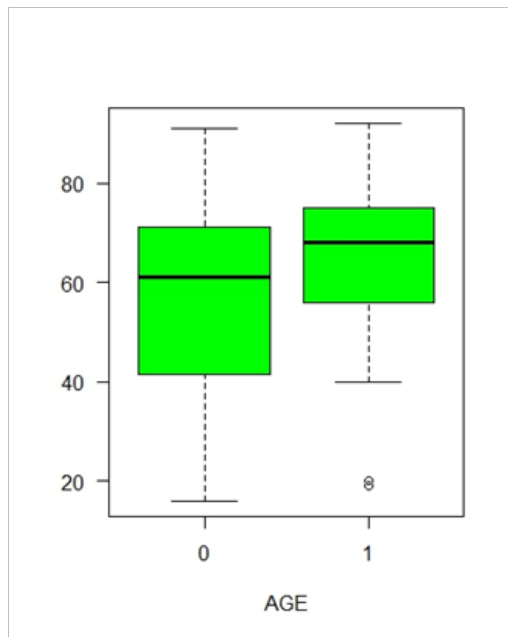
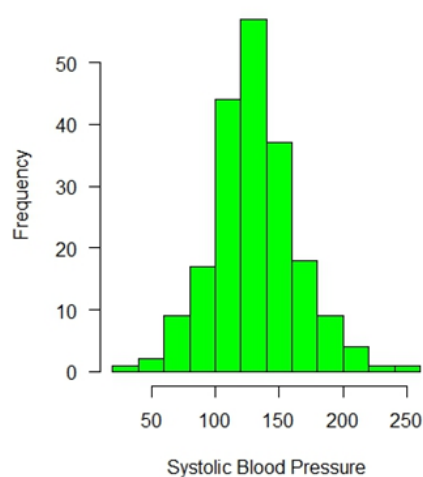
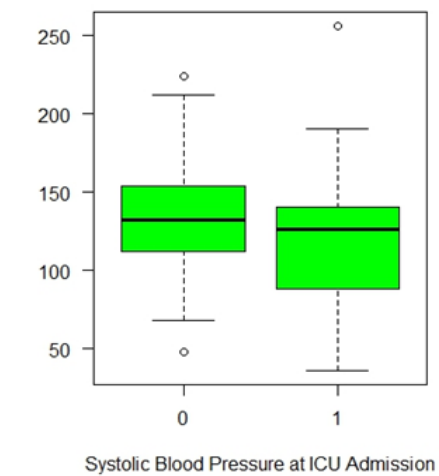
Figure 4 shows the admission to the ICU is higher for the patients with high heart rate with the mean heart rate during the first day of admission being 98/min. The heart rate, however, depends on various factors such as the patient's mood, body temperature, physical activities, etc.

**Table 2** Categorical variables and vital status of ICU patients

		Status frequency			Status percentage		
		Survived	Died	Total	Survived (%)	Died (%)	Total (%)
Gender	Male	100	24	124	50	12	62
	Female	60	16	76	30	8	38
Race	White	138	37	175	69	18.5	87.5
	Black	14	1	15	7	0.5	7.5
	Other	8	2	10	4	1	5
Service	Medical	67	26	93	33.5	13	46.5
	Surgical	93	14	107	46.5	7	53.5
Cancer	No	144	36	180	72	18	90
	Yes	16	4	20	8	2	10
Chronic	No	149	32	181	74.5	16	90.5
	Yes	11	8	19	5.5	4	9.5
Infection	No	100	16	116	50	8	58
	Yes	60	24	84	30	12	42
CPR	No	154	33	187	77	16.5	93.5
	Yes	6	7	13	3	3.5	6.5
PRE	No	137	33	170	68.5	16.5	85
	Yes	23	7	30	11.5	3.5	15
Type	Elective	51	2	53	25.5	1	26.5
	Emergency	109	38	147	54.5	19	73.5
Fracture	No	148	37	185	74	18.5	92.5
	Yes	12	3	15	6	1.5	7.5
PO2	> 60	149	35	184	74.5	17.5	92
	< 60	11	5	16	5.5	2.5	8
PH	> 7.25	151	36	187	75.5	4.5	80
	< 7.25	9	4	13	18	2	20
PCO2	< 45	144	36	180	72	18	90
	> 45	16	4	20	8	2	10
BIC	> 18	150	35	185	75	17.5	92.5
	< 18	10	5	15	5	2.5	7.5
CRE	< 2.0	155	35	190	77.5	17.5	95
	> 2.0	5	5	10	2.5	2.5	5
LOC	No coma	158	27	185	79	13.5	92.5
	Deep stupor	0	5	5	0	2.5	2.5
	Coma	2	8	10	1	4	5

