Project 2: Classification of Sepsis Based on Predictive Models

BME 580.431 Introduction to Computational Medicine I
Fall 2016
Dr. Sri Sarma

Assigned Reading:

- D. C. Angus, W. T. Linde-Zwirble, J. Lidicker, G. Clermont, J. Carcillo, M. R. Pinsky, Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care. Crit. Care Med. 29, 1303–1310 (2001).

- Katharine E. Henry, David N. Hager, Peter J. Pronovost, Suchi Saria (2015). A targeted real-time early warning score (TREWScore) for septic shock. Science Translational Medicine, Vol 7, Issue 299.

Classification of Sepsis

Sepsis is a systemic inflammatory response to infection complicated by acute organ dysfunction, hyperlactatemia, and/or hypotension refractory to fluid resuscitation [1, 2]. It affects over 19 million people worldwide and accounts for ~10% of the admissions to intensive care units (ICUs) [3-5]. The pathophysiology of sepsis involves a dysregulation of pro-inflammatory responses and anti-inflammatory immunosuppressive responses to pathogens [6], and it includes collateral tissue damage, enhanced susceptibility to secondary infections, altered coagulation, and microvascular thrombosis, which results in tissue hypoperfusion [2, 7, 8]. Tissue oxygenation is further impaired by the loss of barrier function of the endothelium and may lead to organ failure [9, 10]. Sepsis progresses rapidly, and without early detection and treatment can easily evolve into a life-threatening condition. It has been estimated that the mortality rate for sepsis is currently between 20% and 30% [4, 11] and mortality increases by ~8% every hour that treatment is delayed [12].

Scoring systems currently used to diagnose sepsis (e.g., APACHE II, SOFA, SAPS, etc. [13-15]) rely on measurements that either modulate with symptomatic manifestations (e.g., temperature and urine output) or require laboratory tests (e.g., measurements of serum biomarkers) [16-20]. These measurements either capture late stages of the disease or are updated very slowly, based on the frequency of the laboratory tests in a hospitalized environment, thus missing the progression of the disease and making an early diagnosis difficult.

For these reasons there is considerable interest in developing approaches for early classification of sepsis. For example, studies [21-27] have investigated the characteristics of heart rate variation under sepsis conditions in infants admitted to the ICU. More recent studies have constructed multivariate feature vectors based on either various physiological measurements (e.g., heart rate variability, respiration rate, oximetry, etc.) [28-31] or electronic medical record data [32, 33], and have then used machine learning tools on these vectors to identify sepsis conditions.

The data you will be using have been shared with you by email. There are in matlab files and a patient description excel file.

a. **Patient Info.xlsx** - This file contains a table that describes attributes for each patient in addition to hospital check-in and check-out times. This file will not really be used. All the information you need to construct and test models are in the matlab files described below.
b. **clinical_data_training.mat** – This is .mat file that contains a table whose rows correspond to each patient in the training data set and whose columns correspond to values of static variables listed in Table 2.
c. **clinical_data_testing.mat** – This is .mat file that contains a table whose rows correspond to each patient in the test data set and whose columns correspond to values of static variables listed in Table 2.
d. waveform_data_training.mat – This is a .mat file that contains the waveform data sampled every minute for training patients. The rows are PTS features computed from the raw waveform data in mimic, which are listed in Table 1.
e. glm_part1.m – This is a matlab function that you will be required to tweak in order to test various "static" models defined in Part 1 below. This function outputs the probability models for each patient and the actual patient classification (sepsis, non-sepsis).
f. test_performance.m – this is a matlab function that takes as inputs the probability models for each patient and the actual patient classification (sepsis, non-sepsis) and computes a sequence of 2x2 confusion matrices and an ROC curve (described below). It outputs the threshold value that will give the best classification.
g. glm_part2.m – This is a matlab function that you will be required to tweak in order to test various "dynamical" models defined in Part 1 below. This function outputs the probability models for each patient and the actual patient classification (sepsis, non-sepsis).
h. classify_test.m – This is a matlab function that builds your final GLM and tests its performance on the test data set.

