Wydawnictwo UR 2023
ISSN 2544-1361 (online)
doi: 10.15584/ejcem.2023.4.27

http://www.ejcem.ur.edu.pl
European Journal of Clinical and Experimental Medicine
Eur J Clin Exp Med 2023; 21 (4): 830–834

ORIGINAL PAPER

Utilizing machine learning to create a blood-based scoring system for sepsis detection

Sadik Aref

Leigh High School, San Jose, United States

ABSTRACT

Introduction and aim. Sepsis, a disease caused by inflammation as a response to infection, often goes undiagnosed due to its heterogeneity and lack of a single diagnostic test. Current sepsis detection scoring systems have low sensitivity and utilize biomarkers that are difficult to obtain from a single test. The goal of this research is to create a scoring system that outperforms current industry standards by utilizing blood-based biomarkers readily available in hospital settings.

Material and methods. Machine learning algorithms were run through Google Colab using Extreme Gradient Boost classifier. The dataset was obtained from NCBI website containing electronic hospital records of intensive care patients. A multivariate linear regression was applied to the dataset to determine statistically significant biomarkers in the detection of sepsis, and their β coefficients. Then, validation testing was performed, and the performance was compared to other scoring systems.

Results. This experiment reveals that a sepsis detection system that utilizes procalcitonin, white blood cells, C-reactive protein, neutrophil to lymphocyte ratio, and albumin can outperform other biomarkers and scoring systems with high sensitivity at a recall score of 0.7922.

Conclusion. These results demonstrate the potential of utilizing a blood-based scoring system for sepsis detection within hospital settings.

Keywords. blood-based, machine learning, sepsis

Introduction

Sepsis is defined as life-threatening organ dysfunction caused by a dysregulated host response to infection.[1] It may lead to shock, multi-organ failure, and death – especially if not recognized early and treated promptly.[2,3] Sepsis is a Global Health crisis with 47–50 million cases each year of which 11 million people die, as of 2020. Most cases of sepsis are caused by bacterial infections, however, other infections like viral infections, such as COVID-19 or influenza, or fungal infections can also cause sepsis.[4] Sepsis is hard to diagnose because it is a heterogeneous syndrome whose course depends on different pathophysiological mechanisms, complexity in clinical context, and clinical phenotypes. Hospitals use scoring systems such as Acute Physiology and Chronic Health Evaluation (APACHE), Sequential Organ Failure Assessment (SOFA), and Quick SOFA (qSOFA) for sepsis detection; however they have low sensitivity and often require biomarkers from multiple different tests.[5] The SOFA score numerically quantifies the number and severity of failed organs. It provides potentially valuable prognostic information on in-hospital survival when applied to patients with severe sepsis with evidence of hypoperfusion at the time of emergency department presentation.[6] The Logistic Organ Dysfunction Score (LODS) is a tool to identify patients at high risk of developing postoperative severe sepsis and the need for early goal-directed therapy.[7] APACHE is a successful scoring system assessing severity

Corresponding author: Sadik Aref, e-mail: sadaref01@gmail.com

Received: 23.07.2023 / Revised: 28.08.2023 / Accepted: 16.09.2023 / Published: 30.12.2023

Aref S. Utilizing machine learning to create a blood-based scoring system for sepsis detection. *Eur J Clin Exp Med*. 2023;21(4):830–834. doi: 10.15584/ejcem.2023.4.27.

of illness and prognosis of intensive care unit patients and has been evaluated and validated in patients for mortality outcome.[8] MODS refers to the clinical syndrome in which at least two systems or organs suffer from simultaneous or sequential dysfunction during major surgery, infection, shock, poisoning and severe trauma, thus undermining the stability of the internal environment.[9] Systemic inflammatory response syndrome (SIRS) is an exaggerated defense response of the body to a noxious stressor (infection, trauma, surgery, acute inflammation, ischemia or reperfusion, or malignancy) to localize and then eliminate the endogenous or exogenous source of the insult. SIRS with a suspected source of infection is termed sepsis.[10] There is still no current standard for sepsis detection scoring systems. Instead, blood culture is considered gold standard for the confirmation of bacteremia which can isolate and identify the causative agent and subsequently test the antimicrobial sensitivity, but the delayed process of bacterial culture emphasizes the importance of a reliable and fast sepsis detection system that utilizes only a few biomarkers that are readily available in hospital and clinical settings.[11]