**Table 3** Continuous variables summary

	Summary					
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Age	16	46.75	63	57.54	72	92
Systolic blood pressure	36	110	130	132.3	150	256
Heart rate	39	80	96	98.92	118.2	192

**Figure 2** Age and vital status.**Figure 3** Systolic blood pressure and vital status.

### Objective identification of risk factors of ICU mortality

Since the response variable in our data (vital status) is binary, binary logistic regression is an appropriate model to consider. Logistic regression is a predictive analysis technique used to illustrate the relationship between a binary response variable and the predictors using the regression on the logarithm of the odds of having a response:<sup>7,8</sup>

Let  $Y_i$  be the binary response variable for  $i^{th}$  patient, with

$Y_i = 0$ , the patient survived

$Y_i = 1$ , the patient died

then, logit

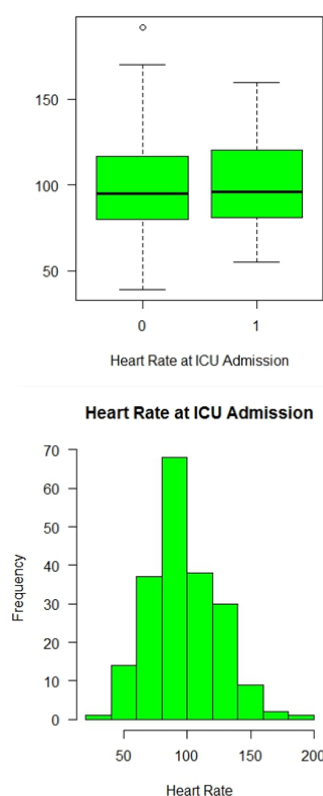
$$\left(\pi_i\right)=\log \left(\frac{\pi_i}{1-\pi_i}\right)$$

$$\beta_0+\beta_1 x_{i1}+\ldots+\beta_k x_{ik} \approx \mathcal{X} \beta$$

Where,

$$\pi_i=\frac{e^{\mathcal{X} \beta}}{1+e^{\mathcal{X} \beta}}$$

We consider multiple variable selection methods in order to identify risk factors in an objective manner. These are manually picked best model, forward selection, backward elimination and least absolute shrinkage and selection operator (LASSO). We discuss these methods briefly as follows.



**Figure 4** Heart rate at ICU admission and vital status.

### Manually picked best model

In order to identify the risk factors for the vital status at ICU, which is a binary response variable, we started with the process of manually picked best model. To correctly leverage this method, we begin with binary logistic regression model. First, we fit a model with all predictor variables. Next we pick the significant variables (in this case risk factors) with the smallest P-value ( $P < .10$ ) manually and remove all the factors which are insignificant. We then refit the model with all the significant variables in the model.

### Forward selection

For forward selection, a null model (a model with no predictors), serves as the starting point. We add one variable at a time to the null model and refitted the model including the added variables. The idea was to keep it if the variable that had been added was significant and then add the next variable. If not, we eliminated it and added the next variable. We refitted the model using the same procedure until the stopping rule was satisfied (all the variables in the model are significant).

### Backward selection

Backward elimination method starts with a full model that contains all the predictors in the model. The least significant risk factors; that is, the ones having the largest P value (greater than 10%) are eliminated, and the model is then refitted. Each step removes the least significant variable from the model until the remaining variables have their P values smaller than the specified 0.10.

### Least absolute shrinkage and selection operator (LASSO)

With this method, there is an automatic selection of predictors of the target variable from the large set of potential predictors. By doing so, the method will return the coefficients of the irrelevant variables to zero thereby performing an automatic selection of variables. The LASSO formulates a curve fitting as a quadratic programming problem with the objective function that penalizes the absolute size of the coefficients based on a value of a tuning parameter, say  $\lambda$ . The method, therefore, shrinks the size of the nonzero coefficients and ends up with the most useful variables.

$$\text{logit}(\pi_i) = \log \left( \frac{\pi_i}{1-\pi_i} \right)$$

$$\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \text{LASSO penalty}$$

$$\approx \mathcal{X} \beta + \text{LASSO penalty}$$

$$\text{LASSO penalty} = \lambda \sum_{j=1}^p |\beta_j|$$

Where,

$$\pi_i = \frac{e^{\mathcal{X} \beta + \lambda \sum_{j=1}^p |\beta_j|}}{1 + e^{\mathcal{X} \beta + \lambda \sum_{j=1}^p |\beta_j|}}$$

$$\pi_i > 0.5, \hat{y}_i = 1 \text{ (dead)}$$

With the predicted probability of the binary response,  $\hat{\pi}_i$  we can predict the response itself using the above cut point.

