Check point: Mortality assessment in ICU with multivariate physiological time-series

4/3/12

Danning He, DHSI, JHSOM
Mentor: Jim Fackler, Harold Lehmann

Overview

- **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)*

- Measurements may be recorded at regular intervals ranging from hourly to daily, or at irregular intervals. Not all time series are available in all cases.
- In a few cases, such as blood pressure, different measurements made using two or more methods or sensors may be recorded with the same or only slightly different time-stamps.

Updated Deliverables

- ***Minimum (Done)***

-- Logistic regression with log odds ratios as risk features

-- Performance evaluation: ROC, AUC

- ***Expected***

-- Model analysis (Done)

-- Incorporating waiting time until the critical events (hold on)

-- Try features constructed from standard HMM, Kalman Filter (code is ready, but how to apply to this study?)

-- Incorporate dependencies between observations (e.g. autoregressive process)

- ***Maximum***

-- Optimize features to achieve better classification performance

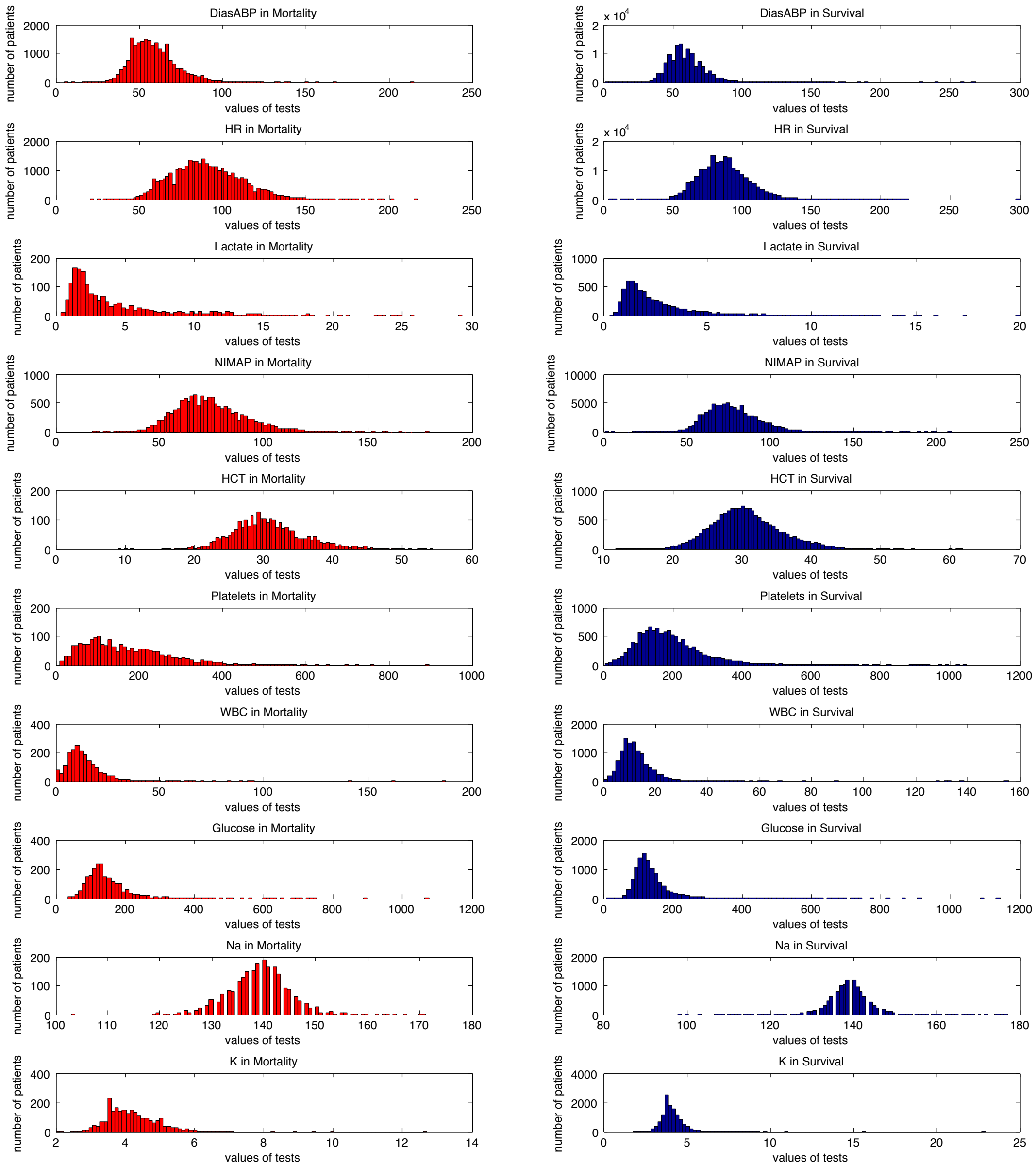
-- Documentation

-- partial AUC

Updated Timeline

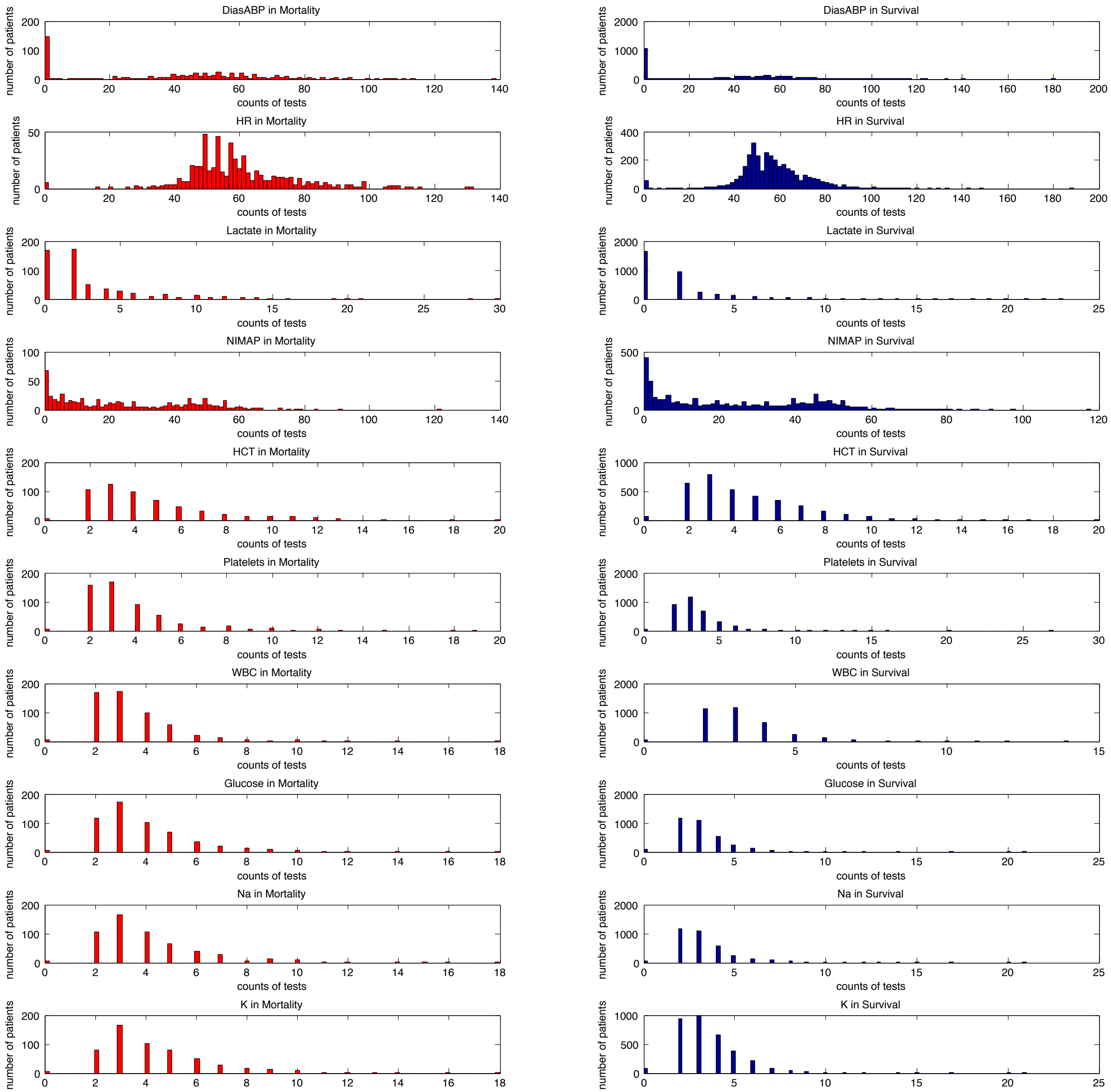
Timeline		Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13
Milestone 1 (By 2/23)	reading list													
	project plan													
	preprocessing data													
Milestone 2 (By 3/26)	Features as Log odds ratio				Milestone is one week's delay, because of new ideas to construct risk features									
	Logistic regression													
	AUC and ROC													
Milestone 3 (By 5/1)	Model Analysis													
	Time dependencies between observations									Autoregressive cannot be applied directly				
	Features from hidden states									Implementation is fast once I know how				
Milestone 4 (By 5/8)	Model optimization													
	Documentation													

