Assessing Risk Factors and Predicting Sepsis Mortality Using Logistic and Survival Models

by

Yunfei Xie

BA, Mathematics, Luther College, 2015

Submitted to the Graduate Faculty of

Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Yunfei Xie

It was defended on

July 24, 2019

and approved by

Abdus S. Wahed, PhD, Professor of Biostatistics
Department of Biostatistics,
Graduate School of Public Health, University of Pittsburgh

Thesis advisor:
Jonathan Yabes, PhD, Assistant Professor of Medicine, Biostatistics
Department of Medicine, School of Medicine
Department of Biostatistics, Graduate School of Public Health,
University of Pittsburgh

Thesis Co-advisor:
Gilles Clermont, MD, Professor of Critical Care Medicine, Industrial Engineering and Mathematics, Department of Critical Care Medicine, School of Medicine
University of Pittsburgh

Jonathan Yabes, PhD
Gilles Clermont, MD

Assessing Risk Factors and Predicting Sepsis Mortality Using Logistic and Survival Models

Yunfei Xie, MS

University of Pittsburgh, 2019

Abstract

Identifying sepsis patients with high risk of death is crucial for making treatment decisions and has a public health significance. Sepsis mortality can be predicted by including clinical features and biomarkers in a predictive model. Hypotheses: (1) Clinical features combined with biomarkers would significantly enhance prediction power over clinical features alone; (2) time-trends of measurements contribute to prediction; (3) Cox proportional hazards model is more informative than logistic model.

Sepsis patients with complete data were identified from the Protocol-based Care in Early Septic Shock (ProCESS) trial. The trial obtained measurements at baseline (0 hours), 6 hours, and 24 hours of hospital admission, as well as patients' within-60-day-of-admission death time. To evaluate biomarkers, logistic regressions with biomarkers and clinical features were compared to logistic regressions with clinical features only. To assess trends, at each time point, trends variables were evaluated in logistic regressions. To compare statistical models, landmark mortality within 3-day, 7-day, 14-day, and 60-day of admission were modeled using logistic regressions; a Cox model was developed to predict mortality over the same period. Areas under the Receiver Operating Characteristic curve (AUC) with bootstrap confidence intervals (CI) were used to evaluate model performance.

There were 528 patients included in baseline cohort (60-day mortality: 25%, mean age: 60 years, mean baseline lactate: 2.41 mmol/L), 534 patients in 6 hours cohort (24%, 60, 2.35), and 432 patients in 24 hours cohort (21%, 60, 2.26). At baseline, the AUC increased significantly from 0.766 [95% CI] = [0.710, 0.826] to 0.812 [0.749, 0.868] when biomarkers were added. In all models, trends were nonsignificant. For logistic models, 3-day model has AUC 0.888 [0.836, 0.939]; 7-day model has AUC 0.827 [0.776, 0.879]; 14-day model has AUC 0.858 [0.820, 0.895]; and 60-day model has AUC 0.795 [0.716, 0.835]. For the Cox model, the time-dependent AUC ranges between (0.859, 0.884).

Biomarkers provided incremental discrimination ability over clinical features alone to predict 60-day mortality at baseline. Trends of time-dependent variables did not increase predictive power. Logistic models and Cox models have similar predictive power in predicting short-term mortality but a Cox model is better in predicting long-term mortality.

Table of Contents

1.0 Introduction **1**

2.0 Dataset And Methods **3**

2.1 Study Population **3**

2.2 Data **4**

2.3 Evaluating Biomarkers **5**

2.3.1 Data **5**

2.3.2 Methods of Evaluating Biomarkers **6**

2.3.3 Logistic Regression **7**

2.3.4 Elastic Net Regularization **7**

2.3.5 ROC Curve and AUC **10**

2.4 Evaluating Trends of Variables **11**

2.4.1 Data **11**

2.4.2 Methods for Evaluating Trends **11**

2.5 Comparing Statistical Models **12**

2.5.1 Data **12**

2.5.2 Methods for Comparing Models **12**

2.5.3 Cox Proportional Hazards Model **15**

2.5.4 Time-dependent AUC **15**

3.0 Results **17**

3.1 Patients **17**

3.2 Evaluation of Biomarkers **19**

3.2.1 Predictive Models 19

3.2.2 Predictive Power of Biomarkers 21

3.3 Evaluation of Time-dependent Variables 22

3.3.1 Predictive Power of Time-dependent Variables 22

3.3.2 Time-independent Variables in the Prediction 22

3.4 Comparison of Statistical Models 23

3.4.1 Predictive Models 23

3.4.2 Model Evaluation 26

4.0 Discussion 27

4.1 Risk Factors of Sepsis Mortality 27

4.2 Model Selection for Sepsis Mortality Prediction 28

4.3 Comparison with Existing Models 29

4.4 Limitations 30

5.0 Conclusion 32

Bibliography 33

List of Tables

Table 1 Baseline Characteristics for Cohorts Based on Survival Status 17

Table 2 Baseline Characteristics for Cohorts Based on Analysis Inclusion..... 18

Table 3 Logistic Models for Mortality Prediction at Different Times..... 20

Table 4 Logistic Models for Landmark Mortality Prediction..... 24

Table 5 Cox Model for 60-day Mortality Prediction..... 25

Table 6 Cox Model Time-dependent AUC..... 25

List of Figures

Figure 1 Flow Chart of Analysis 14

Figure 2 Confidence Intervals for AUCs of Logistic Regressions 21

Figure 3 Plot of Confidence Intervals of AUC 26

1.0 Introduction

Sepsis, as the body's life-threatening complication to an infection, remains a leading cause of death of in-hospital mortality (Angus et al., 2001). Immediate medical care and treatment is necessary for hospitalized patients with signs and symptoms of sepsis. Identifying sepsis patients with a high risk of death is crucial for making treatment decisions.

Many studies analyzed risk factors for sepsis mortality, most of them considered illness severity, comorbidities, and organ failures as predictor variables (Dellinger et al., 2004; Moreno et al., 2008; Ford et al., 2016). However, few of them have considered including inflammatory biomarkers to improve outcome prediction in sepsis patients. Biomarkers may help to diagnose sepsis, indicate stages of disease, and the prognosis of the patient. The clinical signs of sepsis, such as body temperature and blood pressure, are usually unspecific and also occur in noninfectious states (Reinhart et al., 2006). A variety of biomarkers obtained from laboratory tests can help physicians diagnose sepsis and assess the patient's clinical status from sepsis to severe sepsis (Faix, 2013). It is important to recognize biomarkers for the diagnosis and prognosis of sepsis, so that supportive measures can be implemented as soon as possible to reduce the risk of death.

In addition, existing studies have not considered time trends of clinical features or biomarkers in sepsis mortality prediction (Dellinger et al., 2004; Moreno et al., 2008; Ford et al., 2016). Many variables, such as vital signs and biomarkers, are collected over time. Changes in these time-dependent variables may be indicators of prognosis and provide more information.

Besides the selection of predictors, the type of model used could also influence the accuracy of the prediction. Different types of statistical models can be used to predict sepsis mortality (Kasal et al., 2004). The majority of published literature used logistic regression, which

is appropriate to assess associations and predict binary survival outcome; however, it ignores timing of the events in prospective studies. On the other hand, survival models, such as Cox proportional hazards regression, account for the times. Survival models are default methods to analyze time-to-event data in cohort studies. If there is no censoring, either model could be used to predict survival outcome at a fixed time horizon (Staley et al., 2017).

In summary, there are three major objectives of this study with respect to sepsis mortality prediction: (1) to evaluate the predictive power of biomarkers on top of clinical features, (2) to assess whether trends of measurements improve prediction, and (3) to compare the prediction performance of logistic regression and Cox regression. The predictive models were based on measurements at the baseline of hospital admission, and at 6 hours, 24 hours, and 72 hours after the admission. These models may prove useful in identifying sepsis patients with a high risk of mortality to aid in making treatment plans correspondingly.

2.0 Dataset And Methods

2.1 Study Population

We performed a secondary data analysis on data obtained from the Protocol-Based Care for Early Septic Shock (ProCESS) trial which enrolled patients from 31 hospitals in the United States (Process Investigators, 2014). Patients were recruited from the emergency department if sepsis was suspected according to the treating physician.

The enrollment eligibility requires patients had to be at least 18 years of age, who met two or more criteria for systemic inflammatory response syndrome (the criteria are: i. temperature >38° C or <36° C; ii. heart rate >90 beats per minute; iii. respiratory rate >20 breaths per minute or PaCO2 <32 mm Hg; and, iv. white blood cell count >12,000/mm, <4,000/mm, or >10% immature (band)forms), and who had refractory hypotension or a serum lactate level of 4 mmol per liter or higher. Patients did not have to be in shock on arrival of the emergency department but had to be enrolled in the study within 12 hours after the arrival and 2 hours after the earliest detection of shock (Process Investigators, 2014).