Thus, if the predicted probability of dying is greater than 0.5, we coded the predicted response as 1 (died) and 0 (survived) otherwise.

### Validation and k-fold cross validation approach

Validation/cross validation approach is an objective methodology to select an optimal strategy. To select the best model from the variable

selection models selected by four strategies, we implement validation and cross validation. We applied k-fold cross validation approach on the final model of manually picked best model, forward selection and backward elimination to create the confusion matrices and calculate the error rate of each method. Since we have 200 observations, we decided to use 5 folds which split the data into two; training and testing datasets. Training set had 160 observations and testing set had 40 observations (total 5 sets each having 40 observations). k-fold cross-validation approach is applicable where the original sample is partitioned at random into k subsamples and one is left out in every iteration step. Let k parts be  $C_1, C_2, \dots, C_K$ , where  $C_i$  denotes the indices of the observations in part  $i$ . We have the following formula to estimate error rate:

$$CV = \sum_{i=1}^K \frac{n_i}{n} (MSE_i)$$

Where  $MSE_k = \sum_{i \in C_i} \left( y_i - \hat{y}_i \right)^2 / n_i$ ,  $\hat{y}_i$  is the fit for observation  $i$ , and  $n_i = n/K$ . For this study,  $n=200$ ,  $K=5$ . So, there are 200/5 parts of 40. The data is split to two groups of testing and training: testing = 40 and training = 160. We fit logistic models on the training data sets and calculate misclassification error rate on the test data. In

addition, we conducted k- fold cross validation and validation set on LASSO and compared the results of each method. Since we have 200 observations, we decided to use validation set which split the data into two; training and testing datasets. Training set had 100 observations and testing set had 100 observations (total 2 sets each having 100 observations).

The validation set error rate is determined using:

$$CV = \frac{1}{n} \sum_{i=1}^n \left[ \frac{e_i}{1-h_i} \right]^2$$

Where  $e_i$  is the residual obtained from fitting a model to all the n observations.

## Results and discussion

### Manually picked best model

In this method, we fit a model with all predictor variables. We pick the significant variables (in this case risk factors) with the smallest P-value ( $P < 0.10$ ) manually and remove all the factors which are insignificant. We then refit the model with all the significant variables in the model. Table 4 shows the significant and insignificant predictors of the manually picked best model.

**Table 4** Model 1 (all variables)

	Estimate	Std. error	z value	P-value	Significance
(Intercept)	-5.4138	2.121735	-2.552	0.010723	
AGE	0.052432	0.017283	3.034	0.002416	Significant
GENDER	-0.55496	0.501191	-1.107	0.268174	
RACE	-0.00455	0.528253	-0.009	0.993131	
SER	-0.54534	0.575341	-0.948	0.3432	
CAN	2.757739	0.980444	2.813	0.004912	Significant
CRN	-0.10191	0.762361	-0.134	0.893656	
INF	-0.05607	0.534729	-0.105	0.916489	
CPR	0.977885	0.983498	0.994	0.32008	
SYS	-0.01154	0.007854	-1.47	0.141573	
HRA	-0.00364	0.009343	-0.39	0.69674	
PRE	0.929282	0.628867	1.478	0.139486	
TYP	2.744267	0.995251	2.757	0.005827	Significant
FRA	1.152077	0.999318	1.153	0.248967	
PO2	0.38667	0.851053	0.454	0.649582	
PH	2.41525	1.231101	1.962	0.049779	Significant
PCO	-3.17282	1.385963	-2.289	0.022065	Significant
BIC	-0.79384	0.916184	-0.866	0.386239	
CRE	0.233314	1.075334	0.217	0.828233	
LOC	2.706087	0.751819	3.599	0.000319	Significant



We refit the model with the significant predictors (risk factors in this case) and eliminate the insignificant predictors all together.

