**Name of Journal:** World Journal of Clinical Cases

**Manuscript NO:** 84001

**Manuscript Type:** ORIGINAL ARTICLE

*Retrospective Study*

Risk prediction model for distinguishing Gram-positive from Gram-negative bacteremia based on age and cytokine levels: A retrospective study

Zhang W *et al.* RPM for distinguishing G+ from G- bacteremia

Wen Zhang, Tao Chen, Hua-Jun Chen, Ni Chen, Zhou-Xiong Xing, Xiao-Yun Fu
Abstract

BACKGROUND
Severe infection often results in bacteremia, which significantly increases mortality rate. Different therapeutic strategies are employed depending on whether the blood-borne infection is Gram negative (G⁻) or Gram positive (G⁺). However, there is no risk prediction model for assessing whether bacteremia patients are infected with G⁻ or G⁺ pathogens.

AIM
To establish a clinical prediction model to distinguish G⁻ from G⁺ infection.

METHODS
A total of 130 patients with positive blood culture admitted to a single intensive care unit were recruited, and Th1 and Th2 cytokine concentrations, routine blood test results, procalcitonin and C-reactive protein concentrations, liver and kidney function test results, and coagulation function were compared between G⁺ and G⁻ groups. Least absolute shrinkage and selection operator (LASSO) regression analysis was employed to optimize the selection of predictive variables by running cyclic coordinate descent and K-fold cross-validation (K = 10). The predictive variables selected by LASSO regression analysis were then included in multivariate logistic regression analysis to establish a prediction model. A nomogram was also constructed based on the prediction model. Calibration chart, receiver operating characteristic curve (ROC), and decision curve analysis (DCA) were adopted for validating the prediction model.

RESULTS
Age, plasma interleukin-6 (IL-6) concentration, and plasma aspartate aminotransferase concentration were identified from 57 measured variables as potential factors distinguishing G⁺ from G⁻ infection by LASSO regression analysis. Inclusion of these three variables in a multivariate logistic regression model identified age and IL-6 as
significant predictors. In ROC analysis, age and IL-6 yielded an area under the curve of 0.761, and distinguished G⁻ from G⁺ infection with 0.756 specificity and 0.692 sensitivity. Serum IL-6 and IL-10 Levels were upregulated by more than tenfold from baseline in the G⁻ bacteremia group but by less than tenfold in the G⁺ bacteremia group. The calibration curve of the model and Hosmer-Lemeshow test indicated good model fit ($P > 0.05$). When the DCA curve indicated a risk threshold probability between 0% and 68%, a nomogram could be applied in clinical settings.

CONCLUSION
A simple prediction model distinguishing G⁻ from G⁺ bacteremia can be constructed based on reciprocal association with age and IL-6 level.

**Key Words:** Interleukin-6; Cytokine; Bacteremia; Infection; Prediction model

Zhang W, Chen T, Chen HJ, Chen N, Xing ZX, Fu XY. Risk prediction model for distinguishing Gram-positive from Gram-negative bacteremia based on age and cytokine levels: A retrospective study. *World J Clin Cases* 2023; In press

**Core Tip:** This study was designed to assess if cytokine profile and other clinical variables can distinguish Gram positive from Gram negative bacteremia. A reliable predicted model could prove valuable for facilitating early identification of the causative pathogen and the rational use of antibiotics, thereby preventing progression into potentially fatal septic shock.

**INTRODUCTION**
Bacteremia is defined as the presence of bacteria in the blood circulation. If not effectively controlled, these bacteria and associated bacterial toxins (endotoxins and exotoxins) can induce sepsis, a condition characterized by severe tissue-damaging
inflammation with high morbidity, mortality, and treatment cost, mainly due to (septic) shock and multiple organ dysfunction syndrome (MODS)\cite{1,2}. Gram negative (G-) bacteria such as *Escherichia coli*, *Pseudomonas aeruginosa*, *Acinetobacter baumannii*, and *Klebsiella pneumoniae*, and Gram positive (G+) bacteria such as *Staphylococcus aureus*, carbapenem-resistant *Staphylococcus aureus*, and *Enterococcus faecium* are among the most common causative pathogens of septic infection\cite{3,5}. The optimal treatment differs depending on whether the causative pathogen is G+ or G-, so distinguishing between these bacterial types during the early stages of infection is essential for preventing sepsis.