Part I: "Static" Models

Our first modeling effort will be to model the probability that a given patient has sepsis or not. We can define p_i as the probability that patient i has sepsis and then build the following GLM:

$$log\left(\frac{p_i}{1-p_i}\right) = \alpha_0 + \alpha_1 x_1 + \alpha_1 x_2 + \alpha_1 x_3 + \cdots \quad (1)$$

where x_j is some feature j such as gender, age, and comoborbidities. In particular, you will have access to the following information for each patient, which can be candidates for features:

>> load clinical data_training.mat (see Table 1 below)

- column_labels has labels for columns in num
- num is a numerical matrix with values indicating clinical "static" features (e.g. age)

Clinical Data (num)	**Description**
Sepsis Status	0=nonspesis, 1=sepsis
Gender	1=Male, 0=Female
Age	Age
Respitory Comorbidities	1=has comorbidities, 0=doesn't have

Heart Comorbidities	1=has comorbidities, 0=doesn't have
Infection	1=has comorbidities, 0=doesn't have

Table 1

1. Constructing GLMs for Static Model

You may want to explore the dataset to see which variables from the clinical data table separate sepsis from non-sepsis patients. These variables will be good candidates for covariates (features) in the model. Once you feel that you have a collection of informative features, you can start building GLMs to evaluate their predictive power. We divided all subjects into a training set and a test set (~80% training and 20% test). Construct a GLM for a given set of features by modifying **glm_part1** using only training data. You will want to test different models and select the model that, say, has the smallest deviance (or largest likelihood).

Note that the model (1) is a GLM that assumes that the observations are generated by independent samples of a Bernoulli random variable parameterized by p. Use **glm_part1.m** to estimate the parameters $\theta = \{\alpha_0, \alpha_1, \dots\}$ by maximizing the data likelihood function. Examine the parameter estimates and their 5% and 95% confidence bounds to test whether any of the parameters are statistically significantly different from 0. For example, if the p-value associated with $\alpha_1 > 0.05$, then the covariate feature x_1 is likely not informative-hence throw it out of your model. Note that the p-value is a variable in the structure "stats", which is an output of glmfit. In this way, you can prune the set of features that you want to include in your model.

Once you settle on a model, define it and describe it. Plot the model, $\hat{p}_i$ for each patient and overlay the 5% and 95% confidence bounds.

2. Evaluating your GLMs

To test performance, we will select a threshold $0 < \gamma < 1$. Then, for each test patient, we will compute

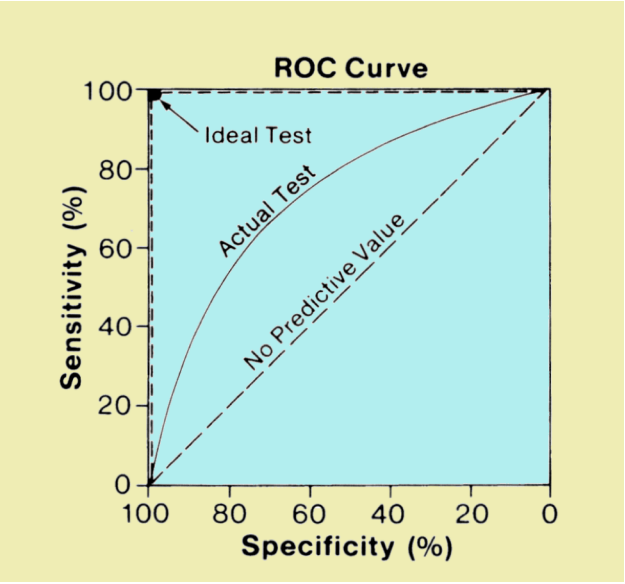
$$\hat{p} = \frac{\exp(\alpha_0 + \alpha_1 x_1 + \alpha_1 x_2 + \alpha_1 x_3 + \cdots)}{1 + \exp(\alpha_0 + \alpha_1 x_1 + \alpha_1 x_2 + \alpha_1 x_3 + \cdots)}.$$

If $\hat{p} > \gamma$, then we will classify as "sepsis", else classify as "non-sepsis". We will tabulate a 2x2 confusion matrix also known as a contingency table or an error matrix as highlighted below.

From the confusion matrix, we will compute TPR (sensitivity) and FPR (specificity) for the chosen γ. Specifically, we will γ from 0.01: 0.99 and draw what is called a Receiver Operating Characteristic (ROC) curve, which is a plot of TPR vs. FPR as a function of γ.