Patients were excluded in the study if they had: a primary diagnosis of acute cerebral vascular event, acute coronary syndrome, acute pulmonary edema, status asthmaticus, major cardiac arrhythmia, active gastrointestinal hemorrhage, seizure, drug overdose, burn or trauma; a requirement of immediate surgery; a known CD4 count < 50/mm; an advance directive that would restrict protocol implementation; a contraindication to central venous catheterization; a high likelihood of blood transfusion rejection; a treating physician who deemed resuscitation to be

futile; on-going participation in another interventional study; known pregnancy; or had been transferred from another hospital (Process Investigators, 2014).

2.2 Data

The original dataset of the ProCESS trial has 1341 subjects with variables including: demographic variables, chronic conditions, vital signs, and biomarkers. In-hospital mortality information was recorded and there were no censored observations.

Demographic variables and chronic conditions are considered as time-independent variables in the analysis. Demographic variables, including age, race, and ethnicity, were included in the analysis. Chronic conditions, such as cancer, immunosuppression status, and cirrhosis, were also included.

Two types of time-dependent variables were included in the analysis, they were vital signs and biomarkers. Vital signs, such as blood pressure and temperature, indicate severity of a patient's condition. Biomarker variables, such as lactate and cytokines, may help identify patients who are developing severe sepsis and thus may help to reduce the risk of death. Measurements occurred at the baseline of admission (0 hours), and at 6 hours, 24 hours, and 72 hours after admission. The unit of measurement for serum biomarkers is mmol/L.

2.3 Evaluating Biomarkers

2.3.1 Data

The variable 60-day mortality was the primary outcome of interest when evaluating biomarkers. All 1341 subjects from the ProCESS trial and 387 variables were initially included in this part of the analysis. Biomarkers were measured at different times. In order to evaluate the predictive power of biomarkers at different times, patients were selected into four cohorts based on their data availability at time of measurement.

Patients with complete data for the initial predictor set measured at time of admission (baseline time) were included in the baseline (0 hours) cohort. Patients with complete data for initial predictors measured at 6 hours, 24 hours, and 72 hours after the enrollment were enrolled in the 6 hours cohort, 24 hours cohort, and 72 hours cohort. Baseline characteristics at time of admission were reported in the baseline cohort, 6 hours cohort, 24 hours cohort, 72 hours cohort, and in the source population. Baseline characteristics were compared between one cohort and patients who were outside that cohort in the source population in order to assess potential selection bias. P-values for comparison were performed using the Student's t-test for normally distributed continuous variables, median test for non-normally distributed continuous variables, and the Chi-square test of independence for categorical variables.

2.3.2 Methods of Evaluating Biomarkers

To evaluate the predictive contribution of biomarkers, logistic regressions with clinical features only (LRC) and logistic regressions with both clinical features and biomarkers (LRCB) were constructed and compared at different hour-based cohorts. For non-normally distributed biomarker variables, such as TMB, they were log transformed.

Before predictor selection, variables with over 80% missing values were first dropped. Univariate logistic regression was then employed for each variable, with 60-day mortality as the outcome. Variables with significant p-value below 0.05 from univariate logistic regression were selected into the initial predictor set.

Logistic regressions with clinical features only (LRC) and logistic regressions with both clinical features and biomarkers (LRCB) were developed for each cohort. Specifically, for the baseline cohort, 6 hours cohort, and 24 hours cohort, since there was a large number of initial predictors, elastic net regularization (Friedman et al., 2010) was additionally applied to refine the initial predictor set. The cohorts were randomly split into training (67%) and testing (33%) sets. While the training set was used for model development, the testing set was used for model validation. Multivariable logistic regression was fitted in the training set. As for the 72 hours cohort, the number of predictors and sample size was small due to a great number of missing data. Hence, the initial set of predictors was determined based on the univariate analysis only. The final logistic models were derived after the backward elimination based on the Akaike Information Criterion (AIC). When fitting logistic regression models with clinical features alone (LRC), only clinical predictors were included; when fitting logistic regression models with both clinical features and biomarkers (LRCB), biomarkers were added to the selected clinical predictors in LRC.

Area under the Receiver Operating Characteristic curve (AUC) with 95% bootstrap confidence intervals (CI) was used to compare model performance. AUCs of LRC and LRCB were compared within each cohort to evaluate the contribution of biomarker predictors at different time points.

Model calibration was assessed via the Hosmer-Lemeshow goodness-of-fit test, where a p-value larger than 0.05 would demonstrate a sound calibration (Lemeshow et al., 1982). Model discrimination was tested by measuring the AUC (Hanley et al., 1982). Statistical analyses were performed using R version 3.4.1. Statistical significance was set at 0.05. All results were provided for the test set.

2.3.3 Logistic Regression

Multivariable logistic regression was used to predict the binary survival outcome of mortality status of a patient. The logistic regression model was defined as:

$$\ln\left(\frac{p}{1-p}\right) = \boldsymbol{X\beta}$$

where p is the probability of death, with covariate matrix $\boldsymbol{X}$ and parameter vector $\boldsymbol{\beta}$. Wald's test was used to test the significance of variables in the model.

2.3.4 Elastic Net Regularization

Elastic net regularization was employed to further screen out predictors from the univariate analysis, it was used to determine a smaller subset of predictors that exhibits the strongest effects.

Elastic net regularization is a regularized regression method that linearly combines penalties of lasso and ridge regressions (Friedman et al., 2010). LASSO (Least Absolute Shrinkage and Selection Predictor) is a method to select variables, it has low bias but large variance in accuracy. The limitation of lasso is that for highly correlated data, lasso tends to pick one of them, but cannot do group selection. Given a response vector $y \in \mathbb{R}_{N\times p}$, the lasso regression is defined as following:

$$\hat{\beta}^{lasso} = \underset{\beta \in \mathbb{R}^p}{argmin} \sum_{i=1}^{N} (y_i - \beta^T x_i)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

Here λ ($\lambda \geq 0$) is a tuning parameter, which controls the strength of the penalty term. Ridge regression is a technique to shrink coefficients, but it does not set any coefficients to 0 and does not give easily interpretable model. The ridge regression is defined as following:

$$\hat{\beta}^{ridge} = \underset{\beta \in \mathbb{R}^p}{argmin} \sum_{i=1}^{N} (y_i - \beta^T x_i)^2 + \lambda \sum_{j=1}^{p} {\beta_j}^2$$

The elastic net regularization combines the penalty of ridge and lasso: the ridge penalty shrinks the coefficients of correlated predictors towards each other while the lasso tends to pick one of them and discard the others. The elastic-net penalty mixes these two and overcomes the limitations of them. Elastic net solves the following problem:

$$\min_{\beta_0,\beta} \frac{1}{N} \sum_{i=1}^{N} w_i \, l(y_i, \beta_0 + \beta^T x_i) + \lambda [\frac{(1-\alpha)}{2} \|\beta\|_2^2 + \alpha \|\beta\|_1]$$

Here N is number of observations, w_i is the weight for observation i, $l(y_i, \beta_0 + \beta^T x_i)$ is the negative log-likelihood contribution for observation i, x_i is covariate for observation i, β is parameter. The penalty is defined as $\frac{(1-\alpha)}{2}\|\beta\|_2^2 + \alpha\|\beta\|_1$, it is controlled by elastic-net mixing parameter α, where $0 \leq \alpha \leq 1$. When α=1, the penalty is the lasso penalty; when α=0, the penalty is ridge penalty. $||\beta||_1 = \sum_{j=1}^{p} |\beta_j|$ is the penalty from lasso regression, which selects only one variable from a group of highly correlated variables. $||\beta||_2^2 = \sum_{j=1}^{p} \beta_j^2$ is the penalty from ridge regression, which shrinks coefficients of correlated variables. Parameter α bridges the gap between lasso and ridge. λ is a tuning parameter that controls the overall strength of the penalty, it has value between 0 and positive infinity (Friedman et al., 2010).