Procalcitonin (PCT) is a protein that is synthesized in the thyroid gland as an immediate response to inflammation, making it a good biomarker for sepsis detection.[11,12] In everyday clinical practice, worldwide, white blood cells (WBC) and C-reactive protein (CRP) still represent the cornerstones on which is based the diagnosis and prognosis of infected patients together with other signs and symptoms like fever, tachycardia, and tachypnea, when a documented or suspected infection is detected.[13] PCT, CRP and WBC can be combined as effective indicators for the identification of acute bacterial or no-bacterial infections in children. Albumin can be another biomarker for sepsis as their levels are associated with short-term and long-term outcomes in sepsis.[15] C-reactive protein (CRP) and serum albumin (ALB) are useful markers that can predict morbidity and mortality among critically ill patients. This is because CRP effectively reflects acute-phase inflammation while ALB may reflect malnutrition among critically ill patients.[16] The neutrophil-lymphocyte ratio (NLR) is an inflammatory biomarker that uses two types of white blood cells (neutrophils and lymphocytes) that when out of equilibrium can indicate systemic inflammation. NLR serves as an effective biomarker, as neutrophils are typically an immediate response to infection, while lymphocytes are not.[17] Together, PCT, CRP, WBC, Alb, and NLR serve as the biomarkers used for sepsis detection in this machine learning model. As of recent, machine learning has been a point of focus in the biomedical research space, as it allows for automation in the healthcare industry, minimizing human error and maximizing efficiency. Still, current research on sepsis diagnostic scoring systems is limited. Previous studies have explored procalcitonin as a possible biomarker for sepsis detection, however sensitivity was below satisfactory, and the biomarkers chosen are not all blood-based and include comorbidities, which may create a less reliable model due to high correlation.[18]

Aim

The aim of this study was to create a sepsis detection artificial intelligence model utilizing only blood-based biomarkers readily available in hospital and clinical settings that outperforms current sepsis scoring systems.

Material and methods

Dataset

The dataset used in this study was obtained from the National Center for Biotechnology Information website.[18] The electronic health records were obtained from Saint Mary's Hospital Luodong, Taiwan and was approved by the Institutional Review Board of Saint Mary's Hospital Luodong (approval # SMHIRB_105012). The data were analyzed anonymously. The dataset consisted of 258 critically ill patients with sepsis group (n=115) and the non-sepsis group (n=143). It had 79 features including demographics, comorbidities, clinical variables, scoring systems, and the Sepsis label.

Preprocessing

The dataset was then preprocessed using Python on Google Colab. First, features irrelevant or highly correlated to sepsis were removed. These features include sex, comorbidities, and features specific to the hospital the data was collected in. Ultimately, 27 features were removed including 'sex', 'IPS' 'ICU_BilT', 'inhos_mortality', 'Reasons_ICUadm', 'Bacteremia', 'UTI', 'Pneumonia', 'Skin_infection', 'Other_infection', "Infection", 'CS_CADMI', 'CS_CHF', 'CS_CKD', 'CS_PADPAOD', 'CS_CVA', 'CS_Dementia', 'CS_CLD', 'CS_PUD', 'CS_HTN',CS_DM', 'CS_LD', 'CS_cancer', 'ICU_day', 'Division_ICUadm,"'VPAP_A"',"CPCR". After analyzing the dataset, it was evident that the feature of CRP had null values for several patients, so data imputation was utilized through filling in blank spaces with median values. The feature 'Sepsis_3', which is the label, was also removed from the dataset before performing the multivariate linear regression. The test procedure flow chart is shown in Figure 1.

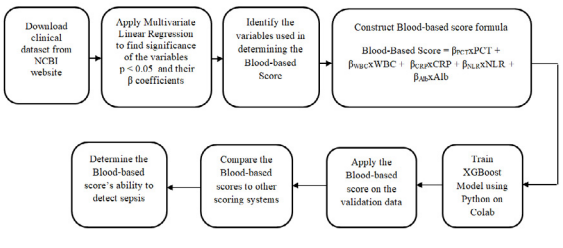


Fig. 1. Flowchart of procedure

Multivariate linear regression

Continuing on Google Colab, a multivariate linear regression analysis was applied to the dataset to determine the p values and the β coefficients for the different features. The five blood-based features that were determined to be statistically significant ($p<0.05$) in detecting sepsis include PCT, CRP, WBC, Alb, and NLR. The formula for the blood-based scoring system was set as Blood-based score=β_{PCT}xPCT + β_{CRP}xCRP + β_{WBC}xWBC + β_{Alb}xAlb + β_{NLR}xNLR. The blood-based scoring system was added back to the dataset as a feature and 'Sepsis_3' was added back as the label.

