Multivariate time series analysis of ICU mortality

600.446 Computer Integrated Surgery II

Project Report

Student: Danning He

Mentors: Dr. Jim Fackler, Dr. Harold Lehmann

Abstract

The objective of this project is to develop a patient-specific mortality prediction model based on physiologic derangement during first 48h of an ICU stay. Under a probabilistic framework, the risk features are defined as log likelihood ratio, and aggregated by a logistic function to generate a probability score. The classification performance of the method is evaluated, contribution from each individual feature is analyzed, and finally, limitations and possible extensions of this study is discussed.

Introduction

The modern ICU is a complex, expensive and resource-intensive environment, admitted patients are usually under life-threatening conditions that require advanced medical care and invasive/noninvasive monitoring. The cost of care for an ICU patient is estimated to be three times the costs of a general patient [1]. Therefore, the primary focus of ICU is on patients whose extreme conditions can be reversed and who have good chances of surviving. Since adjusted mortality rate is a useful marker of ICU quality, tools that quickly and accurately make prediction of a patient's mortality risk of great significance. It allows for better clinical decision-making by the physicians and helps control hospital expenses and manages medical resources. Various general acuity-scoring systems are used for patients with critical illness, and can be calibrated from admission status (age, reasons for admission, previous health status, and etc.), physiological variables, laboratory tests, organ dysfunction, as well as therapeutic intervention (nursing activities, the amount/types of care provided) [2-4].

Motivation and Significance

Instead of focusing on the overall mortality rates in ICU, which vary across units and hospitals due to different patient cohorts and clinical context [5], the goal of our project is to predict patient-specific in-hospital mortality/survival based on 5 general descriptors and 37 time-series physiological variables during the first 48 hours of an ICU stay. However, in many cases, a common difficulty from such high dimensional heterogeneous data is its irregular measurement in terms of time and frequency (the variables are measured from once 30 minutes to once several hours, and not all of them are taken for each individual), Figure 1 corresponds to multiple types of observations in the same patient. In our dataset, out of 4000 samples, only 28 of them have at least one observation for all variables. Moreover, the real-world processes produce series of measureable observations (physiological profiles) as a function of underlying hidden states (latent disease states), the measurement is noisy and

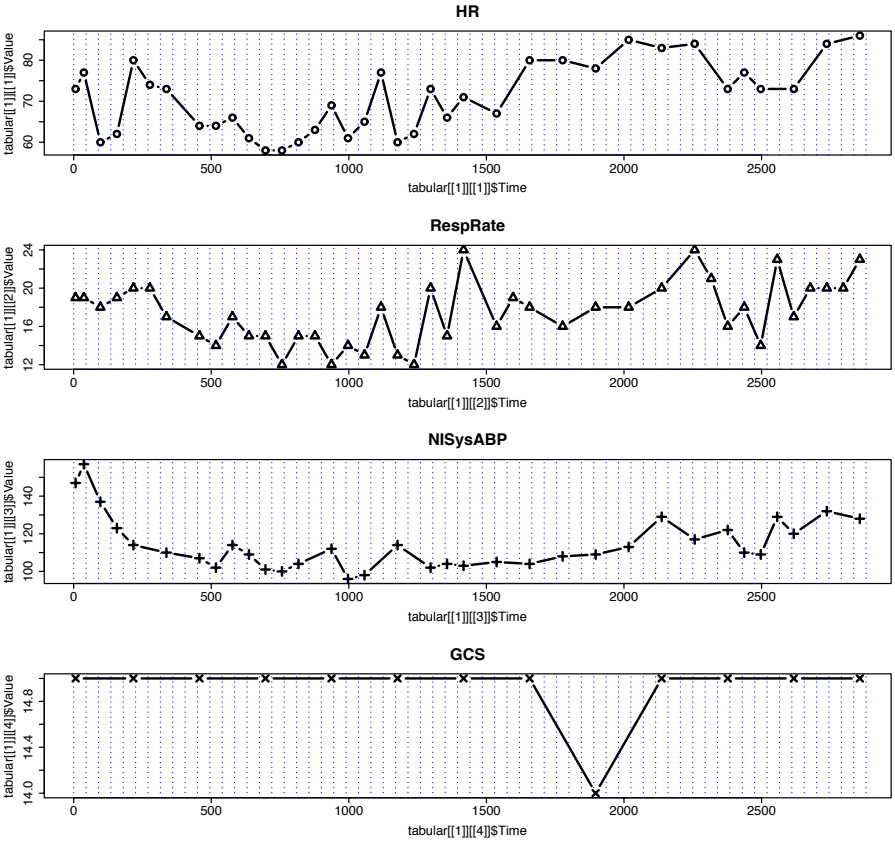


Figure 1. Multiple types of variables measured in the same patient, x-axis: time, y-axis: value