For logistic models defined as following:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta^T x_i, \qquad p = \frac{e^{\beta_0 + \beta^T x_i}}{1 + e^{\beta_0 + \beta^T x_i}}$$

The objective function for elastic net regularization with penalized logistic regression uses the negative binomial log-likelihood, and is defined as:

$$\min_{\beta_0, \beta \in \mathbb{R}^{p+1}} -[\frac{1}{N}\sum_{i=1}^{N} y_i\,(\beta_0 + \beta^T x_i) - \log(1 + e^{(\beta_0 + \beta^T x_i)})] + \lambda[\frac{(1-\alpha)}{2}\|\beta\|_2^2 + \alpha\|\beta\|_1]$$

Logistic regression often has problem that when number of predictors is greater than or close to the number of observations, and when this happens, logistic regressions would not perform well. The elastic net penalty would alleviate this problem and select variables.

For this analysis, elastic net regularization was implemented in R package glmnet. The regularization path for elastic net is computed for the elastic net penalty at a grid of values for the regularization parameter λ. The glmnet algorithms use cyclical coordinate descent, which optimizes the objective function over each parameter with others fixed, and cycles repeatedly until

convergence. 10-fold cross-validation was utilized for parameter tuning and variable selection. In this study, elastic net regularization was only used to refine predictor set to within 20 variables. Final set of predictors was selected using backward stepwise selection with logistic regressions.

2.3.5 ROC Curve and AUC

Receiver Operating Characteristics (ROC) curve and Area Under the ROC Curve (AUC) were used to evaluate model performance. They are the most common evaluation metrics for checking classification model's performance.

ROC curves plot sensitivity against 1 minus specificity of a binary classifier across different thresholds. AUC is the area under the ROC curve, it is the estimate of the probability of the classifier to rank a randomly chosen positive event higher than a randomly chosen negative event using normalized unit (Hanley et al., 1982). AUC ranges between 0 and 1, the higher the value, the better measure of classification.

To compare AUCs of two models, 300 bootstrap samples were sampled and AUCs were calculated for each bootstrap sample for two models respectively. 95% bootstrap confidence interval was calculated based on bootstrap AUCs. To formally test the difference between AUCs of two different models, differences between AUCs of two models from the same bootstrap samples were calculated, bootstrap hypothesis test was performed to calculate a p-value for comparison. ROC and AUC were calculated using R package ROCR for logistic regressions.

2.4 Evaluating Trends of Variables

2.4.1 Data

Hour-based cohorts generated in section 2.3.1 were used in this section to assess the predictive power of trends of time-dependent variables. Time-dependent variables were vital signs and biomarkers, which were measured at the baseline of admission (0 hours), and at 6 hours, 24 hours, and 72 hours after admission. The trends of time-dependent variables were calculated as the difference between two measurements at different time points. Trends of time-dependent variables, the latest measurement of time-dependent variables, as well as other time-independent variables, were included in the analysis as predictors. The outcome variable for this section of analysis is 60-day mortality.

2.4.2 Methods for Evaluating Trends

In order to evaluate predictive contribution of time-dependent variables, logistic regressions with both clinical features and biomarkers (LRCB) were constructed at different hour-based cohorts and trends variables were included as predictors of 60-day sepsis mortality.

The model fitting and variable selection were the same in section 2.3.2, elastic net regularization was applied to refine the initial predictor set. The cohorts were randomly split into training (67%) and testing (33%) sets. While the training set was used for model development, the testing set was used for model validation. Multivariable logistic regression was fitted in the training set. The final logistic models were derived after the backward elimination based on AIC. If trends

variables were dropped during the predictor selection process, then it was considered not contributing in mortality prediction.

2.5 Comparing Statistical Models

2.5.1 Data

This section compares logistic regression and Cox regression in predicting sepsis mortality. Only patients from the baseline cohort of section 2.3.1 were used to develop models in this section. Baseline cohort patients' within-60-days-of-admission death time and 60-day mortality were obtained and considered as the outcome variable. Both clinical features and biomarkers from initial predictor set in section 2.3.2 were included in this section of analysis. Time-independent variables and time-dependent variables measured at baseline were also included as predictors.

2.5.2 Methods for Comparing Models

Cox proportional hazards regression is a semiparametric model that is efficient for analyzing survival data, it estimates only hazard ratios between reference and other groups. Cox regression assumes independent observations, censoring independent of time-to-event, and proportional hazard rates, which means that hazard ratio between two groups is constant over time. Logistic regression is a parametric model to analyze the binary dependent variable, it estimates odds ratio of variables. Logistic regression requires independent observations and cannot be used

when outcome is censored. When competitive events are ruled out, follow-up is complete, and all other assumptions are met, the two methods are valid in predicting survival status.

To get more information on time of survival, logistic regressions were developed to predict mortality at different landmark time points. Data was randomly split into training (67%) and testing (33%) sets to develop and validate prediction models. Landmark mortality within 3 days, 7 days, 14 days, 30 days, and 60 days of admission were modeled using logistic regression models. The initial set of predictors was determined using univariate logistic regression. Final models were derived after backward elimination of non-significant predictors.

A Cox proportional hazards model was developed using the same cohort to predict mortality over the same period with the logistic regressions, from baseline of admission to 60-day. The initial set of predictors was determined using univariate Cox regression for the Cox model. Final model was derived after backward elimination of non-significant predictors.

Area under the Receiver Operating Characteristic curve (AUC) with 95% confidence intervals (CI) was used to evaluate model performance at different time points. To formally test the difference between AUCs of two models, 300 bootstrap samples were sampled, logistic AUC and Cox AUC were calculated for each bootstrap sample and the bootstrap hypothesis testing was performed for hypothesis testing. Statistical analyses were performed using R version 3.4.1. Statistical significance was set at 0.05.

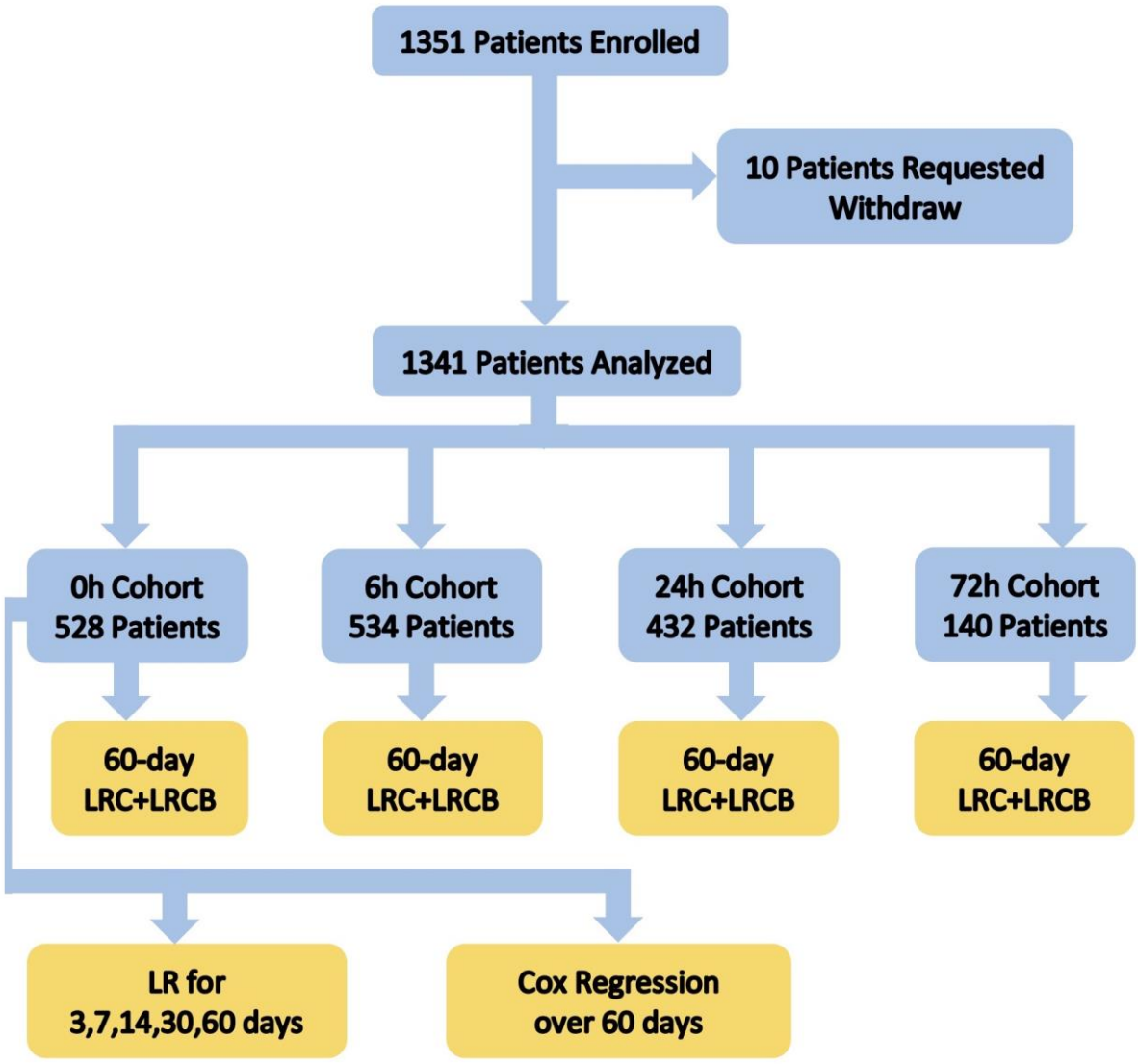


Figure 1 Flow Chart of Analysis

2.5.3 Cox Proportional Hazards Model

The Cox proportional hazards model was used to predict sepsis survival within 60-day period after admission. The model was defined as:

$$h(t|\boldsymbol{X}) = h_o(t)\exp(\boldsymbol{X\beta})$$

where $h_o\,(t)$ is the baseline hazard function with covariate matrix $\boldsymbol{X}$ and parameter vector $\boldsymbol{\beta}$. Wald's test was used to test the significance of variables in the model. Variables selection was based on backward stepwise selection under AIC based pseudolikelihood. Cox regression was developed using R package survival.

