

## Scoring System for the Prognosis of Sepsis Patients in Intensive Care Units

**Authors :** Javier E. García-Gallo, Nelson J. Fonseca-Ruiz, John F. Duitama-Munoz

**Abstract :** Sepsis is a syndrome that occurs with physiological and biochemical abnormalities induced by severe infection and carries a high mortality and morbidity, therefore the severity of its condition must be interpreted quickly. After patient admission in an intensive care unit (ICU), it is necessary to synthesize the large volume of information that is collected from patients in a value that represents the severity of their condition. Traditional severity of illness scores seeks to be applicable to all patient populations, and usually assess in-hospital mortality. However, the use of machine learning techniques and the data of a population that shares a common characteristic could lead to the development of customized mortality prediction scores with better performance. This study presents the development of a score for the one-year mortality prediction of the patients that are admitted to an ICU with a sepsis diagnosis. 5650 ICU admissions extracted from the MIMICIII database were evaluated, divided into two groups: 70% to develop the score and 30% to validate it. Comorbidities, demographics and clinical information of the first 24 hours after the ICU admission were used to develop a mortality prediction score. LASSO (least absolute shrinkage and selection operator) and SGB (Stochastic Gradient Boosting) variable importance methodologies were used to select the set of variables that make up the developed score; each of this variables was dichotomized and a cut-off point that divides the population into two groups with different mean mortalities was found; if the patient is in the group that presents a higher mortality a one is assigned to the particular variable, otherwise a zero is assigned. These binary variables are used in a logistic regression (LR) model, and its coefficients were rounded to the nearest integer. The resulting integers are the point values that make up the score when multiplied with each binary variables and summed. The one-year mortality probability was estimated using the score as the only variable in a LR model. Predictive power of the score, was evaluated using the 1695 admissions of the validation subset obtaining an area under the receiver operating characteristic curve of 0.7528, which outperforms the results obtained with Sequential Organ Failure Assessment (SOFA), Oxford Acute Severity of Illness Score (OASIS) and Simplified Acute Physiology Score II (SAPSII) scores on the same validation subset. Observed and predicted mortality rates within estimated probabilities deciles were compared graphically and found to be similar, indicating that the risk estimate obtained with the score is close to the observed mortality, it is also observed that the number of events (deaths) is indeed increasing as the outcome go from the decile with the lowest probabilities to the decile with the highest probabilities. Sepsis is a syndrome that carries a high mortality, 43.3% for the patients included in this study; therefore, tools that help clinicians to quickly and accurately predict a worse prognosis are needed. This work demonstrates the importance of customization of mortality prediction scores since the developed score provides better performance than traditional scoring systems.

**Keywords :** intensive care, logistic regression model, mortality prediction, sepsis, severity of illness, stochastic gradient boosting

**Conference Title :** ICHCDMHI 2018 : International Conference on Healthcare Data Mining and Health Informatics

**Conference Location :** London, United Kingdom

**Conference Dates :** February 15-16, 2018