Correlated, e.g. pH should go down as the PaCO2 rises, and the HCO3 should rise as the PaCO2 rises in order to bring the pH back to normal, SysABP (Invasive systolic arterial blood pressure) and NISysABP

(Non-invasive systolic arterial blood pressure) both measure the systolic arterial blood pressure, but using different techniques. PCA (principal component analysis) is a commonly used method to identify the most representative variations, which is a linear combination of original variables, and we observe several highly correlated variables, e.g. PaCO2 and PaO2, NISysABP and SysABP, as is shown in Figure 2. For the task of outcome prediction, majority of existing scoring systems are based on such supervised algorithms as logistic regression or artificial neural networks. Since neural network is unstable to different parameter settings (e.g. the number of hidden nodes, decay factor) in our study, we only focus on logistic regression.

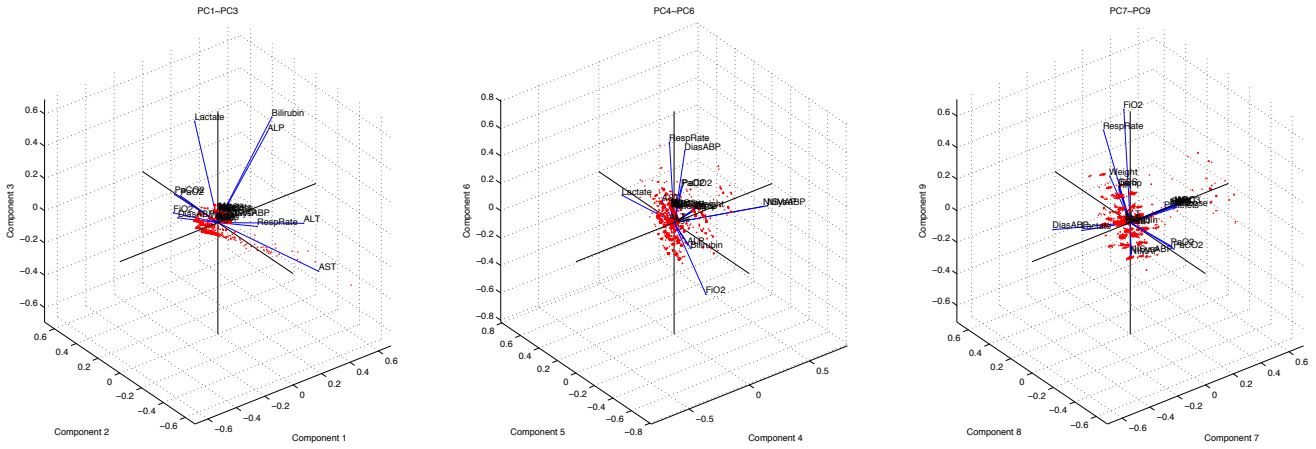


Figure 2. Principle component analysis, each red point corresponds to all records associated with a patient, PC1~PC9 is shown, which accounts for 95% of the total variations.

Technical Approach

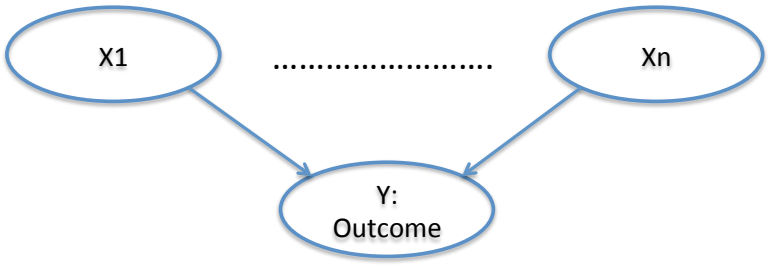


Figure 3. Graphical illustration of logistic regression models

The logistic function models how the probability of an event Y is affected by multiple explanatory variables, $\{\mathbf{x}_1,...,\mathbf{x}_n\}$. Although the event is binary, its probability is a latent variable that generates the observed outcome, e.g. survive/die [6, 7].

The probability of 'death' given original variables (X) can be written as a logistic sigmoid acting on the linear combination of the feature vector $\phi = \phi(X)$ so that

$$P(death \mid \{\mathbf{x}_1,...,\mathbf{x}_M\}) = \sigma(-\mathbf{w}^T\phi) = \frac{1}{1+\exp(-\mathbf{w}^T\phi)},$$

Where $\phi(\cdot)$ is the basis function that transforms original variables X into feature vectors, $\mathbf{w} = (w_0, w_1, ..., w_M)^T$ is model parameter, $M+1$ and is the total number of features? $\mathbf{w}$ Is solved via minimizing the error function, L2 regularization term $\lambda \sum_i w_i^2$ is introduced, the result is not affected with λ ranges from 1.2~1.5.