XGBoost model testing

Due to its speed, ease of use, and superior performance, XGBoost was the machine learning classifier used in this experiment.[19] The scoring system was then validated through the model and performance was compared to PCT alone and other sepsis scoring systems including SOFA, APACHE, LODS, MODS, and SIRS. The AUC curve was plotted, and the accuracy and recall scores were considered.

Results

Multivariate linear regression

The results of the multivariate linear regression along with the features and their p values and β coefficient values are shown in Table1. There were a total of nine features that were statistically significant ($p<0.05$) in detecting sepsis; however, excluding those that are scoring systems and biomarkers that can't be extracted from a blood test, only 5 biomarkers fit the criteria. These blood-based biomarkers include PCT, CRP, WBC, Alb, and NLR. They had p values of 0.042, 0.041, 0.039, 0.001, and 0.001 respectively. Additionally, the β coefficient values for these statistically significant biomarkers were 0.0202, -0.0002, -0.0155, -1.4525, and 0.0419 respectively.

XGBoost model performance

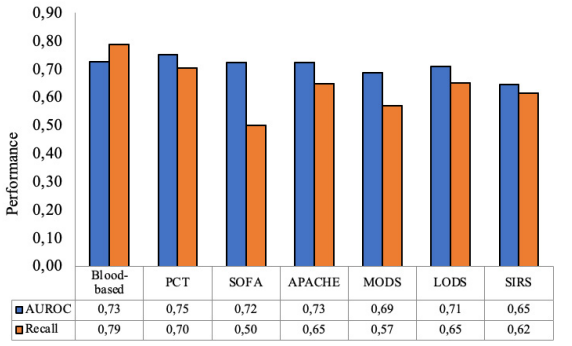
Once the formula β_{PCT}xPCT+β_{CRP}xCRP+β_{WBC}xWBC+β_{Alb}xAlb+β_{NLR}xNLR was applied to create the Blood-based score, training and validation testing was performed. The blood-based scoring system created, outperformed all other scoring systems, as it had a recall of 0.7922, while PCT alone had a recall of 0.7047, SOFA had a recall of 0.5044, APACHE had 0.6480, MODS had 0.5748, LODS had 0.6520, and SIRS had 0.6146. However, the blood-based score underperformed in other metrics with an AUROC of 0.7321, an accuracy of 0.6154, and a precision of 0.4615. The performance of the various scoring systems is shown in Table 2. A graph showing their AUROC and recall is shown in the Figure 2. The ROC (Receiver Operating Characteristic) curve is a graph of false positive rate, which is the x-axis, and true positive rate which is the y-axis. AUROC (Area Under Receiver Operating Characteristic Curve) is a metric for assessing an ROC curve's performance. A greater AUROC indicates a higher model performance, through minimal false positives and a higher rate of true positives. The greatest possible AUROC is 1. The AUROC curves of the scoring systems in shown in Figure 3.

Table 1. Multivariate linear regression results: p values and β coefficients for features