		True class: Positive	True class: Negative	Measures
Predicted class	Positive	True positive *TP*	False positive *FP*	Positive predictive value (PPV) $\frac{TP}{TP+FP}$
	Negative	False negative *FN*	True negative *TN*	Negative predictive value (NPV) $\frac{TN}{FN+TN}$
Measures		Sensitivity $\frac{TP}{TP+FN}$	Specificity $\frac{TN}{FP+TN}$	Accuracy $\frac{TP+TN}{TP+FP+FN+TN}$

Often, the "area under the ROC curve" is used to measure how "good" your classifier is. **test_performance.m** computes the ROC curve and its associated area for each model you build and compares their performance. You will have to complete the missing code in test_performance.m for it to run! Report your results for your final GLM (Accuracy, ROC, AUC etc) and comment.



3. Classification of Sepsis (testing model predictions on test data)

Now that you have a GLM that you believe in, it has to be put to the test on data it has never seen before! Modify **classify_test.m** to implement and test your GLM on the test data set. Report the performance of your final GLM on test data and compare it's performance to that on the training data set. Did it to better or worse than when tested on training data? Discuss.

Part II: "Dynamical" Models

1. Constructing GLMs for Dynamical Model

Now consider a dynamical GLM where p(t)= Pr(sepsis at time t) is a function of your chosen features (some of which vary over time), i.e.,

$$log\left(\frac{p(t)}{1-p(t)}\right) = \alpha_0 + \alpha_1 x_1(t) + \alpha_1 x_2(t) + \alpha_1 x_3(t) + \cdots$$

where $x_i(t)$ is feature i for $i = 1,2,\ldots$summarized at time t. An example could be the mean of heart rate $x_i(t)$ =HR(t). Repeat the feature selection and modeling process above by editing glm_part2.m

This model can include both demographic and waveform data, in which case many rows of X will have the same demographic data, even as the time dependent data differs. Notice, in this case you are now classifying each time point separately, as opposed to just assigning a sepsis status to each patient.

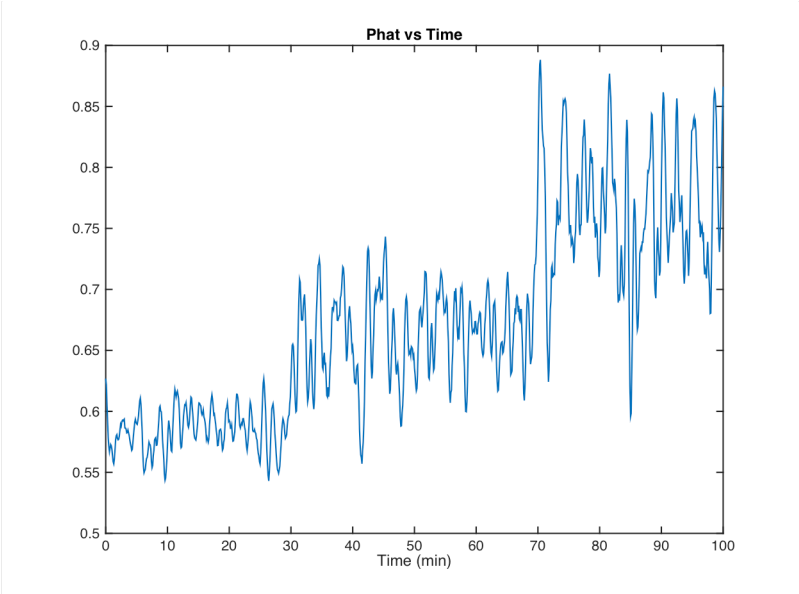
You may also come up with a decision rule that predicts the patient's status based on their whole recording. For instance, all subjects who had more than 50% of their time series recording classified as sepsis could be labeled as septic. This would output a patient-specific classification as in the simpler model.

Justify whichever you think is more useful.

2. Classification of Sepsis (examining and testing your model)

Classification can be accomplished using the same principles as in Part 1.

Once you have a model, you can plot $\hat{p}(t)$ on your test data. Consider why this could be useful. You may get something like the figure below. You can also overlay the 5% and 95% confidence bounds understand how certain your prediction is.