2.5.4 Time-dependent AUC

For survival models, the prediction performance is dependent on time of assessment *t* when the outcome is observed over time. The prediction performance measured by ROC is a function of time *t*. To calculate and plot the time-dependent ROC and AUC, timeROC function in the R timeROC package was used. The Inverse Probability of Censoring Weighting (IPCW) was performed to estimate cumulative dynamic time-dependent ROC curve.

Let $D_i(t)$ denotes the time-dependent outcome status for subject i at time *t*. Let *M* be a marker at baseline. For any threshold *c*, the true positive and false positive rates are time-dependent functions defined as:

$$TPR(c,t) = P(M > c|D(t) = 1)$$

$$FPR(c,t) = P(M > c|D(t) = 0)$$

The time-dependent ROC curve *ROC(t)* plots *TPR(c, t)* against *FPR(c, t)* for any threshold *c*. Consider two individuals i and j ($i \neq j$). Time-dependent AUC at time t can be defined as following:

$$AUC(t) = E_{i,j}[I\{\hat{F}_i(t|X_i) > \hat{F}_j\left(t\middle|X_j\right)\}|T_i \leq t, T_j > t]$$

Where T_i and T_j is the event time for i and j, respectively. *F(t)* is the estimated failure probability for individuals. A case is defined as a subject i with $T_i \leq t$. A control is defined as a subject i with $T_i > t$.

Confidence intervals for areas under time-dependent ROC curves (time-dependent AUC) were calculated using confint function in the timeROC package. The method was implemented for inverse probability of censoring weights computed from a Kaplan-Meier estimator. Time-dependent AUC estimators were asymptotically normally distributed. Then, confidence intervals were computed using an estimate of the variance and the quantiles of the standard normal distribution. Pointwise confidence intervals and simultaneous confidence bands were computed from the asymptotic normality of time-dependent AUC estimators. To compute the variance estimates, the function computes the empirical variance estimates of the estimated iid-representations of the time-dependent AUC estimators.

3.0 Results

3.1 Patients

There were 1351 patients recruited by the Protocol-Based Care for Early Septic Shock (ProCESS) trial, 10 of them requested to withdraw, 1341 of them included in this analysis. 528 patients were included in baseline cohort (60-day mortality: 25%, mean age: 60 years, mean baseline lactate: 2.41 mmol/L), 534 patients in 6 hours cohort (24%, 60, 2.35), 432 patients in 24 hours cohort (21%, 60, 2.26), and 140 patients in 72 hours cohort (24%, 61, 2.89).

Important baseline characteristics between patients with different survival statuses were compared in each cohort (Table 1). Variables, including comatose status, cancer and metastatic cancer, immuno-suppressed status, renal disease, cirrhosis, chronic liver disease, mechanical ventilation, non-invasive ventilation, and sofa score, were significantly different between survivors and non-survivors across different cohorts (Table 1).

To demonstrate that the cohorts represented the 1341 patients, baseline characteristics of patients in the cohorts were compared with patients excluded (Table 2). Baseline characteristics were similar for baseline cohort and patients who were excluded from baseline cohort. As for 6 hours cohort and 24 hours cohort, some predictors were significantly different between those included in the cohorts and those not included. Predictors, such as mental status, comatose state, and mechanical ventilation, were significantly different, indicating patients included in 6 hours cohort and 24 hours cohort were less sick than those excluded. The 60-day mortality for those two cohorts was significantly lower than that of the patients not in the cohorts.

Table 1 Baseline Characteristics for Cohorts Based on Survival Status

	Baseline (0 Hours) Cohort		6 Hours Cohort		24 Hours Cohort		72 Hours Cohort		All	
Baseline Variables[1]	Survivor (N=396)	Non-Survivor (N=132)	Survivor (N=407)	Non-Survivor (N=127)	Survivor (N=340)	Non-Survivor (N=92)	Survivor (N=106)	Non-Survivor (N=34)	Survivor (N=977)	Non-Survivor (N=364)
age	58.34 (16.01)*[2]	65.82 (14.0)	58.37 (15.89)*	64.77 (14.19)	58.52 (16.30)*	65.68 (14.31)	60.38 (14.19)	63.79 (16.98)	59.08 (16.08)*	66.58 (14.77)
mental status	285 (72%)*	82 (62.1%)	306 (75.1%)	86 (67.7%)	258 (75.9%)	62 (67.4%)	78 (73.6%)	21 (61.8%)	681 (69.7%)	235 (64.6%)
comatose status	18 (4.5%)*	21 (15.9%)	14 (3.4%)*	19 (15%)	13 (3.8%)*	11 (12%)	6 (5.7%)	4 (11.8%)	55 (5.6%)*	51 (14%)
pleural effusion	28 (7.1%)*	28 (21.2%)	31 (7.6%)*	27 (21.3%)	25 (7.4%)*	24 (26.1%)	4 (3.8%)*	11 (32.4%)	88 (9.0%)*	90 (24.7%)
metastatic cancer	22 (5.6%)*	33 (25%)	21 (5.2%)*	31 (24.4%)	18 (5.3%)*	21 (22.8%)	3 (2.8%)*	9 (26.5%)	53 (5.4%)*	76 (20.9%)
immune-suppressed	55 (13.9%)*	38 (28.8%)	58 (14.3%)*	39 (30.7%)	45 (13.2%)*	27 (29.3%)	8 (7.5%)*	10 (29.4%)	128 (13.1%)*	91 (25%)
renal disease	49 (12.4%)*	30 (22.7%)	53 (13%)*	27 (21.3%)	43 (12.7%)*	19 (20.7%)	14 (13.2%)	8 (23.5%)	133 (13.6%)*	80 (22%)
cirrhosis	17 (4.3%)*	14 (10.6%)	21 (5.2%)*	22 (17.3%)	15 (4.4%)*	17 (18.5%)	7 (6.6%)	5 (14.7%)	43 (4.4%)*	44 (12.1%)
hepatic disease	7 (1.8%)	5 (3.8%)	9 (2.2%)*	9 (7.1%)	7 (2.1%)	6 (6.5%)	2 (1.9%)	1 (2.9%)	16 (1.6%)*	200 (55.0%)
chronic liver disease	24 (6.1%)*	16 (12.1%)	37 (9.1%)*	25 (19.7%)	27 (7.9%)	17 (18.5%)	8 (7.5%)	5 (14.7%)	64 (6.6%)*	47 (12.9%)
mechanical ventilation	44 (11.1%)*	59 (44.6%)	35 (8.6%)*	33 (26.0%)	31 (9.1%)*	23 (25.0%)	15 (14.2%)*	10 (29.4%)	110 (11.3%)*	105 (28.8%)
non-invasive ventilation	14 (3.5%)*	12 (9.9%)	13 (3.2%)*	9 (7.1%)	10 (2.9%)	4 (4.3%)	4 (3.8%)	–	36 (3.7%)*	26 (7.1%)
urine output	81.76 (131.01)	79.91 (149.60)	88.06 (143.39)	80.15 (143.42)	89.16 (141.69)	70.40 (122.43)	83.84 (130.86)	105.44 (135.87)	87.02 (140.56)	68.20 (120.16)
respiration rate	22.23 (6.83)	23.18 (7.13)	22.38 (6.98)	23.36 (7.30)	22.21 (6.33)	23.20 (7.27)	22.42 (7.32)	22.50 (7.96)	22.59 (6.60)*	23.43 (7.20)
temperature	37.28 (1.32)	37.02 (1.69)	37.28 (1.24)*	36.97 (1.56)	37.24 (1.26)*	36.88 (1.61)	36.95 (1.23)	36.62 (2.10)	37.38 (1.30)*	37.01 (1.48)
heart rate	101.42 (19.00)*	105.9 (21.24)	100.65 (18.98)*	105.25 (22.87)	101.48 (18.65)	104.72 (21.98)	104.54 (19.27)	106.44 (20.30)	102.34 (19.58)*	105.56 (22.42)
sbp	99.12 (25.34)	98.14 (21.99)	97.79 (23.64)	98.11 (22.05)	97.97 (23.87)	97.51 (20.66)	98.94 (24.61)	97.32 (17.14)	100.07 (25.03)	99.36 (22.47)
dbp	56.83 (15.47)	56.00 (15.02)	55.77 (14.63)	55.87 (14.94)	55.73 (14.14)	55.83 (15.26)	54.96 (14.71)	56.88 (11.40)	56.97 (15.85)	56.24 (15.45)
O2sat	95.30 (7.24)	93.55 (8.60)	95.67 (6.16)*	94.00 (8.03)	95.81 (5.38)	95.61 (6.99)	95.07 (0.79)	95.06 (6.24)	95.51 (6.13)*	94.37 (7.24)
Sofa	6.44 (3.29)*	9.13 (3.79)	6.32 (3.20)*	9.09 (3.90)	6.24 (3.24)*	9.13 (3.63)	6.75 (3.72)*	8.71 (3.71)	6.54 (3.27)*	8.84 (3.80)
Vasopressor	58 (14.6%)*	37 (28.0%)	57 (14.0%)*	33 (26.0%)	47 (13.8%)*	26 (28.3%)	11 (10.4%)*	13 (38.2%)	173 (17.7%)*	109 (29.9%)
lactate	1 (0, 2)*	2 (1,3)	1 (0,2)	1 (0,2)	1 (0,2)	1 (0,2)	–	–	–	–
IL6	28 (18, 50)	38.22 (25.45)	–	–	–	–	–	–	–	–
IL10	15 (12, 35)	25.35 (21.16)	–	–	–	–	–	–	–	–
WBC	15.51 (9,17)	14.40 (12.10)	15.25 (8.79)	15.14 (13.05)	15.48 (8.63)	14.41 (9.82)	16.23 (8.17)	14.70 (10.92)	15.27 (9.50)*	17.28 (15.39)
platelet	231.51 (132.73)	225.65 (166.96)	227.65 (126.02)	214.89 (164.68)	229.69 (127.71)*	198.91 (149.46)	242.27 (141.46)	209.17 (151.44)	232.82 (128.83)	224.88 (161.66)
14 day mortality	–	89 (67%)	–	78 (61.4%)	–	52 (56.5%)	–	15 (44.12%)	–	239 (65.7%)
60 day mortality	–	132 (100%)	–	127 (100%)	–	92 (100%)	–	34 (100%)	–	364 (100%)