We use a probabilistic framework to define the basis function, ϕ, which is primarily based on previously developed nonlinear model of risk factors [8]. There're several assumptions of this transformation: 1) The true probability distribution, P_i^{death} or $P_i^{survive}$, of each original variable, $\mathbf{x}_i$, in each class of patients can be approximated by one of five long-tail parametric distributions (normal, gamma, exponential, weibull and lognormal) using maximum likelihood estimation, each variable has multiple observations, $\{o_{i1},...,o_{iT}\}$, within 48 hours—which means the underlying distribution, P_i^{death}, is invariant over time. 2) The number of observations, T_i, follows Poisson distribution, and is independent of the values taken by the observations. 3) If a measurement is completely missing in one patient, $T_i = 0$, it's assumed $P(T_i = 0 \mid death)$ equals $E_{Data}[\mathbf{1}\{T_i = 0\} \mid death]$ the expectation of such event in the 'death' patient population.

$$\phi(\mathbf{x}_i) = \begin{cases} \log \dfrac{P(\mathbf{x}_i \mid death)}{P(\mathbf{x}_i \mid survive)}, \log \dfrac{P(T_i \mid death)}{P(T_i \mid survive)} \\ \log \dfrac{P(T_i = 0 \mid death)}{P(T_i = 0 \mid survive)} \end{cases}$$

Where $\log \frac{P(\mathbf{x}_i \mid death)}{P(\mathbf{x}_i \mid survive)} = \sum_{t=1}^{T} \log \frac{P(o_{it} \mid death)}{P(o_{it} \mid survive)}$ and $\{o_{i1},...,o_{iT}\}$ is the set of observations for a variable $\mathbf{x}_i$.

Ten-fold cross validation is used to evaluate classification performance. 90% samples are used to fit distributions and learn model parameters, while 10% held-out samples are used to evaluate predictive accuracy of the classifier. The ROC (Receiver operating characteristic) curve plotted true positive rate (TPR) vs. the false positive rate (FPR) as the cutoff threshold changes, which indicates the accuracy of the prediction method at various cutoff thresholds used to discriminate survive vs. death. The AUC (Area Under the Curve) is equal to the probability that the classifier assigns a higher score for a randomly selected positive instance than that for a randomly selected negative one [9].

Results

Classification evaluation

The left figure shows the ROC curve and associated area under the curve values for our method (10-fold cross-validation) and SAPS-I and SOAP (based on all available samples). The overall classification performance of our method (AUC=0.801) is better than SAPS-I (AUC=0.638) and SOAP (AUC=0.628). The mortality rate in our data is $554/4000 \approx 0.138$, but a threshold of 0.05 achieves a sensitivity of 0.945 and specificity of 0.351 (marked out by an arrow in Figure 4). Since $sensitivity = \frac{TP}{TP + FN}$ and $specificity = \frac{TN}{FP + TN}$, this means if we predict patients with $P(death) > 0.05$ to die and those with $P(death) < 0.05$ to survive, we correctly identify 94.5% patients that died and 35.1% patients that survived. Alternatively, threshold of 0.5 (marked out by an arrow in Figure 4) that assigns patients with $P(death) > 0.5$ to die, and those with $P(death) < 0.5$ to survive correctly identify 98.0% patients that survived but only 23.6% patients that died. The use of lower threshold improves sensitivity at the price of specificity, falsely rejecting some of the patients whose life-threatening conditions could have been reversed with intensive care, while the higher threshold increases specificity at the price of sensitivity,

falsely accepting some of the patients that won't have good chance of survive even with advanced treatment.