APPENDIX

Subjects

A total of 34 adult ICU patients from the MIMIC-II database were included in the study. A list of the MIMIC-II identification numbers (IDs) associated with the patients in the database is reported in Table 3 (one ID per patient). For each patient, the following information is reported: the overall duration of every PTS sequence used in the study, whether or not the patient was diagnosed as septic during the ICU stay, and a list of relevant comorbidities diagnosed at the time of admittance to ICU (Table 3). The ICD-9 code (International Classification of Diseases) of each comorbidity is reported. Each patient included in the study satisfied the following criteria: (i) the patient was an adult (i.e., age 21 or older); (ii) the patient was admitted to ICU one or more times and in at least one case the admission included an ICD-9 code for sepsis, severe sepsis, or sepsis shock (ICD-9 codes 995.x) [1]; and (iii) in at least one ICU stay, continuous electrocardiogram (ECG), respiratory (RESP), photoplethysmograph (PPG), and blood oxygen saturation (SpO2) measurements were recorded simultaneously for at least 3 consecutive hours.

Preprocessing of Waveform Data:

For each patient, MIMIC-II ECG (lead II), PPG and RESP signals were acquired with a sampling rate of 125 Hz. The S_pO_2 signal, instead, was averaged over 60-s-long consecutive windows with no overlap before storage, thus resulting in a sampling rate of 1 Hz. R-wave peaks (R-peaks) were extracted from the ECG signal as in and then used to compute heart rate (HR) and heart rate variability (HRV). In particular, HR was estimated as the number of R-peaks over consecutive non-overlapping windows (1-min-long each) and HRV was estimated as the variance of the R-R intervals over consecutive 5-min-long windows (sliding by 1 min), thus resulting in a sequence of HR and HRV values, one per minute. Similarly, the RESP signal was processed to extract the wave peaks and the respiratory rate (r_{RESP}) was computed as the number of wave peaks over consecutive non-overlapping 1-min-long windows. For the S_pO_2 signal, instead, 10-min-long consecutive windows were used (sliding by 1 min) and, for each window, the mean and variance of the S_pO_2 samples (μ_{SO} and σ_{SO}, respectively) were computed, thus resulting in a sequence of μ_{SO} and σ_{SO} values, one per minute.

Because of the heterogeneity of the data sources, the HR and r_RESP series, the R-R intervals, and SpO2 samples were normalized according to the formula:

$$\hat{s}(t) = \frac{s(t) - m_s}{M_s - m_s} \quad (1)$$

where s is any of HR, r_RESP, R-R intervals, and SpO2, s(t) is the sample at the generic time t≥ 0, and m_s and M_s are the lower bound and upper bound of the expected range of values of s(t) in normal subjects, respectively, see Table 4. The PPG

signal was high-pass filtered (4th order Butterworth filter, cutoff at 0.02 Hz) to remove the DC component and divided into consecutive 5-min-long windows (sliding by 1 min). Then, in each window, the power spectrum was computed with the Welch method (sub-window size: 1 min, no overlap) and the fraction of power in the frequency bands [0.04, 0.15] Hz (P_{LF}), [0.15, 0.3] Hz (P_{MF}), and [0.3, 0.5] Hz (P_{HF}) was computed. Frequency bands LF and HF were chosen as in to capture the sympathetic control over the peripheral circulation (LF) and the effects of the respiratory rhythm on the venous return (HF), respectively. Frequency band MF, instead, aims to capture the activity of baroreceptors and blood pressure control.

PTS Measurements in Waveform Structure (wav.Fn, wav.Fs)	Description
$\widehat{HR}$	Heart rate
$\widehat{HRV}$	Heart Rate Variability
$\hat{r}_{RESP}$	Respiratory rate
$\hat{\mu}_{SO}$	the mean of the S_pO_2
$\hat{\sigma}_{SO}$	the variance of the S_pO_2
P_{LF}	Low frequency power of PPG
P_{MF}	Mid frequency power of PPG
P_{HF}	High frequency power of PPG