Distributions of values in two classes



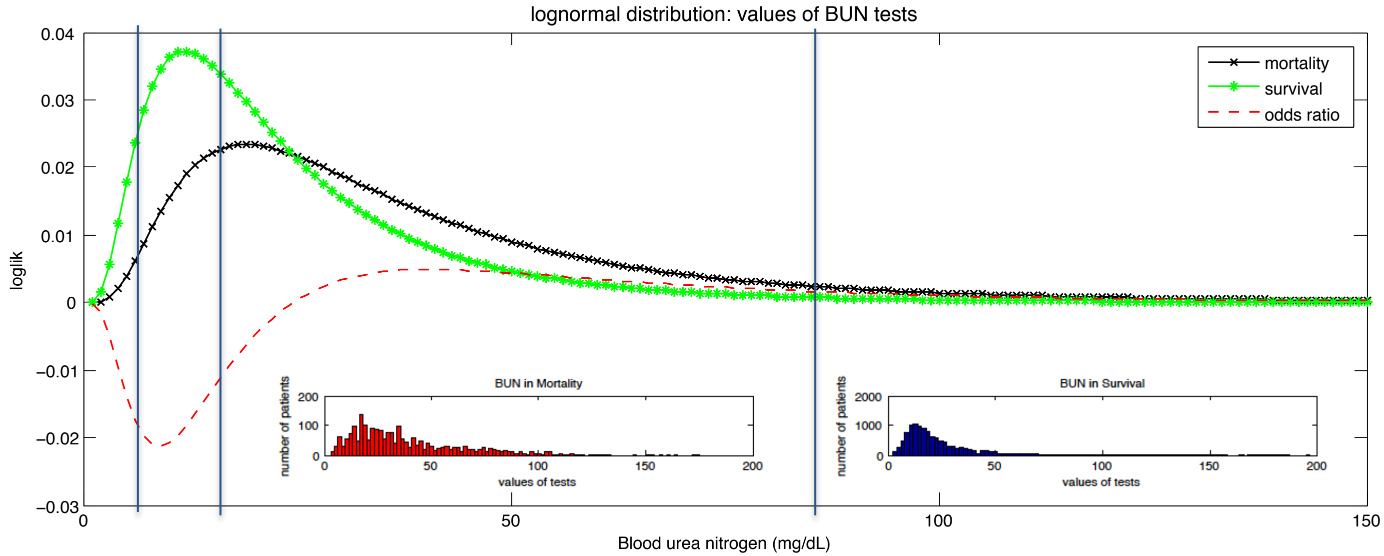
Fit with five candidate long-tailed distributions: Exponential, lognormal, gamma, normal or weibull

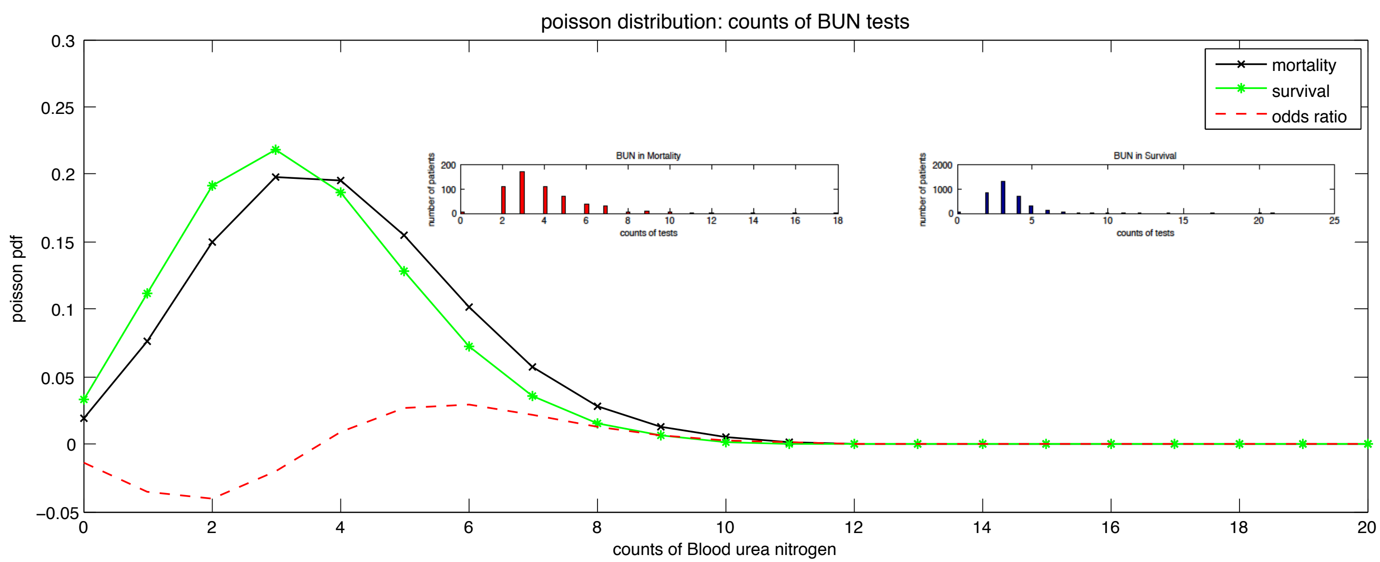
Distributions of counts of tests in two classes