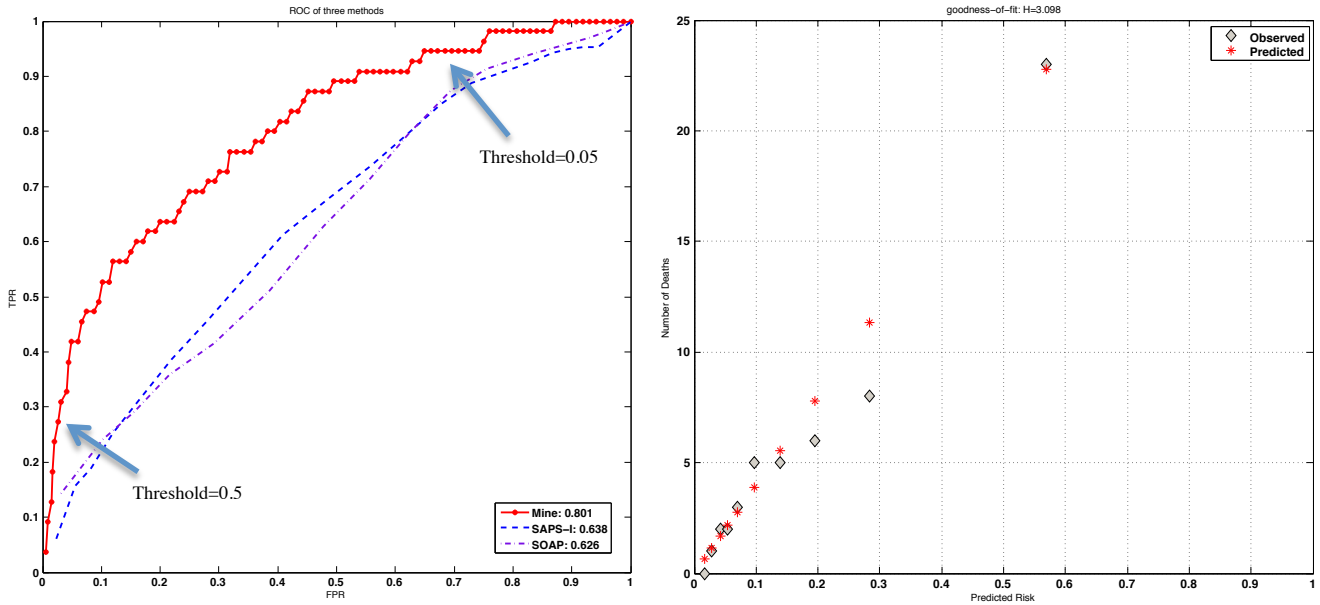


Figure 4. Left: classification performance evaluation using ROC and associated AUC. Right: Hosmer–Lemeshow statistics.

Goodness of fit

The right figure plotted the Hosmer–Lemeshow statistics [7], which tests for goodness of fit for logistic regression [10]. The population is discretized into several subgroups based on the predicted risk, and the test assesses whether the observed number of events match the expected number of events in different risk deciles,

$$H = \sum_{d=1}^{D} \frac{(O_d - E_d)^2}{N_d \pi_d (1 - \pi_d)},$$

Where O_d, E_d, N_d and π_d correspond to the observed events, expected evens, observations and predicted risk for the d^{th} risk deciles (D=10). The test statistic follows asymptotically the chi-squired distribution, with $D-2$ degree of freedom, while an ideal and unattainable score is 0, which corresponds to 0 cumulative probability, our H statistics equals 3.10, corresponds to 0.0719 cumulative probability, the maximum accepted H statistics for this model is 15.507, corresponds to 0.95 cumulative probability.

Non-linear transformation of risk factors

The basis function $\phi(\cdot)$ defined previously is the log-likelihood ratio of two statistical models, which expresses how many times more likely the set of observations $\{o_{i1},...,o_{iT}\}$ are under one model ('death') than the other ('survive'). The model specifies both the family of distribution (one of five long-tail distributions for the observed values and Poisson distribution for the number of observations) and the parameters that maximize the observation likelihood.

The logistic classifier consists of 24 time-series variables and 2 general descriptors (Table 1), the unselected variables are either missing in majority of the patients or highly correlated with the already selected variables (e.g. NIMAP and MAP, NISysABP and SysABP). Besides examining the learned weights $\mathbf{w}$ of features (Table 1), representing relative contribution from both the values and the number of observations for each physiological variable to the observed outcome, the Bayesian modeling enables the nonlinear transformation from values of physiological parameters to the posterior probability of outcome (e.g. from $P(\mathbf{x}_i \mid death)$ to $P(death \mid \mathbf{x}_i)$), visualization of such transformation is shown in Figure 5 (only for values of observations),

$$P(death \mid \mathbf{x}_i) = \frac{P(\mathbf{x}_i \mid death) \cdot P(death)}{P(\mathbf{x}_i \mid death) \cdot P(death) + P(\mathbf{x}_i \mid survive) \cdot P(survive)}$$