Table 2: Waveform Data

Table 3: Patient Information

	ID	Sex / Age (y)	Time (h)	State	Comorbidities
1	s00618	M / 78	72	S	CHF, UTI
2	s00801	F / 46	768	S	ARF, UTI, BI
3	s01995	F / 73	96	S	CHF, DM, PNM
4	s05646	M / 57	120	SS	DM, BI, UTI
5	s07614	M / 81	48	S	ARF, CHF, DM, PNM
6	s08141	M / 82	48	S	ARF, CHF, PNM
7	s10124	F / 83	48	S	ARF, DHF, DM, BI, PNM
8	s10188	M / 62	48	S	ARF, CHF, UTI
9	s10769	M / 82	24	S	PVD, DM
10	s14325	F / 56	72	S	BI
11	s17423	M / 48	96	S	ARF, DM, PNM, UTI
12	s22035	F / 90	60	S	SND, DM, BI
13	s00402	F / 49	192	NS	CHF, DM, UTI
14	s02187	M / 42	48	NS	ARF, CHF, DM, PNM, BI
15	s02513	M / 39	24	NS	CHF, BI
16	s03491	F / 44	48	NS	DM, PNM, UTI
17	s05937	M / 41	240	NS	ARF, CHF, PNM, UTI
18	s06428	F / 21	24	NS	CHF, DHF
19	s14828	F / 60	96	NS	BI
20	s18082	F / 90	192	NS	ARF, DM, UTI, BI
21	s20124	M / 55	72	NS	CHF, DM, BI
22	s10152	M / 74	144 24	S NS	CHF, PNM CHF, SHF, UTI, PNM
23	s10653	F / 87	14 72	S NS	ARF, DHF, DM, BI CHF, DHF, DM, UTI
24	s11342	F / 44	192 72	S NS	ARF, DHF CHF, DHF, HPT
25	s14579	M / 65	96 48	S NS	CHF, DM, UTI, PNM, BI CHF, BI
26	s19411	M / 59	7 72	S NS	CHF, SHF, DM, PNM CHF, SHF, DM

"ID" is the unique identifier of the patient in the MIMIC-II database. "Time" is the duration of the PTS data. "State" indicates whether the patient was diagnosed septic (S), severe sepsis (SS) or non-septic (NS) at the time of recording. "Comorbidities" denote relevant diagnosis annotated in the patients' own records at the time of recording. ARF=acute respiratory failure (ICD-9: 518.81); BI=bacterial infection (ICD-9: 038.11, 0.38.42, 041.04, 041.11, 041.3, or 041.4); CHF=congestive heart failure (ICD-9: 428); CLG=cholangitis (ICD-9: 576.1) DHF=diastolic heart failure (ICD-9: 428.33); DM=diabetes mellitus (ICD-9: 250); F=female; HPT=hepatitis (ICD-9: 070.54 or 573.3); M=male; MNG=chronic meningitis (ICD-9: 322.2); PNM=pneumonia (ICD-9: 481, 482.83, 486, or 487); PVD=peripheral vascular disease (ICD-9: 443.9); SHF=systolic heart failure (ICD-9: 428.2); SND=sinoatrial node dysfunction (ICD-9: 427.81); UTI=urinary tract infection (ICD-9: 599).

Table 4. Range of physiologic signals in normal subjects. HR=heart rate (bpm); r_{RESP}=respiratory rate (breath per min); R-R intervals (ms); S_pO_2 (%).

Signal	Lower Bound	Upper Bound
R-R interval	60	100
r_{RESP}	12	20
HR	60	100
S_pO_2	90	100

Table 5. Population-averaged PTS components. Average value (mean±S.D.) of the components of the normalized vector $\boldsymbol{v}(t)$ across the population in sepsis (n =133,615 samples) and non-sepsis state (n =102,848 samples). For each entry in the table, the p-value for the comparison sepsis vs. non-sepsis was $p < 0.001$ (Wilcoxon rank-sum test).

PTS	Non-sepsis	Sepsis
$\widehat{HR}$	0.0945±0.0760	0.0823±0.0746
$\widehat{HRV}$	0.1469±0.0654	0.1339±0.0621
$\hat{r}_{RESP}$	0.1075±0.0723	0.0950±0.0726
$\hat{\mu}_{SO}$	0.0965±0.1345	0.04851±0.0858
$\hat{\sigma}_{SO}$	1.8232±2.1821	1.8151±1.5706
P_{LF}	0.6273±0.4928	0.6801±0.4703
P_{MF}	0.6347±0.2973	0.6364±0.2969
P_{HF}	0.0208±0.070	0.02547±0.1476

References

[1] R. P. Dellinger, M. M. Levy, A. Rhodes, D. Annane, H. Gerlach, S. M. Opal*, et al.*, "Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012," *Crit Care Med,* vol. 41, pp. 580-637, Feb 2013.

[2] D. C. Angus and T. van der Poll, "Severe sepsis and septic shock," *N Engl J Med,* vol. 369, pp. 840-51, Aug 29 2013.