Table 5 below shows the final model with all significant variables. It shows that the risk factors age, cancer part of present problem, type of admission, PH from Initial blood gases, PCO2 from initial blood gases and Level of consciousness at ICU admission as the risk factors.<sup>9,10</sup>

$$= -6.75128 + 0.04018 \text{ AGE} + 2.14668 \text{ CAN} + 2.81592 \text{ TYP} + 1.76830 \text{ PH} + (-2.13254) \text{ PCO} + 2.30890 \text{ LOC} \approx \chi\beta$$

**Table 5** Model 2 (final model of manually picked best model)

	Estimate	Std. error	z value	P-value	Significance
(Intercept)	-6.75128	1.29885	-5.198	2.02E-07	
AGE	0.04018	0.01311	3.066	0.00217	Significant
CAN	2.14668	0.84582	2.538	0.01115	Significant
TYP	2.81592	0.89512	3.146	0.00166	Significant
PH	1.7683	0.85459	2.069	0.03853	Significant
PCO	-2.13254	0.98844	-2.157	0.03097	Significant
LOC	2.3089	0.57504	4.015	5.94E-05	Significant

$$\pi_i = \frac{e^{\chi\beta}}{1 + e^{\chi\beta}}$$

Considering the median age =63, presence of cancer =1, type of admission was elective =0, had no coma or stupor =0, and keep the rest factors fixed, we found:

$$\pi_i = \frac{e^{\chi\beta}}{1 + e^{\chi\beta}} = 11.2\% \text{ chance of mortality.}$$

Considering the same as above except the type of admission was emergency =1, we found:  $\pi_i = \frac{e^{\chi\beta}}{1 + e^{\chi\beta}} = 67.8\% \text{ chance of mortality.}$

In addition, if type of admission is emergency =1, had coma =2, and keep the remaining factors fixed we found the chance of mortality increased from 67.8% to 99.5%.

### Forward selection model

For forward selection, a null model (a model contains no predictors), served as the starting point. We added one variable at a time to the null model and refitted the model including the added variables. The idea was to keep it if the variable that had been added was significant and then add the next variable. If not, we eliminated it and added the next variable. We refitted the model using the same procedure until the stopping rule was satisfied (all the variables in the model are significant meeting the level of 10%).

The table above shows the final model of forward selection method. It shows that the risk factors age, type of admission and level of consciousness at ICU admission statistically significant for the ICU status.

From table 6, our model equation can be written as:

$$\text{logit} \left( \pi_i \right) = \log \left( \frac{\pi_i}{1 - \pi_i} \right)$$

From table 5, our model equation can be written as:

$$\begin{aligned} \text{logit} \left( \pi_i \right) &= \log \left( \frac{\pi_i}{1 - \pi_i} \right) \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 - \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 \approx \chi\beta \end{aligned}$$

$$\begin{aligned} &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\ &= -5.52063 + 0.03291 \text{ AGE} + 2.18842 \text{ TYP} + 1.83445 \text{ LOC} \\ &\approx \chi\beta \end{aligned}$$

Considering the median age =63, type of admission was elective =0, had no coma or stupor =0, and keep the rest factors fixed, we found:

$$\pi_i = \frac{e^{\chi\beta}}{1 + e^{\chi\beta}} = 3.09\% \text{ chance of mortality.}$$

Considering the same as above except the type of admission was emergency =1, we found:  $\pi_i = \frac{e^{\chi\beta}}{1 + e^{\chi\beta}} = 22.12\% \text{ chance of mortality.}$

### Backward elimination model

In this method, we began with the full model which includes all predictors in the model and eliminate variables one at a time. The least significant risk factors; that is, the ones having the largest P value (greater than 10%) are eliminated, and the model is then refitted.

Table 7 shows the final model of backward elimination method. It shows that the risk factors age, cancer part of present problem, systolic blood pressure at ICU admission, type of admission, PH from initial blood gases, PCO2 from initial blood gases, and level of consciousness at ICU admission statistically significant for the ICU status.