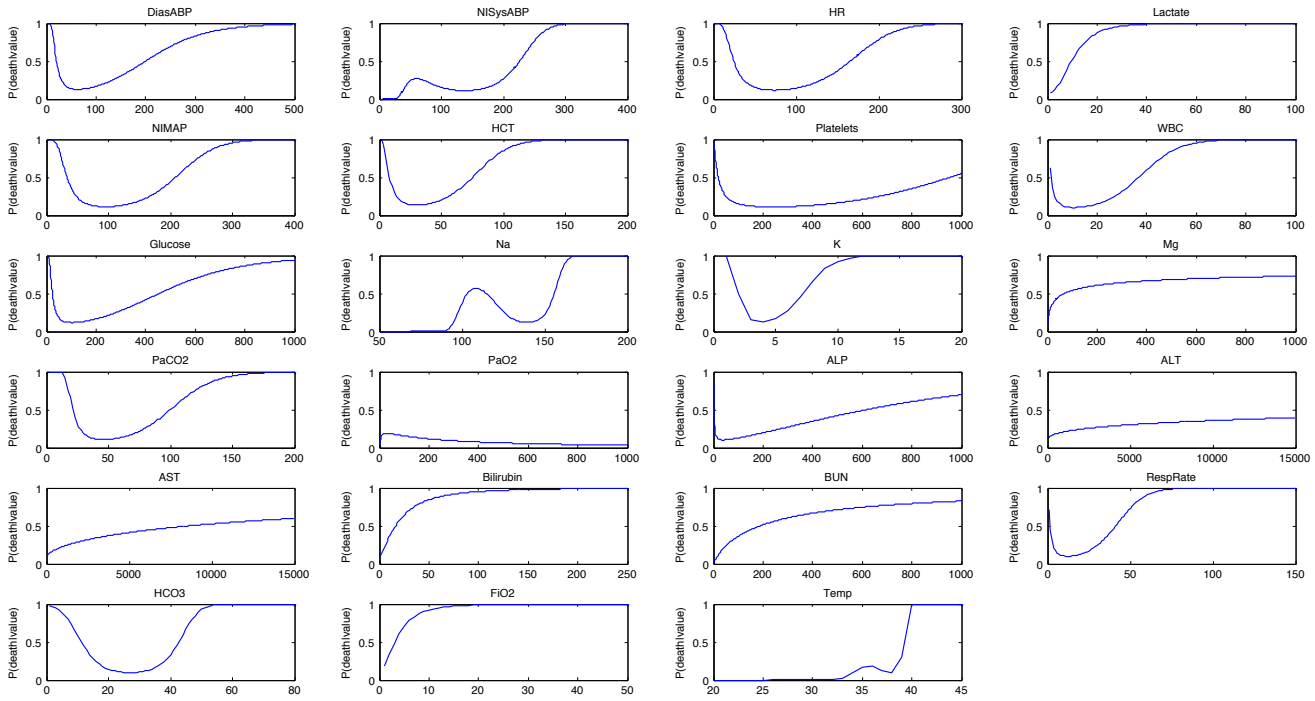


Figure 5. the nonlinear transformation from values of physiological parameters to the posterior probability of death

vname	weight_value	weight_count
'DiasABP'	0.11	-0.24
'NISysABP'	0.14	0.25
'HR'	0.14	0.15
'Lactate'	0.22	0.03
'NIMAP'	0.16	-0.10
'HCT'	0.06	-0.01
'Platelets'	0.17	-0.27
'WBC'	0.13	0.09
'Glucose'	-0.01	0.35
'Na'	0.01	0.36
'K'	0.06	0.03
'Mg'	0.06	-0.05
'PaCO2'	0.02	-0.11
'PaO2'	0.12	-0.08
'ALP'	0.06	-0.05
'ALT'	-0.05	0.42
'AST'	0.09	0.14
'Bilirubin'	0.21	-0.49
'BUN'	0.40	-0.06

'RespRate'	0.19	0.18
'HCO3'	0.10	-0.65
'FiO2'	0.19	-0.10
'Temp'	0.30	0.20
GCS	-0.78*	0.13
Age	0.19*	-
Weight	-0.18*	-

Table 1. Parameter weights for logistic regression, *: using average value

Discussion and Conclusion

The observed physiological signals and their dynamics over time are affected by many factors, from the intrinsic state of disease, the setup of the monitoring instruments, to the medical interventions received by the patients [11, 12]. Most importantly, our data is a mixture of different patient cohorts, having different baseline characteristics, but detailed information regarding their origins is not available. Also, the disease state manifests itself through physiology, which is measured by various digital equipment, and our observations are, again, a mixture of all these confounding variables. However, the logistic model only adopts the strongest assumption (Figure 3): there's only one latent variable Y controlling the patient's binary outcome, all the physiological variables are independent and directly affect the outcome, and the model remains exactly the same over time and across patients. Further investigation can be related to eliminate some of the assumptions and generalize the model.

Metrics for scoring the critically ill have several purposes: 1) patient outcome prediction, 2) measure disease severity, 3) resource management, 4) ICU performance assessment in different cohorts [13, 14]. Due to limited information in the data, our study only modeled the impact of physiological status to the patient-specific observed outcome, and such intrinsic genetic factors usually change slowly over decades or even centuries. However, outcome-predicting model in terms of accuracy is also strongly affected by population characteristics and healthcare delivery systems, which is changing continuously and become more and more important, justifying the need to 'reinvent the wheel' from time to time. Given the diversity and complexity of medical interventions we can offer today, the physiological impact is actually much lower compared with what it used to be in the past, justifying the definition of 'death' as those whose extreme conditions cannot be reversed. Moreover, in terms of pragmatic usage, prediction

models differ in the number and types of variables required have different data collection burden [15]. All of the above points are worth further investigation in the future.