Fit with Poisson distribution

Built up intuition





Nonlinear models of risk factors

- Mainly based on Suchi's paper.

$$f(\mathbf{v}_i) = \begin{cases} \boxed{\log \dfrac{P(\mathbf{v}_i \mid HM)}{P(\mathbf{v}_i \mid LM)}} + \boxed{\log \dfrac{P(T \mid HM)}{P(T \mid LM)}} \\ \log \dfrac{P(T = 0 \mid HM)}{P(T = 0 \mid LM)} \end{cases}$$

where $$\log \frac{P(\mathbf{v}_i \mid HM)}{P(\mathbf{v}_i \mid LM)} = \sum_{t=1}^{T} \log \frac{P(v_{it} \mid HM)}{P(v_{it} \mid LM)}$$

- Method in original paper:

$$f(v_i) = \begin{cases} \log \dfrac{P(v_i \mid HM, m_i = 0) \cdot P(m_i = 0 \mid HM)}{P(v_i \mid LM, m_i = 0) \cdot P(m_i = 0 \mid LM)} \\ \log \dfrac{P(m_i = 1 \mid HM)}{P(m_i = 1 \mid LM)} \end{cases}$$

Nonlinear models of risk factors

- Mainly based on Suchi's paper.

$$f(\mathbf{v}_i)=\begin{cases} \boxed{\log\dfrac{P(\mathbf{v}_i \mid HM)}{P(\mathbf{v}_i \mid LM)}}+\boxed{\log\dfrac{P(T \mid HM)}{P(T \mid LM)}} \\ \log\dfrac{P(T=0 \mid HM)}{P(T=0 \mid LM)} \end{cases}$$

Poisson distribution

where $\log\dfrac{P(\mathbf{v}_i \mid HM)}{P(\mathbf{v}_i \mid LM)}=\sum_{t=1}^{T}\log\dfrac{P(v_{it} \mid HM)}{P(v_{it} \mid LM)}$

Five candidate long-tailed distributions:
Exponential, lognormal, gamma, normal or weibull

- Method in original paper:

$$f(v_i)=\begin{cases} \log\dfrac{P(v_i \mid HM, m_i=0)\cdot P(m_i=0 \mid HM)}{P(v_i \mid LM, m_i=0)\cdot P(m_i=0 \mid LM)} \\ \log\dfrac{P(m_i=1 \mid HM)}{P(m_i=1 \mid LM)} \end{cases}$$