[1] Mean (standard deviation) calculated for normally distributed continuous variables, median (1st quantile, 3rd quantile) calculated for non-normally distributed continuous variables, count(percentage) calculated for categorical variables

[2] * Indicates significant P-value less than 0.05, P-values are calculated from Student's t test and median test for continuous variables and Chi-square test of independence for categorical varia

Table 2 Baseline Characteristics for Cohorts Based on Analysis Inclusion

	T0	T0-	p-value[2]	T6	T6-	p-value	T24	T24-	p-value	T72	T72-	p-value	ALL
Baseline Variable[1]	N=528 (39.4%)	N=813 (60.6%)	—	N=534 (39.8%)	N=807 (60.2%)	—	N=432 (32.2%)	N=909 (67.8%)	—	N=140 (10.4%)	N=1201 (89.6%)	—	N=1341
age	60.21 (15.86)	61.72 (16.20)	0.093	59.89 (15.73)	61.93 (16.26)	0.0229*	60.05 (16.15)	61.63 (16.03)	0.0927	61.21 (14.93)	61.11 (16.21)	0.9445	61.12 (16.08)
ethnicity	70 (13.3%)	73 (9.0%)	0.013*	70 (13.1%)	73 (9.1%)	0.018*	52 (12.0%)	91 (10.0%)	0.261	22 (15.7%)	121 (10.1%)	0.041*	143 (10.7%)
Race													
White	374 (70.8%)	544 (66.9%)	0.131	375 (70.2%)	543 (67.3%)	0.257	308 (71.3%)	610 (67.1%)	0.123	102 (72.8%)	815 (67.9%)	0.229	918 (68.5%)
Black	117 (22.2%)	217 (26.7%)	0.061	120 (22.5%)	215 (26.6%)	0.084	94 (21.8%)	240 (26.4%)	0.066	26 (18.6%)	307 (25.6%)	0.07	334 (24.9%)
Other	37 (7.0%)	52 (6.4%)	0.660	39 (7.3%)	49 (6.1%)	0.406	30 (6.9%)	59 (6.5%)	0.755	12 (8.6%)	79 (6.6%)	0.375	89 (6.6%)
mental status	367 (69.5%)	549 (67.5%)	0.446	392 (73.4%)	524 (64.9%)	0.001*	320 (74.1%)	596 (65.6%)	0.002*	99 (70.7%)	817 (68.0%)	0.518	916 (68.3%)
comatose state	39 (7.4%)	67 (8.2%)	0.571	33 (6.2%)	73 (9.0%)	0.057	24 (5.6%)	82 (9.0%)	0.028*	10 (7.1%)	62 (5.2%)	0.325	106 (7.9%)
cancer	93 (17.6%)	141 (17.4%)	0.438	92 (17.2%)	142 (17.6%)	0.862	70 (16.2%)	164 (18.1%)	0.451	18 (12.9%)	216 (18.0%)	0.130	234(17.5%)
metastatic cancer	55 (10.4%)	74 (9.1%)	0.425	52 (9.7%)	77 (9.6%)	0.905	39 (9.0%)	90 (9.9%)	0.612	12 (8.6%)	117 (9.8%)	0.657	129 (9.7%)
immno-suppresed	93 (17.6%)	126 (15.5%)	0.306	97 (18.2%)	122 (15.2%)	0.139	72 (16.7%)	147 (16.2%)	0.819	18 (12.9%)	201 (16.8%)	0.240	219 (16.3%)
renal disease	79 (15.0%)	134 (16.5%)	0.457	80 (14.9%)	133 (16.5%)	0.462	62 (14.4%)	151 (16.6%)	0.29	22 (15.7%)	191 (15.9%)	0.954	213 (16.3%)
cirrhosis	31 (5.9%)	56 (6.9%)	0.460	43 (8.1%)	44 (5.5%)	0.057	32 (7.4%)	55 (6.0%)	0.346	12 (8.6%)	75 (6.3%)	0.290	87 (6.5%)
hepatic failure	12 (2.3%)	24 (2.9%)	0.452	18 (3.4%)	18 (2.2%)	0.206	13 (3.0%)	23 (2.5%)	0.612	3 (2.1%)	33 (2.7%)	0.675	36 (2.7%)
chronic liver disease	40 (7.6%)	71 (8.7%)	0.095	62 (11.6%)	49 (6.1%)	0*	44 (10.2%)	67 (7.4%)	0.080	13 (9.3%)	98 (8.2%)	0.647	111 (8.3%)
connective tissue disease	22 (4.2%)	37 (4.5%)	0.737	28 (5.2%)	31 (3.8%)	0.220	22 (5.1%)	37 (4.1%)	0.963	9 (6.4%)	50 (4.2%)	0.216	59 (4.4%)
pulse oximetry (%)	94.86 (7.64)	95.42 (5.58)	0.1217	95.27 (6.69)	95.16 (6.33)	0.7608	95.55 (5.77)	95.03 (6.78)	0.1694	95.24 (5.59)	95.18 (6.57)	0.9174	95.20 (6.47)
non-invasive ventilation	26 (4.9%)	36 (4.4%)	0.672	22 (4.1%)	40 (4.9%)	0.475	13 (3.2%)	48 (5.3%)	0.062	4 (2.9%)	58 (4.8%)	0.293	62 (4.6%)
mechanical ventilation	83 (15.7%)	132 (16.3%)	0.801	68 (12.7%)	147 (18.3%)	0.007*	54 (12.5%)	161 (17.8%)	0.015*	25 (17.8%)	190 (15.8%)	0.534	215 (16.0%)
Urine output	81.30 (135.75)	82.30 (135.46)	0.8950	86.18 (132.31)	79.93 (139.06)	0.4116	85.16 (137.89)	80.33 (134.42)	0.5421	89.08 (131.92)	81.06 (135.97)	0.5078	81.90 (135.53)
temperature	37.22 (1.43)	37.32 (1.32)	0.1900	37.21 (1.32)	37.32 (1.39)	0.1481	37.16 (1.35)	37.33 (1.37)	0.0331*	36.88 (1.47)	37.33 (1.34)	0.0002*	37.28 (1.36)
heart rate	102.54 (19.66)	103.65 (20.92)	0.2366	101.72 (20.03)	104.19 (20.64)	0.0301*	102.15 (19.41)	103.71 (20.88)	0.1913	105.00 (19.47)	102.99 (20.53)	0.2706	103.21 (20.42)
respiratory rate	22.46 (6.36)	23.05 (8.68)	0.1789	22.61 (7.06)	22.95 (6.58)	0.3685	22.41 (6.54)	23.00 (6.88)	0.1363	22.43 (7.45)	22.86 (6.69)	0.4773	22.81 (6.77)
systolic blood pressure	98.88(24.23)	100.54 (24.24)	0.2527	97.86 (23.26)	101.21 (24.98)	0.0136*	97.87 (23.22)	100.83 (24.83)	0.0375*	98.55 (22.97)	100.04 (25.52)	0.5092	99.88 (24.36)
diastolic blood pressure	56.63 (15.35)	56.87 (15.99)	0.7851	55.79 (14.78)	57.42 (16.37)	0.0639	55.75 (14.36)	57.25 (16.33)	0.1028	55.43 (13.96)	56.93 (15.93)	0.2860	56.77 (15.74)
Lactate	2 (1, 3)	—	—	1 (0, 2)	—	—	1 (0, 2)	—	—	—	—	—	—
IL10	15 (12, 35)	—	—	—	—	—	—	—	—	—	—	—	—
IL6	28 (18, 50)	—	—	—	—	—	—	—	—	—	—	—	—
white blood cell	15.23 (9.98)	16.19 (12.27)	0.1329	15.23 (9.94)	16.18 (12.27)	0.0228*	15.25 (8.89)	16.07 (12.43)	0.2190	15.86 (8.89)	15.79 (11.67)	0.9453	15.80 (11.41)
Platelet	230.06 (141.84)	231.14 (135.96)	0.8889	224.65 (136.03)	234.75 (139.77)	0.1907	223.13 (133.06)	234.36 (140.72)	0.1649	234.29 (144.06)	230.28 (137.68)	0.7456	230.70 (138.32)
60-day mortality	132 (25.0%)	232 (28.5%)	0.155	127 (23.8%)	237 (29.4%)	0.024*	92 (21.3%)	272 (29.9%)	0.001*	34(24.28%)	330 (27.48%)	0.422	364 (27.1%)