Feature	p	β coefficient
ICU_BT (body temperature)	0.473	0.1408
ICU_HR (heart rate)	0.463	0.0089
ICU_RR (respiratory rate)	0.516	0.0189
ICU_SPO_2 (oxygen Saturation)	0.002	0.1056
ICU_MAP (mean arterial pressure)	0.029	-0.0209
ICU_GCS (Glascow coma scale)	0.155	-0.1419
ICU_Cr (creatinine)	0.759	0.0336
ICU_PCT (procalcitonin)	0.042	0.0202
bCr (breakpoint cluster region)	0.97	0.4722
ICU_CRP (C-reactive protein)	0.041	-0.0002
ICU_BUN (blood urea nitrogen)	0.853	0.0015
ICU_eGFR (estimated glomerular filtration)	0.774	-0.0012
ICU_Ca (calcium)	0.539	0.1477
ICU_GLU (glucose)	0.968	-5.206e-05
ICU_Alb (albumin)	0.001	-1.4525
ICU_GOT (glutamic-ocaloacetic transaminase)	0.934	0.0002
ICU_GPT (glutamic-pyruvic trandaminase)	0.413	-0.0041
ICU_Na (sodium)	0.707	-0.0083
ICU_K (potassium)	0.661	-0.0938
ICU_PH (potential of hydrogen)	0.221	-2.8951
ICU_PCO_2 (partial pressure of oxygen)	0.403	-0.0151
ICU_ PO_2 (partial pressure of oxygen)	0.073	-0.0089
ICU_HCO_3 (bicarbonate)	0.401	0.0343
ICU_SO_2 (sulfur dioxide)	0.03	0.0380
ICU_WBC (white blood cells)	0.039	-0.0155
ICU_HB (hemoglobin)	0.665	0.2543
ICU_HCT (hematocrit)	0.784	-0.0530
ICU_NLR (neutrophil to lymphocyte ratio)	0.001	0.0419
ICU_PLT (platelet count)	0.514	0.0015
qSOFA (quick sequential organ failure assessment)	0.384	-0.3493
MODS (multiple organ dysfunction score)	0.656	-0.0873
APACH (APACHE II)	0.474	-0.0397
SIRS (systemic inflammatory response syndrome)	0.566	0.1769
LODS (logistic organ dysfunction system)	0.158	-0.2167
SOFA (sequential organ failure assessment)	0.02	0.3389

Table 2. Validation testing results for Blood-based Score, PCT and other scoring systems

Scoring System	Recall	AUROC	Accuracy	Precision
Blood-based Score	0.7922	0.7321	0.6154	0.4615
PCT	0.7047	0.7549	0.75	0.5909
SOFA	0.5044	0.7370	0.6731	0.5926
APACHE	0.6480	0.7370	0.5577	0.5833
MODS	0.5748	0.6851	0.7115	0.6522
LODS	0.6520	0.7144	0.7115	0.6429
SIRS	0.6146	0.6488	0.5769	0.5667



	Blood-based	PCT	SOFA	APACHE	MODS	LODS	SIRS
AUROC	0,73	0,75	0,72	0,73	0,69	0,71	0,65
Recall	0,79	0,70	0,50	0,65	0,57	0,65	0,62

Fig. 2. Graph of scoring systems' performance (AUROC in blue and recall in red)

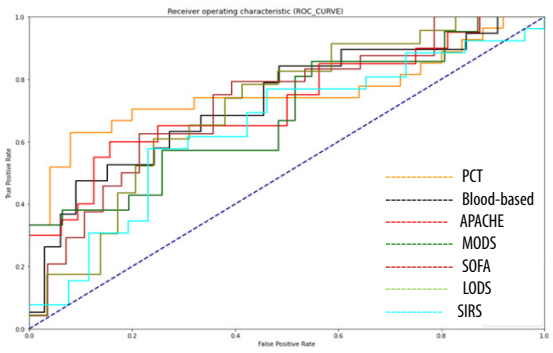


Fig. 3. AUROC Curve

Discussion

Hospital mortality from sepsis has ranged from 25% to 80% over the last few decades. Although mortality may be lower in recent years, sepsis is clearly still a very serious condition.[20] This study has developed a machine learning model that detects sepsis utilizing a combination of five blood-based biomarkers procalcitonin, white blood cells, C-reactive protein, neutrophil to lymphocyte ratio, and albumin which outperformed procalcitonin biomarker as well as other scoring systems such as APACHE, MODS, SOFA, LODS and SIRS with a higher sensitivity. A study conducted by Luo et al., 2009 found that LODS, SOFA and MODS show a good discrimination power, while maximum LODS is of the highest discrimination power to predict the outcome of patients with severe sepsis.[21] The blood-based biomarker developed in this study outperformed the LODS scoring system. A recent study conducted by Peipei Liang and Feng Yu supports the effective use of blood-based biomarkers in sepsis prognosis. They concluded that NLR, CRP, and PCT have important clinical applications in the assessment of the extent of disease and prognosis of patients with bloodstream infection and sepsis.[22] Interestingly, another study comparing the effectiveness of PCT and CRP concluded that PCT concentrations during multiple organ dysfunction syndrome provides more information about the severity and the course of the disease than that of CRP.[23] The current study is in line with the research conducted by Claudia Gregoriano. He concluded that PCT is very effective in detecting sepsis, but yet should be used as a complementary tool combined with available clinical and diagnostic parameters.[24] A similar study conducted by Tsui et al., 2021 used machine learning to develop a novel PCT-based score which outperformed existing sepsis biomarkers and scoring systems in sepsis detection. This PCT-based score was composed of five predictors, including higher serum PCT level, higher NLR, lower albumin level, diabetes mellitus, and necessitating vasopressor and performed at a sensitivity of 0.70.[18]