From table 7, backward elimination model equation can be written as:

$$\begin{aligned} \text{logit} \left( \pi_i \right) &= \log \left( \frac{\pi_i}{1 - \pi_i} \right) \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 \end{aligned}$$

$$= -5.278880 + 0.040425 \text{ AGE} + 2.164740 \text{ CAN} + (-0.010994) \text{ SYS} + 2.753050 \text{ TYP} \\ + 1.809602 \text{ PH} + (-2.297444) \text{ PCO} + 2.343905 \text{ LOC} \\ \approx \mathcal{X}\beta$$

**Table 6** Final model of forward selection

	Estimate	Std. error	z value	P-value	Significance
(Intercept)	-5.52063	1.09373	-5.048	4.48E-07	
AGE	0.03291	0.01179	2.791	0.005247	Significant
TYP	2.18842	0.76276	2.869	0.004117	Significant
LOC	1.83445	0.51609	3.555	0.000379	Significant

**Table 7** Final model of backward elimination method

	Estimate	Std. error	z value	P-value	Significance
(Intercept)	-5.27888	1.55063	-3.404	0.000663	
AGE	0.040425	0.013084	3.09	0.002004	Significant
CAN	2.16474	0.853723	2.536	0.011224	Significant
SYS	-0.01099	0.006753	-1.628	0.103512	Significant
TYP	2.75305	0.909096	3.028	0.002459	Significant
PH	1.809602	0.874858	2.068	0.038598	Significant
PCO	-2.29744	1.027075	-2.237	0.025294	Significant
LOC	2.343905	0.618393	3.79	0.00015	Significant

Considering the median age =63, presence of cancer =1, median systolic blood pressure =130, type of admission is elective =0, had no coma or stupor=0, and keep the rest factors fixed, we found:

$$\pi_i = \frac{e^{\mathcal{X}\beta}}{1 + e^{\mathcal{X}\beta}} = 11.96\% \text{ chance of mortality.}$$

Considering the same as above except the type of admission was emergency =1, we found:  $\pi_i = \frac{e^{\mathcal{X}\beta}}{1 + e^{\mathcal{X}\beta}} = 68.06\% \text{ chance of mortality.}$

### Least absolute shrinkage and selection operator (LASSO)

We fit LASSO with validation set approach. The results are presented in Table 8.

Table 8 shows the final model of LASSO (validation set). We can see type of admission, and level of consciousness at ICU admission as the risk factors.

We also fitted LASSO with applying 5-fold cross validation approach. Table 9 has the results of this approach.

Table 9 shows the final model of LASSO (5-fold cross validation). Cancer part of present problem, previous admission to an ICU within 6 months, type of admission, PH from initial blood gases, PCO2 from initial blood gases, and level of consciousness at ICU admission are identified as the risk factors by this approach.

### Misclassification error rate

The cross-validation approach allows to compute misclassification error rate by calculating the confusion matrix. Tables 10 – 14 present confusion matrices for manually picked best model, forward selection model, backward elimination and LASSO.

**Table 8** Final model of LASSO (validation set)

	Coefficient
(Intercept)	0.031276
AGE	0.001951
GENDER	
RACE	
SER	
CAN	0.001493
CRN	0.017163
INF	0.018822
CPR	
SYS	-0.00054
HRA	
PRE	
TYP	0.111793
FRA	
PO2	
PH	
PCO	
BIC	
CRE	0.020118
LOC	0.277094



**Table 9** Final model of LASSO (5-fold cross validation)

	coefficient
(Intercept)	-0.09257
AGE	0.003733
GENDER	-0.02657
RACE	
SER	-0.02175
CAN	0.178869
CRN	0.016181
INF	0.02257
CPR	0.036143
SYS	-0.00092
HRA	
PRE	0.068878
TYP	0.191538
FRA	0.019172
PO2	0.006562
PH	0.137662
PCO	-0.15182
BIC	
CRE	0.035652
LOC	0.329276

From the confusion matrix in table 10, we calculate the misclassification error rate of manually picked best model as  $\left(\frac{7+68}{200}\right) * 100 = 37.5\%$ .

**Table 10** Manually picked best model confusion matrix

	Actual	
Predicted	0	1
0	92	7
1	68	33

From the confusion matrix in table 11, we calculate the misclassification error rate of forward selection model as  $\left(\frac{4+80}{200}\right) * 100 = 42.0\%$ .

**Table 11** Forward selection confusion matrix

	Actual	
Predicted	0	1
0	80	4
1	80	36

From the confusion matrix in table 12, we calculate the misclassification error rate of backward elimination model as  $\left(\frac{7+68}{200}\right) * 100 = 37.5\%$ .