[1] Mean (standard deviation) calculated for normally distributed continuous variables, median (1st quantile, 3rd quantile) calculated for non-normal distributed continuous variables, count (percentage) calculated for categorical variables

[2] P-values are calculated from Student's t test or median test for continuous variables and Chi-square test of independence for categorical variables

* Indicates significant P-value less than 0.05

The survival time for patients ranges from within 6 hours after admission to over 60 days after admission. The overall 60-day mortality for all 1341 patients recruited by the trial is 27.14%. 15% of the death occurs within 1 day of admission, 30% of the death occurs within 3 days of admission, 56% of the death occurs within 14 days of admission.

3.2 Evaluation of Biomarkers

3.2.1 Predictive Models

LRC and LRCB were fitted for each cohort (Table 3). Metastatic cancer and age were significant predictors in all models across different times. Some other chronic conditions, such as immuno-compromised status, cirrhosis, and chronic liver diseases, also increased the risk for 60-day sepsis mortality. On the other hand, among time-dependent vital sign predictors, higher diastolic blood pressure and higher Glasgow coma scale score (GCS) lowered the risk of death. As for biomarker predictors, lactate was significant in all LRCBs within 24 hours, TMB was significant in the baseline model, and total bilirubin level was significant in the 72 hours LRCB model. Lactate is significant all the time within 24 hours of admission, high lacate level is associated with high risk of death.

Table 3 Logistic Models for Mortality Prediction at Different Times

Predictors	OR (95% CI)	
Baseline (0h)	**LRC[1]**	**LRCB[2]**
Metastatic Cancer	3.72 (1.60, 8.88)	3.53 (1.40, 8.99)
Immuno-Compromised	2.29 (1.10, 4.70)	2.89 (1.29, 6.34)
Cirrhosis	4.95 (1.61, 15.20)	4.39 (1.27, 14.92)
Age	1.04 (1.02, 1.06)	1.04 (1.02, 1.06)
Mechanical Ventilation 0h	3.97 (2.05, 7.74)	3.19 (1.52, 6.22)
Lactate 0h	—	1.30 (1.16, 1.45)
TMB 0h	—	1.09 (1.02, 1.17)
6h		
Metastatic Cancer	6.03 (2.57, 14.58)	4.87 (2.02, 11.99)
Immuno-Compromised	3.75 (1.81, 7.73)	3.96 (1.92, 8.19)
Cirrhosis	8.67 (3.45, 22.64)	7.11 (2.75, 18.90)
Age	1.03 (1.01, 1.06)	1.03 (1.01, 1.05)
GCS 6h	0.82 (0.77, 0.89)	0.86 (0.79, 0.93)
Heart Rate 6h	1.01(1.00, 1.03)	—
Lactate 6h	—	1.18 (1.07, 1.33)
24h		
Metastatic Cancer	10.08 (3.96, 26.67)	4.27 (1.50, 12.00)
Chronic hepatitis	5.46 (2.18, 13.77)	6.44 (2.47, 17.02)
Age	1.04 (1.01, 1.06)	1.04 (1.02, 1.07)
Diastolic blood pressure 24h	0.97 (0.95, 0.99)	0.97 (0.94, 0.99)
GCS 24h		—
score 3-8	4.48 (1.81, 11.09)	
score 9-12	3.83 (1.01, 13.73)	
Lactate 24h	—	1.42 (1.23, 1.69)
72h		
Pleural Effusion	19.70 5.47, 87.98)	15.47 (4.02, 70.33)
Metastatic Cancer	10.75 (2.40, 60.62)	11.33 (2.44, 66.72)
Glasgow Coma Score	0.82 (0.71, 0.95)	0.81 (0.69, 0.93)
Systolic Blood Pressure 24h	0.96 (0.93, 0.98)	0.96 (0.92, 0.99)
Bilirubin 72h	—	2.41 (1.33, 4.94)

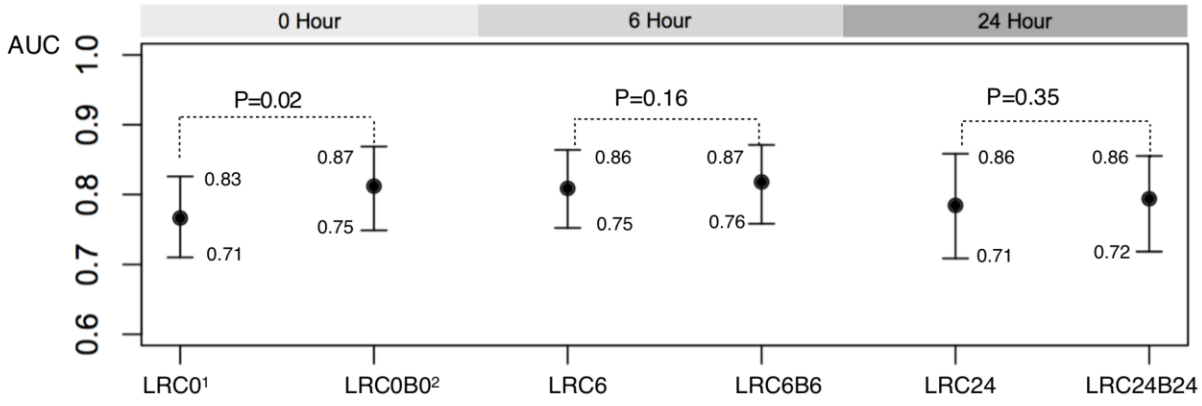
[1] LRC denotes logistic regression models of clinical variables only

[2] LRCB denotes logistic regression of clinical variables and biomarker variables

3.2.2 Predictive Power of Biomarkers

Area under the Receiver Operating Characteristic curve (AUC) with 95% bootstrap confidence intervals (CI) was used to compare model performance. Point estimates of AUCs and their 95% confidence intervals were calculated and plotted (Figure 2). At baseline time, the AUC increased from 0.766 (95% CI: 0.710, 0.826) of the LRC to 0.812 (95% CI: 0.749,0.869) of the LRCB, with a p-value of 0.023. For 6 hours cohort, 24 hours cohort, and 72 hours cohort, there was no significant difference between the AUCs of LRC and LRCB. Therefore, only at baseline, adding biomarkers to clinical features increased predictive power. As for comparisons between the cohorts, there was no significant difference between the AUCs of cohorts at different times.