Management Summary

Project Timeline

Timeline		Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13
Milestone 1 (By 2/23)	reading list	done												
	project plan													
	preprocessing data		done											
Milestone 2 (By 3/26)	Features as Log likelihood ratio				done									
	Logistic regression													
	AUC and ROC													
Milestone 3 (By 5/1)	Feature Analysis									done				
	Time dependencies between observations									Substitute with PCA				
	Features from hidden states									Hold on due to limitations of the data itself				
Milestone 4 (By 5/8)	Model optimization												done	
	partial AUC												ROC threshold analysis	
	Documentation												done	

Project Deliverables

Minimum (Done)

-- Logistic regression with log likelihood ratios as risk features

-- Performance evaluation: ROC, AUC

Expected

-- Feature analysis (Done)

-- Try features constructed from standard HMM, Kalman Filter (Hold on due to limitations of the data itself)

-- Incorporate dependencies between observations **(Substitute with PCA)**

Maximum

-- Optimize features to achieve better classification performance (Done)

-- Documentation (Done)

-- partial AUC **(Substitute with ROC threshold analysis)**

All work is implemented in MATLAB.

Acknowledgements

Thanks to Dr. Jim Fackler and Dr. Harold Lehmann for their clinical insights, and thanks to Dr. Suchi Saria for her guidance in methodology.

Reference

1. Cooper LM, WT L-Z: **Medicare intensive care unit use: analysis of incidence, cost, and payment.** *Crit Care Med* 2004, **32**:2247-2253.
2. Zimmerman JE, Kramer AA, McNair DS, FM M: **Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients.** *Crit Care Med* 2006, **34**:1297-1310.
3. Timsit JF, Fosse JP, Troche G, De Lassence A, Alberti C G-O, Bornstain C, Adrie C, Cheval C, S C: **Calibration and discrimination by daily Logistic Organ Dysfunction scoring comparatively with daily Sequential Organ Failure Assessment scoring for predicting hospital mortality in critically ill patients.** *Crit Care Med* 2002, **30**:2003-2013.
4. Padilha KG, de Sousa RM, Queijo AF, Mendes AM, MD R: **Nursing Activities Score in the intensive care unit: analysis of the related factors.** *Intensive Crit Care Nurs* 2008, **24**:197-204.
5. Glance LG, Osler T, T S: **Effect of varying the case mix on the standardized mortality ratio and W statistic: A simulation study.** *Chest* 2000, **117**:1112-1117.
6. http://en.wikipedia.org/wiki/Logistic_regression.
7. David W. Hosmer, Lemeshow, S: **Applied Logistic Regression, Second Edition**. *Wiley Series in Probability and Statistics Texts and References Section* 2005.
8. Saria S, Rajani AK, Gould J, Koller D, Penn AA.: **Integration of early physiological responses predicts later illness severity in preterm infants.** *Sci Transl Med* 2010, **2**(48):48ra65.
9. Fawcett, Tom: **An introduction to ROC analysis**. *Pattern Recognition Letters* 2006, **27**:861–874.
10. http://en.wikipedia.org/wiki/Hosmer%E2%80%93Lemeshow_test.
11. S. Saria, D. Koller, Penn A: **Learning individual and population level traits from Clinical Temporal data**. *Neural Information Processing Systems, Predictive Models in Personalized Medicine workshop* 2010.

12. John A. Quinn, Christopher K.I. Williams, Neil McIntosh: **Factorial Switching Linear Dynamical Systems applied to Physiological Condition Monitoring**. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2009, **31**(9):1537-1551.

13. Vincent JL, Moreno R.: **Clinical review: scoring systems in the critically ill.** *Crit Care* 2010, **14**(2):207.

14. Metnitz PG, Moreno RP, Almeida E, Jordan B, Bauer P CR, Iapichino G, Edbrooke D, Capuzzo M, Le Gall JR; SAPS 3 Investigators.: **SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description.** *Intensive Care Med* 2005, **31**(10):1336-1344.