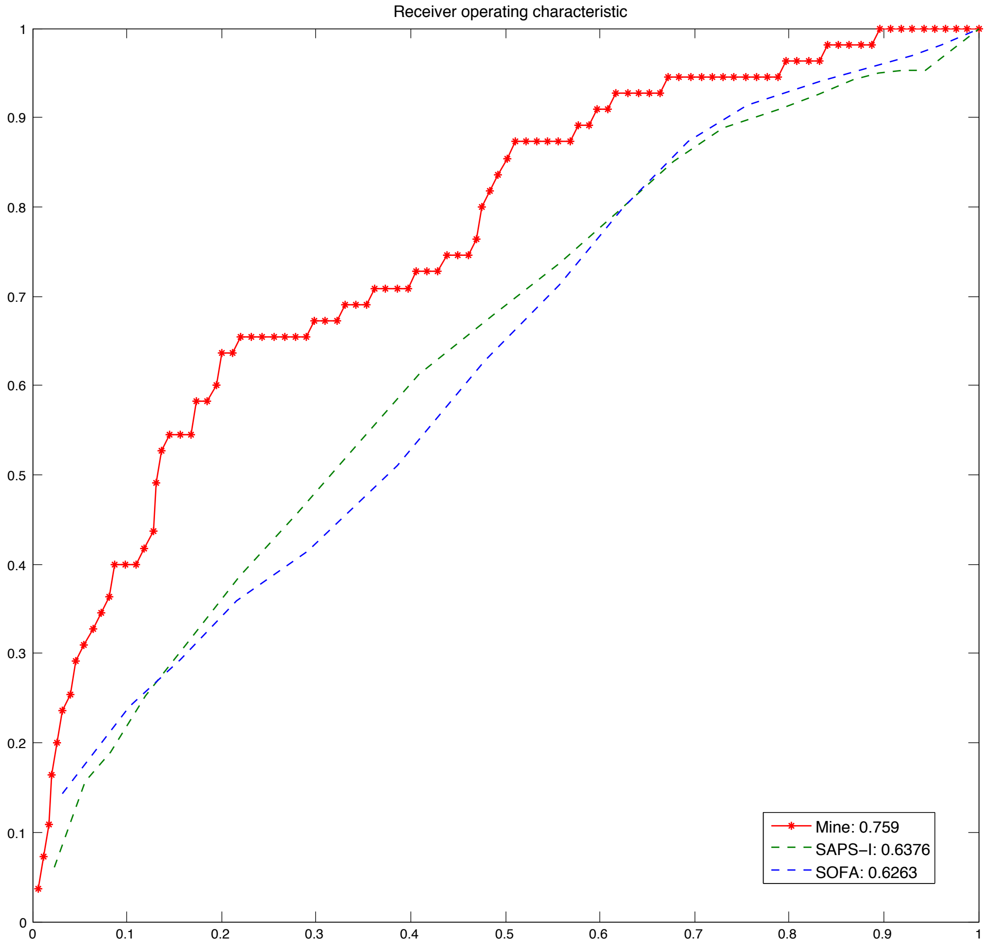
Assuming the values of tests and the counts of tests are independent

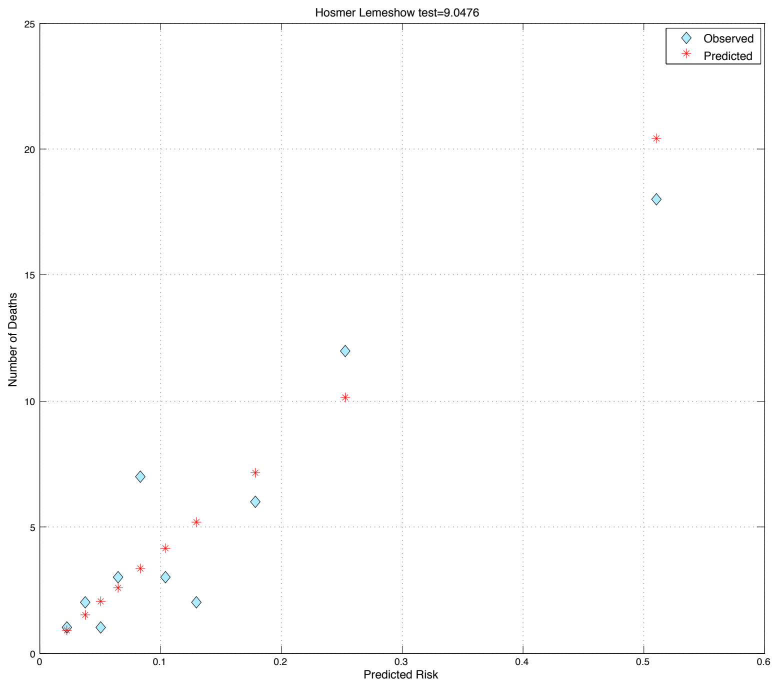
Logistic regression

$$P\left(HM \mid \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{20}\right) = \frac{1}{1 + \exp\left(b + \sum_{i=1}^{n} w_i \cdot f\left(\mathbf{v}_i\right)\right)}$$