Figure 2 Confidence Intervals for AUCs of Logistic Regressions



[1] LRC0, LRC6, LRC24 are logistic regression models with clinical variables only at 0h, 6h, 24h respectively
[2] LRC0B0, LRC6B6, LRC24B24 are logistic regression models with clinical variables and biomarker variables at 0h, 6h, 24h respectively

3.3 Evaluation of Time-dependent Variables

3.3.1 Predictive Power of Time-dependent Variables

To evaluate trends of time-dependent variables, for all time-dependent variables, changes over time in measurements were included in predictor selection process. The changes were calculated as latest measurement minus the earlier measurement. In the final models, none of those changes of time-dependent variables were significant, so they were all dropped. Latest measurements of time-dependent predictors were significant in the models. For example, for 6 hours cohort, lactate level measured at 6 hours were significant, but its change over time from baseline time to 6 hours was not. When trends of time-dependent variables were added to the models, none of them were significant, after dropping non-significant variables, the final models were exactly the same with Table 1. Therefore, based on the final models, only the latest measurements of time-dependent variables matter in the prediction.

3.3.2 Time-independent Variables in the Prediction

According to the final logistic regressions (Table 3), metastatic cancer is highly significant in all models, indicates that having metastatic cancer would higher the risk of sepsis mortality. Other chronic condition predictors, such as immune-compromised status, cirrhosis, and old age, would also increase the probability of death.

3.4 Comparison of Statistical Models

3.4.1 Predictive Models

To make a comparison between logistic regression and Cox proportional hazards regression in sepsis mortality prediction, only 528 patients from the baseline cohort was included in analysis (60-day mortality: 25%, mean age: 60 years, mean baseline lactate: 2.41 mmol/L).

For logistic regressions, five unique logistic regressions were developed for predicting 3-day mortality, 7-day mortality, 14-day mortality, 30-day mortality, 60-day mortality. All these logistic regressions including both clinical features and biomarker. 3-day mortality model has AUC 0.888 with 95% CI (0.836, 0.939); 7-day mortality model has AUC 0.827 with 95% CI (0.776, 0.879); 14-day mortality model has AUC 0.858 with 95% CI (0.820, 0.895); and 60-day mortality model has AUC 0.795 with 95% CI (0.716, 0.835).

For the Cox proportional hazards regression, a single model for 60-day survival was developed. None of the final set of predictors were found to violate the proportional hazards assumption. The time-dependent AUC ranges between (0.859, 0.884). The 95% confidence intervals of Cox regression AUC at different time points is shown in table 6.

Model predictors and coefficients are shown in table 4 and table 5. For 60-day mortality logistic regression and 60-day mortality Cox regression, the variables were almost the same: metastatic cancer, immunocompromised status, cirrhosis, age, lactate, and TMB were significant in both models. The coefficients of predictors were different in those two models.

Table 4 Logistic Models for Landmark Mortality Prediction

Predictors	OR (95% CI)
3-day mortality	
Mechanical Ventilation 0h	5.89 (2.06, 16.80)
Metastatic Cancer	4.16 (1.65, 10.47)
Lactate	1.29 (1.15, 1.46)
TNF	1.02 (1.01, 1.05)
log(TMB)	1.99 (1.04, 3.81)
7-day mortality	
O2 Saturation	0.94 (0.90, 0.98)
Age	1.03 (1.01, 1.06)
Heart Rate	1.02 (1.01, 1.03)
Glasgow Coma Score	0.88 (0.81, 0.95)
Lactate	1.33 (1.19, 1.50)
VEGF	1.02 (1.01, 1.04)
TMB	1.08 (1.02, 1.16)
14-day mortality	
Pleural Effusion	2.47 (1.01, 5.99)
Metastatic Cancer	3.33 (1.39, 7.95)
Glasgow Coma Score	0.90 (0.84, 0.97)
Noninvasive Ventilation	4.62 (1.44, 14.77)
Lactate	1.36 (1.21, 1.54)
log(TMB)	2.63 (1.57, 4.41)
30-day mortality	
Pleural Effusion	5.51 (2.45, 12.42)
Metastatic Cancer	3.82 (1.71, 8.55)
Age	1.03 (1.01, 1.05)
Glasgow Coma Score	0.88 (0.81, 0.95)
Lactate	1.38 (1.23, 1.56)
60-day mortality	
Metastatic Cancer	5.35 (1.40, 8.99)
Immuno-Compromised	2.89 (1.29, 6.34)
Cirrhosis	4.39 (1.27, 14.92)
Age	1.04(1.02, 1.06)
Lactate	1.30 (1.16, 1.45)
TMB	1.09 (1.02, 1.17)
Mechanical Ventilation	3.19 (1.52, 6.22)

Table 5 Cox Model for 60-day Mortality Prediction

Predictors	Hazard Ratio (95% CI)
Metastatic Cancer	2.52 (1.53, 4.17)
Immuno-Compromised	2.10 (1.20, 3.67)
Cirrhosis	3.08 (1.61, 5.89)
Age	1.03 (1.01, 1.05)
Lactate	1.15 (1.09, 1.22)
TMB	1.06 (1.02, 1.10)
Glasgow Coma Score	0.90 (0.85, 0.95)
Noninvasive Ventilation	2.76 (1.32, 5.74)

Table 6 Cox Model Time-dependent AUC

Day	COX AUC% (95% CI)
3	87.97 (79.81, 96.13)
18	87.12 (81.37, 92.87)
30	86.70 (81.15, 92.26)
45	88.35 (83.24, 93.47)
60	85.95 (80.01, 91.89)

3.4.2 Model Evaluation

Area under the Receiver Operating Characteristic curve (AUC) with 95% bootstrap confidence intervals (CI) was used to evaluate the two models. According to confidence intervals of AUC (Figure 3), logistic model and Cox model have similar predictive power in predicting short-term sepsis mortality (mortality within 30 days of admission), where the majority of events occur. There is no significant difference between AUCs of two models before 30-days. At 60-day, the AUC of logistic regression is 0.795 with 95% CI (0.716, 0.835), the AUC of Cox regression is 0.859 with 95% CI (0.800, 0.919). 300 bootstrap samples were taken, the difference between AUCs is calculated as Cox regression AUC minus logistic regression AUC for each bootstrap sample. The 95% confidence intervals for the 60-day AUC difference is (0.039, 0.048). Only at 60-day, the bootstrap hypothesis test has significant p-value. Therefore, at 60-day, AUCs of logistic regression are significantly lower than AUCs of Cox regression, and a Cox regression is preferable in predicting long-term sepsis mortality.

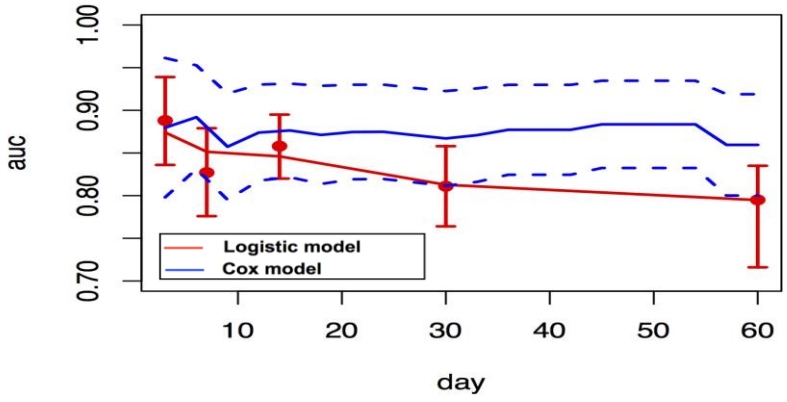


Figure 3 Plot of Confidence Intervals of AUC

4.0 Discussion

4.1 Risk Factors of Sepsis Mortality

This study suggests that biomarkers provide incremental discrimination ability to clinical features for predicting 60-day mortality at the time of baseline. However, at later time points, biomarkers did not have significant contribution to mortality prediction. On the contrary to the original hypotheses that time-trends of measurements contribute to prediction, trends of change in time-dependent variables did not increase predictive power, while the latest measurements did matter in the prediction. Chronic morbid conditions would increase the risk of death, metastatic cancer and organ failures would especially do so. Age was also an important risk factor. As for biomarkers, lactate, TMB, and total bilirubin level might help identify high risk of death.