The blood-based scoring system developed in this study had a slightly lower AUROC than the PCT biomarker. But it had the highest recall score of 0.79, which is considered the most important metric of performance, as higher recall scores indicate minimal false negatives to minimize mortality. Even though the blood-based scoring system demonstrated satisfactory results relative to the other scoring systems, recall and AUC scores for the model were still considerably low. It also had lower precision than other scoring systems, which could lead to excessive medical intervention for healthy patients. To optimize performance, a greater balance between recall and precision scores must be met. The major limitation of this study is the limited data availability of sepsis patients. Further research would include testing of the model with larger datasets from Medical Information Mart for Intensive Care (MIMIC), which will lead to more accurate β coefficients, hence improving the performance. In addition to XGBoost, various other machine learning algorithms could be developed and tested utilizing the same blood-based biomarkers which may lead to higher performance.

Conclusion

This research has explored a novel blood-based scoring system based on PCT, CRP, NLR, WBC, and Alb. Because sepsis is a disease that is often acquired in hospital settings, this scoring system has the potential to be used within hospital and clinical settings, due to its sole use of blood-based biomarkers. This study also shows the potential for multivariate linear regression to be used as a tool for creating disease detection scoring systems. In conclusion, this study demonstrates the potential to utilize a blood-based scoring system to detect one of the deadliest and most widespread diseases: sepsis.

Declarations

Funding

This research received no external funding.

Author contributions

Conceptualization, S.A.; Methodology, S.A.; Software, S.A.; Validation, S.A.; Formal Analysis, S.A.; Investigation, S.A.; Resources, S.A.; Data Curation, S.A.; Writ-

ing Original Draft Preparation, S.A.; Writing - Review & Editing, S.A.; Visualization, S.A.; Supervision, S.A.; Project Administration, S.A.

Conflicts of interest
Author declares no conflict of interest.

Data availability
Data will be made available on reasonable request.

Ethics approval
Non applicable.

