

Rapid Sepsis Detection Using Routine Laboratory Data: A Logistic Regression Approach

Annalisa Belli⁽¹⁾, Chiara Della Franca⁽²⁾, Daniela Ligi⁽²⁾, Francesca Salvatori⁽²⁾, Marisol Palomino Huaman⁽²⁾, Marc Vasse⁽³⁾, Rocchi Marco B. L.⁽¹⁾, Davide Sisti⁽¹⁾, Ferdinando Mannello⁽²⁾

 Annalisa Belli 0009-0008-8561-6088

(1) Department of Biomolecular Sciences, Section of Biostatistics, University of Urbino Carlo Bo, Urbino, Italy

(2) Department of Biomolecular Sciences, Section of Biochemistry and Biotechnology, Unit of Clinical Biochemistry, University of Urbino Carlo Bo, Urbino, Italy

(3) Clinical Biology Department, Hôpital Foch, Suresnes, France

CORRESPONDING AUTHOR: Annalisa Belli, Department of Biomolecular Sciences, Section of Biostatistics, University of Urbino Carlo Bo, Urbino, Italy. E-Mail: a.belli7@campus.uniurb.it

INTRODUCTION

Sepsis is a life-threatening condition associated with high morbidity and mortality, in which early diagnosis is crucial for improving clinical outcomes. Routine haematological parameters generated by automated analyzers have emerged as promising biomarkers for sepsis detection due to their rapid availability and low cost.

OBJECTIVES

The principal aim was to investigate the predictive value of routine hematochemical parameters for sepsis diagnosis and to construct a validated multivariable logistic regression model. Particular attention was given to variable selection, model validation, and cut-off optimization through the weighted Youden index to maximize sensitivity while maintaining high specificity.

METHODS

A total of 597 subjects admitted to the Emergency Department were included and the sample was randomly divided into training ($n = 427$; 71.5%) and validation datasets ($n = 170$; 28.5%). Dataset homogeneity was assessed through independent samples t-tests and chi-square tests with effect size estimation (Cohen's d and Cramer's V). Multicollinearity was evaluated using variance inflation factors ($VIF > 10$ threshold) and univariate logistic regression analyses were initially performed to identify candidate predictors associated with sepsis, with variable preselection based on statistical significance and

sensitivity > 0.70 . A forward stepwise multivariable logistic regression approach was subsequently applied (entry $p = 0.05$; removal $p = 0.10$). Model fit was evaluated using Nagelkerke's R^2 and Akaike Information Criterion (AIC). Discriminative ability was assessed using ROC curves, area under the curve (AUC), sensitivity, specificity, and weighted Youden index (Jw). Since minimizing false negatives represented the main clinical priority, multiple probability thresholds (0.25–0.75) were tested to optimize classification performance.

RESULTS

Most hematochemical parameters differed significantly between sepsis and non-sepsis groups. Nine variables demonstrated strong discriminatory performance in univariate analyses (sensitivity > 0.70). The final multivariable model included ten predictors and demonstrated excellent goodness-of-fit (Nagelkerke's $R^2 = 0.947$; AIC = 84.689). At the conventional threshold (0.50), sensitivity and specificity reached 97.12% and 97.26% in the training dataset, and 92.96% and 96.67% in validation. Threshold optimization identified 0.31 as the optimal cut-off (Jw = 0.9796), increasing sensitivity to 99.04% while maintaining specificity at 96.35%, substantially reducing false negatives.

CONCLUSIONS

The proposed statistical framework enabled the development of a highly discriminative and robust predictive model for sepsis detection. The prioritization of sensitivity-driven threshold optimization improved clinical applicability by minimizing missed diagnoses while preserving high specificity.