[3] N. K. Adhikari, R. A. Fowler, S. Bhagwanjee, and G. D. Rubenfeld, "Critical care and the global burden of critical illness in adults," *Lancet,* vol. 376, pp. 1339-46, Oct 16 2010.

[4] D. C. Angus, W. T. Linde-Zwirble, J. Lidicker, G. Clermont, J. Carcillo, and M. R. Pinsky, "Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care," *Crit Care Med,* vol. 29, pp. 1303-10, Jul 2001.

[5] W. T. Linde-Zwirble and D. C. Angus, "Severe sepsis epidemiology: sampling, selection, and society," *Crit Care,* vol. 8, pp. 222-6, Aug 2004.

[6] T. van der Poll and S. M. Opal, "Host-pathogen interactions in sepsis," *Lancet Infect Dis,* vol. 8, pp. 32-43, Jan 2008.

[7] M. Levi and T. van der Poll, "Inflammation and coagulation," *Crit Care Med,* vol. 38, pp. S26-34, Feb 2010.

[8] J. S. Boomer, K. To, K. C. Chang, O. Takasu, D. F. Osborne, A. H. Walton*, et al.*, "Immunosuppression in patients who die of sepsis and multiple organ failure," *Jama,* vol. 306, pp. 2594-605, Dec 21 2011.

[9] H. F. Galley, "Oxidative stress and mitochondrial dysfunction in sepsis," *Br J Anaesth,* vol. 107, pp. 57-64, Jul 2011.

[10] N. M. Goldenberg, B. E. Steinberg, A. S. Slutsky, and W. L. Lee, "Broken barriers: a new take on sepsis pathogenesis," *Sci Transl Med,* vol. 3, p. 88ps25, Jun 22 2011.

[11] G. Kumar, N. Kumar, A. Taneja, T. Kaleekal, S. Tarima, E. McGinley*, et al.*, "Nationwide trends of severe sepsis in the 21st century (2000-2007)," *Chest,* vol. 140, pp. 1223-31, Nov 2011.

[12] A. Kumar, D. Roberts, K. E. Wood, B. Light, J. E. Parrillo, S. Sharma*, et al.*, "Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock," *Crit Care Med,* vol. 34, pp. 1589-96, Jun 2006.

[13] F. Geier, S. Popp, Y. Greve, A. Achterberg, E. Glockner, R. Ziegler*, et al.*, "Severity illness scoring systems for early identification and prediction of in-hospital mortality in patients with suspected sepsis presenting to the emergency department," *Wien Klin Wochenschr,* vol. 125, pp. 508-15, Sep 2013.

[14] N. O. Ghanem-Zoubi, M. Vardi, A. Laor, G. Weber, and H. Bitterman, "Assessment of disease-severity scoring systems for patients with sepsis in general internal medicine departments," *Crit Care,* vol. 15, p. R95, 2011.

[15] F. Innocenti, S. Bianchi, E. Guerrini, S. Vicidomini, A. Conti, M. Zanobetti*, et al.*, "Prognostic scores for early stratification of septic patients admitted to an emergency department-high dependency unit," *Eur J Emerg Med,* vol. 21, pp. 254-9, Aug 2014.

[16] F. A. Bozza, J. I. Salluh, A. M. Japiassu, M. Soares, E. F. Assis, R. N. Gomes*, et al.*, "Cytokine profiles as markers of disease severity in sepsis: a multiplex analysis," *Crit Care,* vol. 11, p. R49, 2007.

[17] T. Chan and F. Gu, "Early diagnosis of sepsis using serum biomarkers," *Expert Rev Mol Diagn,* vol. 11, pp. 487-96, Jun 2011.

[18] A. Nierhaus, J. Linssen, D. Wichmann, S. Braune, and S. Kluge, "Use of a weighted, automated analysis of the differential blood count to differentiate sepsis from non-infectious systemic inflammation: the intensive care infection score (ICIS)," *Inflamm Allergy Drug Targets,* vol. 11, pp. 109-15, Apr 2012.

[19] P. Paulus, C. Jennewein, and K. Zacharowski, "Biomarkers of endothelial dysfunction: can they help us deciphering systemic inflammation and sepsis?," *Biomarkers,* vol. 16 Suppl 1, pp. S11-21, Jul 2011.