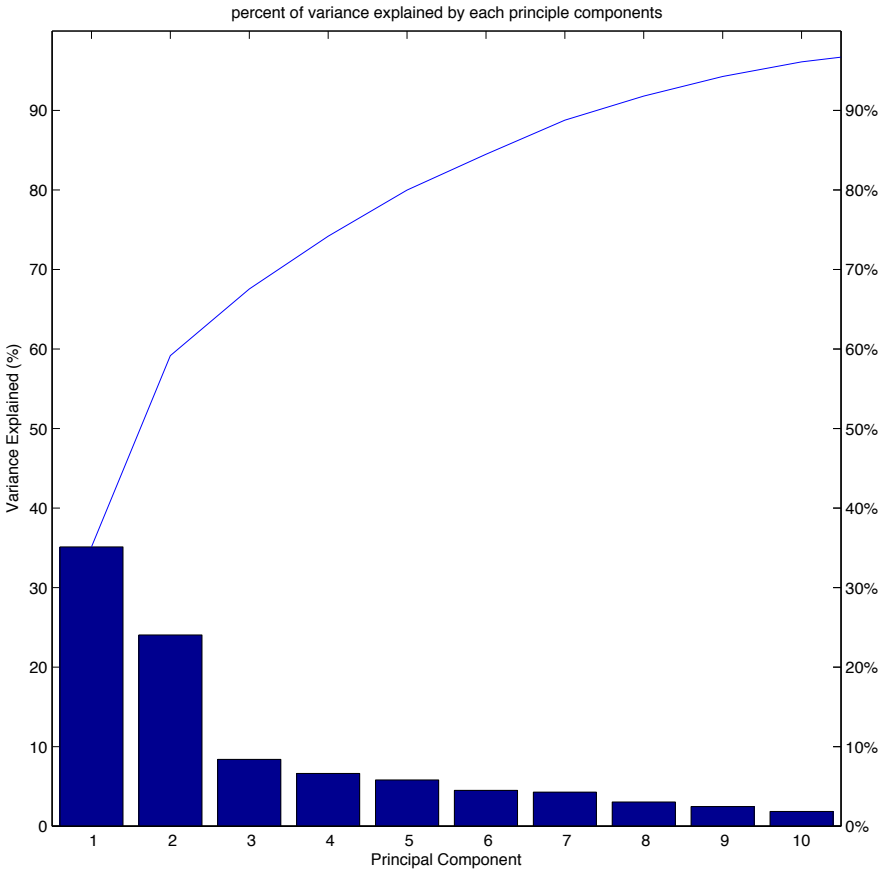
15. Kuzniewicz MW, Vasilevskis EE, Lane R DM, Trivedi NG, Rennie DJ CT, Kotler PL, Dudley RA: **Variation in ICU risk-adjusted mortality: impact of methods of assessment and potential confounders.** *Chest* 2008, **133**(6):1319-1327.

Appendix

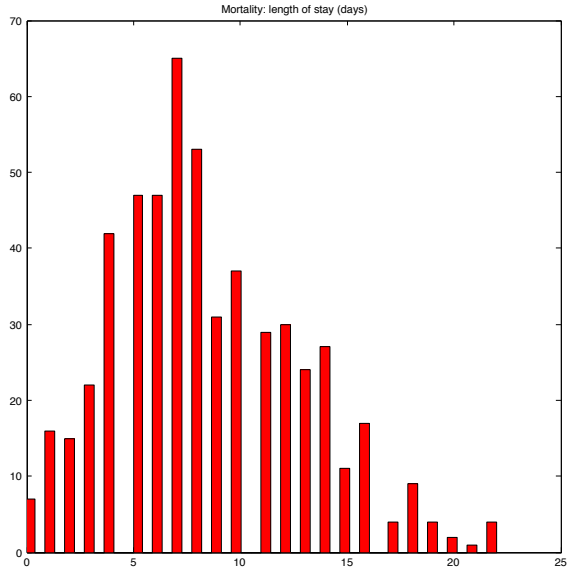
1. Overview of original variables:

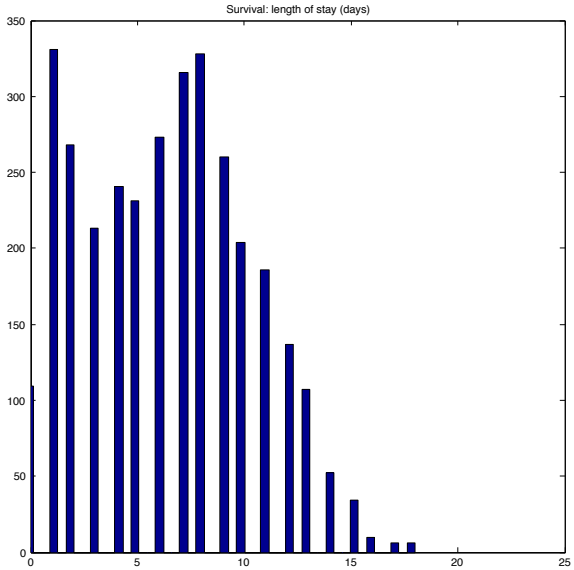
- **Albumin** (g/dL)
- **ALP** [Alkaline phosphatase (IU/L)]
- **ALT** [Alanine transaminase (IU/L)]
- **AST** [Aspartate transaminase (IU/L)]
- **Bilirubin** (mg/dL)
- **BUN** [Blood urea nitrogen (mg/dL)]
- **Cholesterol** (mg/dL)
- **Creatinine** [Serum creatinine (mg/dL)]
- **DiasABP** [Invasive diastolic arterial blood pressure (mmHg)]
- **FiO2** [Fractional inspired O_2 (0-1)]
- **GCS** [Glasgow Coma Score (3-15)]
- **Glucose** [Serum glucose (mg/dL)]
- **HCO3** [Serum bicarbonate (mmol/L)]
- **HCT** [Hematocrit (%)]
- **HR** [Heart rate (bpm)]
- **K** [Serum potassium (mEq/L)]
- **Lactate** (mmol/L)
- **Mg** [Serum magnesium (mmol/L)]
- **MAP** [Invasive mean arterial blood pressure (mmHg)]
- **MechVent** [Mechanical ventilation respiration (0:false, or 1:true)]
- **Na** [Serum sodium (mEq/L)]
- **NIDiasABP** [Non-invasive diastolic arterial blood pressure (mmHg)]
- **NIMAP** [Non-invasive mean arterial blood pressure (mmHg)]
- **NISysABP** [Non-invasive systolic arterial blood pressure (mmHg)]
- **PaCO2** [partial pressure of arterial CO_2 (mmHg)]
- **PaO2** [Partial pressure of arterial O_2 (mmHg)]
- **pH** [Arterial pH (0-14)]
- **Platelets** (cells/nL)
- **RespRate** [Respiration rate (bpm)]
- **SaO2** [O_2 saturation in hemoglobin (%)]
- **SysABP** [Invasive systolic arterial blood pressure (mmHg)]
- **Temp** [Temperature (°C)]
- **TropI** [Troponin-I (μg/L)]
- **TropT** [Troponin-T (μg/L)]
- **Urine** [Urine output (mL)]
- **WBC** [White blood cell count (cells/nL)]
- **Weight** (kg)*

2. Percent of variance explained by each principle components



3. Waiting time until death (left figure) or discharge (right figure)





4. Maximum likelihood estimation of parametric distributions for each selected variable

vname	Mortality	param1.M	param2.M	Survival	param1.S	param2.S	poisson.M	poisson.S
'DiasABP'	'logn'	4.06	0.22	'logn'	4.07	0.21	40.22	36.20
'NISysABP'	'logn'	4.73	0.21	'gam'	28.78	4.15	24.14	24.39
'HR'	'gam'	19.25	4.71	'gam'	23.93	3.63	59.92	56.74
'Lactate'	'logn'	1.12	0.79	'logn'	0.74	0.61	3.26	2.03
'NIMAP'	'gam'	20.83	3.55	'gam'	25.94	2.99	23.87	24.01
'HCT'	'gam'	35.02	0.88	'gam'	38.25	0.80	4.68	4.58
'Platelets'	'gam'	2.48	70.05	'gam'	3.80	50.63	3.83	3.51
'WBC'	'gam'	2.27	6.19	'gam'	4.11	3.01	3.55	3.21
'Glucose'	'logn'	4.93	0.41	'logn'	4.87	0.36	3.96	3.20
'Na'	'logn'	4.94	0.04	'norm'	139.08	5.05	4.14	3.30
'K'	'logn'	1.42	0.17	'logn'	1.40	0.15	4.33	3.54
'Mg'	'logn'	0.71	0.20	'logn'	0.68	0.20	3.91	3.38
'PaCO2'	'logn'	3.62	0.25	'logn'	3.69	0.21	6.94	5.83
'PaO2'	'logn'	4.78	0.49	'logn'	4.89	0.51	6.94	5.83
'ALP'	'logn'	4.67	0.70	'logn'	4.45	0.62	1.45	0.93
'ALT'	'logn'	4.54	1.71	'logn'	4.12	1.60	1.50	0.95
'AST'	'logn'	5.00	1.70	'logn'	4.45	1.49	1.50	0.95
'Bilirubin'	'logn'	0.62	1.49	'logn'	-0.01	1.11	1.49	0.95
'BUN'	'logn'	3.40	0.71	'logn'	2.97	0.71	4.01	3.41
'RespRate'	'gam'	11.79	1.82	'gam'	13.74	1.42	8.22	14.52
'HCO3'	'norm'	21.54	5.48	'norm'	23.43	4.42	3.96	3.33
'FiO2'	'logn'	-0.63	0.31	'logn'	-0.66	0.30	10.80	7.73
'Temp'	'norm'	36.92	1.13	'wbl'	37.47	42.58	19.89	22.12
GCS	NA	NA	NA	NA	NA	NA	16.15	15.32