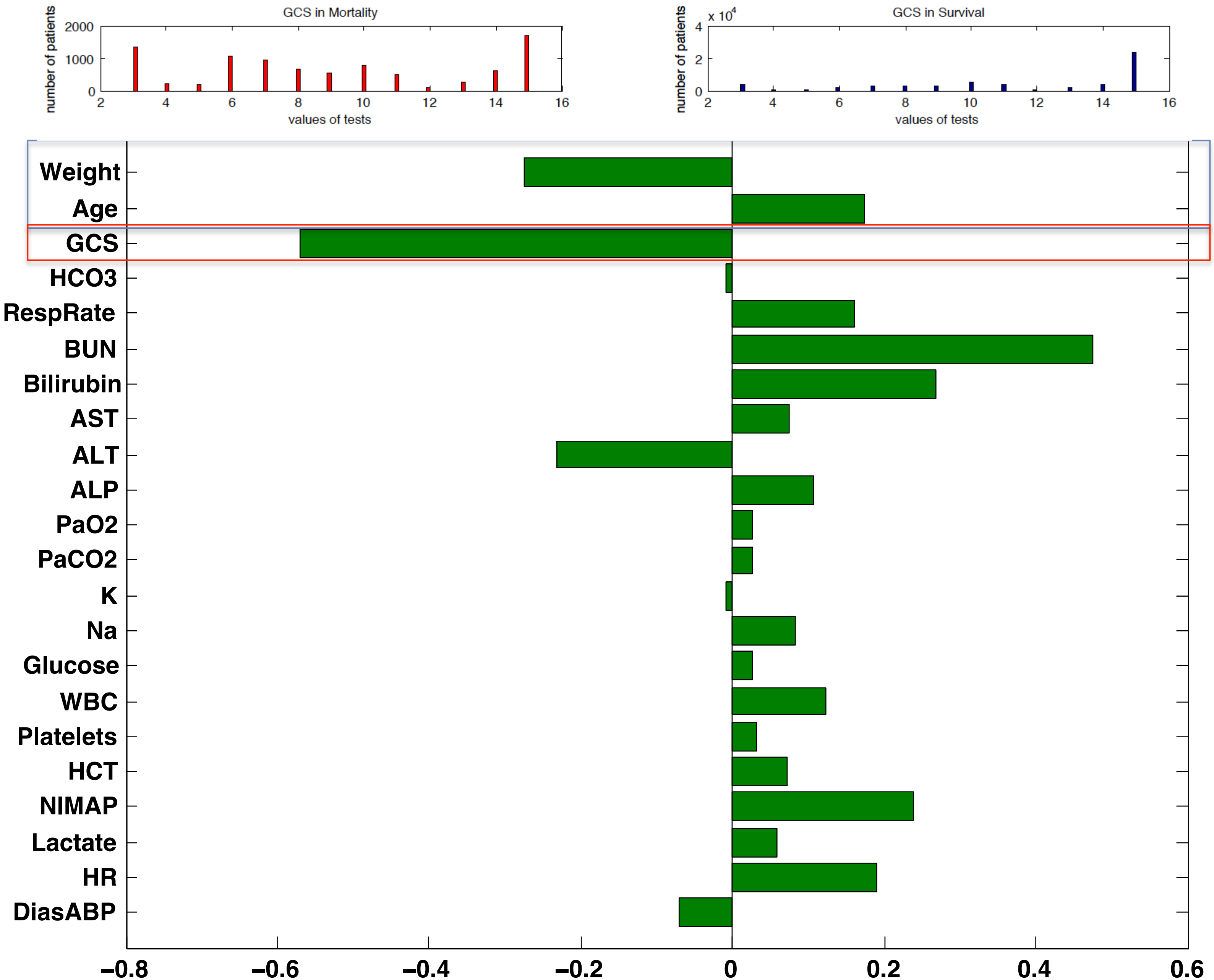
$$\arg\max_{w,b} \sum_{j=1}^{n} \log P\left(HM \mid v_1^j, v_2^j \ldots v_{20}^j\right) - \lambda \sum_i w_i^2$$

Classification evaluation and goodness-of-fit (10-fold CV)





Not very satisfying currently, AUC=0.76 on held-out samples.



Scale and center the variables, so that the weights are comparable.

What's next?

- ***Expected***

-- Try features constructed from standard HMM, Kalman Filter (Implementation is fast, but how to apply to this study?)

-- Incorporate time dependencies between observations (e.g. autoregressive, but cannot be applied directly)

- ***Maximum***

-- Optimize features to achieve better classification performance

-- Documentation

-- partial AUC

Thank you !