Compared with other published literature, this study is unique in including biomarkers in sepsis mortality prediction. Biomarkers are commonly used to diagnose sepsis, but rarely used in mortality prediction. Including biomarkers as sepsis mortality predictors may help identify patients who are developing severe sepsis before organ dysfunction has advanced to far (Faix, 2013). Based on the study results, biomarkers, especially lactate and TMB, are supportive to predicting sepsis mortality at the time of admission when sepsis is suspected by the treating physician.

4.2 Model Selection for Sepsis Mortality Prediction

Logistic regression and Cox proportional hazards regression have been previously compared in cohort studies (Staley et al., 2017). It was reported that: (i) the Cox model yields more precise estimates of association; (ii) odds ratios and hazard ratios diverge as follow-up time, cumulative disease incidence and the strength of the association increases; (iii) in certain situations, the Cox regression has greater statistical power (Green et al., 1983).

This study investigated the differences between logistic regression and Cox regression when analyzing sepsis survival data. Compared with Cox regression, which analyzes information on the entire time period and provides time for the event to happen, logistic regression only explains survival below or above a fixed time point. To make these two methods comparable, the 60-day mortality period was divided into 3-day, 7-day, 14-day, 30-day, and 60-day intervals, different logistic regressions were developed for different intervals. According to the AUC results (Figure 3), logistic regressions have similar predictive power with Cox regression for short term mortality within 30 days. When time is considered, Cox regression has higher predictive power for long term mortality.

Based on the study results in section 3.4.1, the model coefficients differ, because hazard ratios and odds ratios are different measures of association and have different interpretations. Cox regressions incorporate the length of time the patients survived and measure whether the risk factors affect the time at which the disease event occurs. Logistic regression assesses whether the risk factor affects the odds of disease, and hence does not take into account the time of death. In logistic regressions, early and late death are given the same weight in the analysis. Therefore, they have lower power than the Cox regression in making long-term predictions, they are better used for short-term variability of events (Staley et al., 2017).

4.3 Comparison with Existing Models

The predictive models in this study are not sepsis specific, they can also be generalized to predict mortality in all ICU patients. The predictors are not sepsis specific, except for biomarker lactate. Those predictors are easily obtained, routinely recorded, and less likely to be missing. An existing model, the Mortality Prediction Model (MPM) was developed in 1985 for mortality prediction in ICU patients, it uses patients' clinical variables at admission, and 24 and 48 hours after the admission to predict the probability of in-hospital mortality with logistic regressions (Lemeshow et al., 1985). Compare LRC0B0 with MPM at admission, and LRC24B24 with MPM at 24 hours, most of the predictors are the same: cancer, organ failures, and age. In addition, the coefficients for age and cancer are very close between our models and MPM models. MPM models also include type of hospital admission and infection as predictors (Lemeshow et al., 1985). For our study, all patients had infection and all the admissions were emergency. AUCs were calculated to compare our models and MPM models for baseline cohort and 24 hours cohort. At baseline time, LRC0B0 has AUC of 0.78 with 95% CI (0.71, 0.85), while MPM has an AUC of 0.71, which is lower. At 24 hours, AUC of our model is 0.74 with 95% CI (0.66, 0.83), and the AUC of MPM is also 0.74.

Since the outcome is 60-day mortality, which is long-term, the models in this study may be more accurate for predicting long-term mortality. Most significant risk factors in our study are chronic conditions, while vital signs seem less significant. Chronic conditions, such as comorbidities and baseline health, may greatly influence outcome of a patient and are determinant in predicting late mortality (Kennelly et al., 2016). In comparison, vital signs are more likely to predict death in a short period of time rather than in distant future. By applying the baseline models (LRC0 and LRC0B0) to predict mortality for patients who had survived over 24 hours and 72

hours, the resulted AUCs were higher than the AUC for all patients, this result indicates that the models have better performance for predicting late mortality. Thus, predictive models in this study may provide help in predicting long-term mortality in ICU patients irrespective of specific illness.

4.4 Limitations

One major limitation of this study is missing data. For some variables, the missing value is more than 50% or even 80%. As a result, some variables were eliminated from the initial set of predictor variables and not included in the statistical analysis process. In addition, patients with missing values for selected predictors were excluded from the cohorts, therefore the sample size of each cohorts was greatly reduced. Another potential problem is that patients in different cohorts are not comparable. Patients in 6 hours cohort and 24 hours cohort are patients who had survived 6 hours and 72 hours, they were significantly healthier than patients excluded from those two cohorts (Table 2). This might due to patients who were sicker tend to have less measurements and more missing data for the selected predictors and therefore were not be able to be included in the cohorts.

The definitions and clinical criteria of sepsis and septic shock were revised in 2016 (Singer et al., 2016). The new definition defines sepsis as life-threatening organ dysfunction caused by dysregulated host response to infection and eliminates the systemic inflammatory response syndrome (SIRS) criteria. Patients with suspected infection can be identified with beside qSOFA score, alteration in mental status, low systolic blood pressure, and high respiratory rate(Singer et al., 2016). Although the data of this study were collected before the definition change and the study

enrollment eligibility were based on SIRS, the major clinical criteria is not changed, the results of this study should not be affected and thus remain the same.

5.0 Conclusion

Predictive models for death within 60 days of the hospital admission of sepsis patients were developed. The models were based on clinical features and biomarker variables obtained during hospital admission. Biomarkers provided incremental discrimination ability over clinical features alone to predict 60-day mortality at the baseline time of hospital admission. Latest measurements of time-dependent variables matter in the prediction, while trends of time-dependent variables do not contribute to mortality prediction. This study also found that Cox regression and logistic regression have similar predictive power for short-term mortality within 30 days, while Cox regression has better predictive power for long-term mortality. These predictive models are not sepsis specific and may be generalized to help identify ICU patients with high risk of death, while significant work remains to confirm the contribution of biomarkers and time trends of time-dependent predictors in sepsis prognosis predicting, other statistical models could be explored to analyze sepsis survival.

Bibliography

Angus, D. C., Wax, R. S. (2001). Epidemiology of sepsis: an update. *Crit Care Med*, 29,109-116.

Dellinger, R. P., Carlet, J. M., Masur, H., et al. (2004). Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. *Crit Care Med*, 32, 858–73.

Faix, J. D. (2013). Biomarkers of sepsis. *Crit Rev Clin Lab Sci*, 50,23–36.

Friedman, J., Hastie, T., Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*, 33, 1–22.

Ford, D. W., Goodwin, A. J., Simpson, A. N., et al. (2016). A Severe Sepsis Mortality Prediction Model and Score for Use With Administrative Data. *Crit Care Med*, 44,319–327.

Green, M. S, Symons, M. J. (1983). A comparison of logistic risk function and the proportional hazards model in prospective epudemiologic studies. *J Chronic Dis*, 36, 715-723.

Hanley, J. A., McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29–36.

Kennelly, P. J., Martin-Loeches, I. (2016). Long term mortality following sepsis. *Ann Transl Med*, 4, 387.

Kasal, J., Jovanovic, Z., Clermont, G., et al.(2004). Comparison of Cox and Gray's survival models in severe sepsis. *Crit Care Med*, 32,700–707.

Lemeshow, S., Hosmer, D. W. (1982). A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol*, 115:92–106.

Lemeshow, S., Teres, D., Pastides, H., et al. (1985). A method for predicting survival and mortality of ICU patients using objectively derived weights. *Crit Care Med* , 13, 519–25.

Moreno, R. P., Metnitz, B., Adler, L., et al. (2008). Sepsis mortality prediction based on predisposition, infection and response. *Intensive Care Med*, 34, 496–504.

Process Investigators (2014). A Randomized Trial of Protocol-Based Care for Early Septic Shock *N Engl J Med*, 370,1683–1693.

Reinhart, K., Meisner, M., Brunkhorst, F. M. (2006). Markers for Sepsis Diagnosis: What is Useful? *Crit Care Clin*, 22, 503–519.

Stanley, J. R., Edmund, J., Kaptoge, S., Butterworth, A. S., Sweeting, M. j., Wood, A. M., Howson, J. M. M. (2017). A comparsion of Cox and Logistic Regression for use in genome-wide association studies of cohort and case-cohort design. *European Journal of Human Genetics,* 25, 854-862.

Singer, M., Deutschman, C. S., Seymour, C. W., et al. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315:801

Teres, D., Lemeshow, S., Avrunin, J. S., et al.(1987). Validation of the mortality prediction model for ICU patients. *Crit Care Med* , 15, 208–13.