References

1. Network J. Consensus Definitions for Sepsis and Septic Shock from JAMA Network. Jamanetwork.com. https://sites.jamanetwork.com/sepsis/. Accessed March 10, 2023.
2. Mollura M, Lehman LW, Barbieri R. Assessment of Sepsis in the ICU by Linear and Complex Characterization of Cardiovascular Dynamics. *Annu Int Conf IEEE Eng Med Biol Soc.* 2021;2021:862-865. doi: 10.1109/EMBC46164.2021.9630521
3. Sepsis. Global Sepsis Alliance. https://www.global-sepsis--alliance.org/sepsis. Accessed March 10, 2023.
4. CDC. Sepsis is a medical emergency. ACT FAST. Centers for Disease Control and Prevention. http://www.cdc.gov/sepsis/what-is-sepsis.html. Accessed March 10, 2023.
5. Camacho-Cogollo JE, Bonet I, Gil B, Iadanza E. Machine Learning Models for Early Prediction of Sepsis on Large Healthcare Datasets. *Electronics.* 2022;11(9):1507. doi: 10.3390/electronics11091507
6. Jones AE, Trzeciak S, Kline JA. The Sequential Organ Failure Assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation. *Crit Care Med.* 2009;37(5):1649-1654. doi: 10.1097/ccm.0b013e31819def97
7. Heldwein MB, Badreldin AM, Doerr F, et al. Logistic Organ Dysfunction Score (LODS): A reliable postoperative risk management score also in cardiac surgical patients? *J Cardiothorac Surg.* 2011;6:110. doi: 10.1186/1749-8090-6-110
8. Dahhan T, Jamil M, Al-Tarifi A, Abouchala N, Kherallah M. Validation of the APACHE IV scoring system in patients with severe sepsis and comparison with the APACHE II system. *Crit Care.* 2009;13(Suppl 1):P511. doi: 10.1186/cc7675
9. Sun GD, Zhang Y, Mo SS, Zhao MY. Multiple Organ Dysfunction Syndrome Caused by Sepsis: Risk Factor Analysis. *Int J Gen Med.* 2021;14:7159-7164. doi: 10.2147/ijgm.s328419
10. Chakraborty RK, Burns B. Systemic Inflammatory Response Syndrome. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing; May 29, 2023.
11. Vijayan AL, Vanimaya, Ravindran S, et al. Procalcitonin: a promising diagnostic marker for sepsis and antibiotic therapy. *J Intensive Care*. 2017;5(1):51. doi: 10.1186/s40560-017-0246-8
12. Becker KL, Snider R, Nylen ES. Procalcitonin assay in systemic inflammation, infection, and sepsis: Clinical utility and limitations. *Crit Care Med.* 2008;36(3):941-952. doi: 10.1097/ccm.0b013e318165babb
13. Magrini L, Gagliano G, Travaglino F, et al. Comparison between white blood cell count, procalcitonin and C reactive protein as diagnostic and prognostic biomarkers of infection or sepsis in patients presenting to emergency department. *Clin Chem Lab Med*. 2014;52(10):1465-1472. doi: 10.1515/cclm-2014-0210
14. Li Y, Min L, Zhang X. Usefulness of procalcitonin (PCT), C--reactive protein (CRP), and white blood cell (WBC) levels in the differential diagnosis of acute bacterial, viral, and mycoplasmal respiratory tract infections in children. *BMC Pulm Med*. 2021;21(1):386. doi: 10.1186/s12890-021-01756-4
15. Cao Y, Su Y, Guo C, He L, Ding N. Albumin Level is Associated with Short-Term and Long-Term Outcomes in Sepsis Patients Admitted in the ICU: A Large Public Database Retrospective Research. *Clin Epidemiol.* 2023;15:263-273. doi: 10.2147/clep.s396247
16. Oh TK, Song IA, Lee JH. Clinical usefulness of C-reactive protein to albumin ratio in predicting 30-day mortality in critically ill patients: A retrospective analysis. *Sci Rep.* 2018;8(1):14977. doi: 10.1038/s41598-018-33361-7
17. Liu X, Shen Y, Wang H, Ge Q, Fei A, Pan S. Prognostic Significance of Neutrophil-to-Lymphocyte Ratio in Patients with Sepsis: A Prospective Observational Study. *Mediators Inflamm.* 2016;2016:1-8. doi: 10.1155/2016/8191254
18. Tsui TL, Huang YT, Kan WC, et al. A novel procalcitonin-based score for detecting sepsis among critically ill patients. *PLOS ONE.* 2021;16(1):e0245748. doi: 10.1371/journal.pone.0245748
19. scikit-learn: machine learning in Python — scikit-learn 1.0.2 documentation. scikit-learn.org. http://scikit-learn.org/stable/. Accessed March 20, 2023.
20. Angus DC, Wax RS. Epidemiology of sepsis: An update. *Crit Care Med.* 2001;29:S109-S116. doi: 10.1097/00003246-200107001-00035
21. Luo X, Wang H, Hu S, et al. Comparison of three different organ failure assessment score systems in predicting outcome of severe sepsis. *Zhonghua Wai Ke Za Zhi*. 2009;47(1):48-50. doi: 10.3760/cma.j.issn.0529-5815.2009.01.017
22. Liang P, Yu F. Value of CRP, PCT, and NLR in Prediction of Severity and Prognosis of Patients With Bloodstream Infections and Sepsis. Front Surg. 2022;9:857218. doi: 10.3389/fsurg.2022.857218
23. Meisner M, Tschaikowsky K, Palmaers T, Schmidt J. Comparison of procalcitonin (PCT) and C-reactive protein (CRP) plasma concentrations at different SOFA scores during the course of sepsis and MODS. *Crit Care.* 1999;3(1):45. doi: 10.1186/cc306
24. Gregoriano C, Heilmann E, Molitor A, Schuetz P. Role of procalcitonin use in the management of sepsis. *J Thorac Dis.* 2020;12:5-15. doi: 10.21037/jtd.2019.11.63