**Table 12** Backward elimination confusion matrix

	Actual	
Predicted	0	1
0	92	7
1	68	33

From the confusion matrix in table 13, we calculate the misclassification error rate of LASSO under validation set approach as  $\left(\frac{30+2}{200}\right) * 100 = 16.0\%$ .

**Table 13** LASSO confusion matrix "validation set"

	Actual	
Predicted	0	4
0	158	30
1	2	10

From the confusion matrix in table 14, we calculate the misclassification error rate of LASSO under 5-fold cross-validation as  $\left(\frac{27+2}{200}\right) * 100 = 14.5\%$ .

**Table 14** LASSO confusion matrix "5-fold"

	Actual	
Predicted	0	1
0	158	27
1	2	13

From the results above we conclude that, LASSO (with applying 5-fold cross validation) is the best model for identifying the risk factors associated with the ICU mortality with the lowest error rate (14.5%). We can see cancer part of present problem, previous admission to an ICU within 6 months, type of admission, PH from initial blood gases, PCO2 from initial blood gases, and level of consciousness at ICU admission as the risk factors.<sup>11,12</sup>

## Conclusion

The major objective of this study is to identify the risk factors associated with medical and surgical ICU mortality. In order to identify the risk factors without subjective bias, we considered different variable selection methods and recommended the method that had the lowest misclassification error rate. The variable selection methods considered in this study were manually picked best model, forward selection, backward elimination and least absolute shrinkage and selection operator (LASSO). Cross validation and validation set approach are applied to the final model of manually picked best model, forward selection, backward elimination, and conducted both validation set and 5-fold cross validation on LASSO. Validation set and 5-fold cross validation approaches allow us to calculate the

misclassification error rates for each method and finalize the decision by choosing the model with the lowest misclassification error rate. The procedure determines a reliable model that would identify the risk factors associated with ICU mortality, in an objective manner.

From the results obtained in this study we recommend LASSO (with applying 5-fold cross validation) as the best model that identifies the risk factors associated with the ICU mortality since it has the lowest error rate (14.5%). The model identified cancer part of present problem, previous admission to an ICU within 6 months, type of admission, PH from initial blood gases, PCO<sub>2</sub> from initial blood gases, and level of consciousness at ICU admission as the risk factors. One limitation of this study is that the methodology is applied to a limited publicly available data on ICU mortality from a single hospital. In order to confirm the results of this study an elaborative study on ICU mortality should be performed on a randomly selected hospitals throughout the country.

## Acknowledgement

None.

## Conflict of interest

None.

## References

1. <https://www.umass.edu/statdata/statdata/data>
2. Morandi A, Jackson JC, Wesley Ely E. Delirium in the intensive care unit. *International Review of Psychiatry*. 2009;21(1):43–58.
3. Girard TD, Pandharipande PP, Ely EW. Delirium in the intensive care unit. *Critical Care*. 2008;12(S3).
4. Belayachi J, Dendane T, Madani N, et al. Factors predicting mortality in elderly patients admitted to a Moroccan medical intensive care unit. *Southern African Journal of Critical Care*. 2012;28(1):22–27.
5. Lima EQ, Zanetta DMT, Castro I, et al. Mortality risk factors and validation of severity scoring systems in critically ill patients with acute renal failure. *Renal failure*. 2005;27(5):547–556.
6. Iwuafor AA, Ogunsola FT, Oladele RO, et al. Incidence, Clinical Outcome and Risk Factors of Intensive Care Unit Infections in the Lagos University Teaching Hospital (LUTH), Lagos, Nigeria. *PLoS One*. 2016;11(10):e0165242.
7. Bursac Z, Gauss CH, Williams DK, et al. Purposeful selection of variables in logistic regression. *Source code for biology and medicine*. 2008;3(1):17.
8. Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression: Third Edition. 2013;528.
9. <http://www.mayoclinic.org/healthy-lifestyle/fitness/expert-answers/heart-rate/faq-20057979>
10. <https://www.cdc.gov/bloodpressure/measure.htm>
11. Leong IYO, Tai DYH. Is Increasing Age Associated with Mortality in the Critically ill Elderly. *Singapore Med J*. 2002;43(1):33–36.
12. Toufen C, Franca SA, Okamoto VN, et al. Infection as an independent risk factor for mortality in the surgical intensive care unit. *Clinics*. 2013;68(8):1103–1108.