The profile of Th1 and Th2 cytokine release by activated neutrophils, lymphocytes, and mononuclear phagocytes is also a critical determinant of whether bacteremia progresses to sepsis. Cytokines can be broadly divided into pro-inflammatory and anti-inflammatory, and the balance between these classes has a major influence on clinical course, including the incidence of sepsis\cite{6}. Among patients with bacterial infection, interleukin-6 (IL-6), IL-10, tumor necrosis factor (TNF)-α, and interferon (IFN)-γ are all elevated. In contrast, only IFN-γ level may be elevated significantly (> 100 pg/mL) in patients with fungal infections or tuberculosis, although IL-6 level may also increase in severe cases. Among patients with viral infection. However, IFN-γ level may increase only slightly (< 100 pg/mL), accompanied by elevations of IL-6 or IL-10\cite{7,8}.

This study was designed to assess if cytokine profile and other clinical variables can distinguish G+ from G- bacteremia. A reliable predicted model could prove valuable for facilitating early identification of the causative pathogen and the rational use of antibiotics, thereby preventing progression into potentially fatal septic shock.

**MATERIALS AND METHODS**

*Study design and conception*

The data used to construct the predictive model were obtained from a retrospective review of sepsis patients admitted to the intensive care unit (ICU) of our hospital from December 2018 to February 2020.
Study subjects
A total of 130 patients with sepsis were enrolled. Inclusion criteria were as follows: (1) Basic conditions for infection; (2) positive for bacteremia according to blood bacterial culture; (3) age ≥ 18 and ≤ 80 years; (4) initial quick Sequential Organ Failure Assessment score (qSOFA) ≥ 2 (severe infection and organ dysfunction); and (5) two or more of (a) body temperature > 38 °C or < 36 °C; (b) heart rate > 90/min; (c) respiratory frequency > 20/min or arterial partial pressure of carbon dioxide (PaCO₂) < 32 mmHg; (d) peripheral white blood cell count > 12 × 10⁹/L or < 4 × 10⁹/L; and (e) infected site confirmed by imaging examination. Exclusion criteria were: (1) age < 18 or > 80 years; and (2) infection complicated by other serious physical diseases (malignant tumors or nervous system diseases, etc.), chronic renal insufficiency, or other infectious diseases.

Data collection
Age, sex, family medical history, patient medical history, concentrations of plasma cytokines (IL-2, IL-4, IL-6, IL-10, TNF-α and IFN-γ), bacterial culture results (blood, sputum, secretions, and effusions, etc.), blood biochemistry results, C-reactive protein, procalcitonin, and fungal D-glucan concentrations, routine blood test results, and chest or other imaging data were collected.

Venous blood samples was collected using EDTA anticoagulation tubes and sent for laboratory detection of plasma Th1/Th2 cytokine levels. Briefly, the plasma sample was centrifuged at 3000 rpm for 20 min at room temperature to separate plasma. A Th1/Th2 cytokine detection kit was purchased from Cellgene Biotech (Hangzhou, China). Plasma cytokine contents were determined using a BD FACS flow cytometer.

Statistical analysis
All statistical analyses were conducted using R language software (version 4.0.5). Patients were divided into G⁻ and G⁺ groups according to blood culture results. Normally distributed continuous variables are expressed as mean ± SD, and non-
normally distributed continuous variables by median and interquartile range M (P25, P75). Continuous variables normal distributions and homogeneity of variance were compared between groups by independent sample t-test while non-normally distributed continuous variables were compared between groups by Mann-Whitney U-test. Counting data were compared by Chi-square test.

Least absolute shrinkage and selection operator (LASSO) regression analysis using the package is suitable for reducing high-dimensional data and identifying the optimal variables for further construction of multivariate prediction models. Variables with "non-zero" coefficients from the LASSO regression model were selected and ten-fold cross validation was used to define an appropriate adjustment parameter (\( \lambda \)). The most significant variables screened by LASSO regression analysis were then included in a multivariate logistic regression model. Variables without significant correlation coefficients were excluded.

Model performance for distinguishing G+ from G- bacteremia was evaluated by receiver operating characteristic (ROC) curve analysis using the qROC package for R language. The calibration curve function of the rms package was utilized to evaluate the sepsis risk nomogram and the Hosmer-Lemeshow test to determine the model goodness of fit. Finally, the decision curve analysis (DCA) package ggDCA was used to determine the clinical utility of the model according to the net benefit under different threshold profiles. Finally, a visual nomogram was drawn using the rms package.

RESULT

Clinical characteristics of the patient

A total of 130 patients with sepsis were recruited (42 females and 88 males), including 52 cases caused by G+ bacteria and 78 cases caused by G- bacteria as determined by blood culture. All patients completed relevant examinations, yielding 57 variables, and each was compared between G- and G+ groups. Age, IL-2, IL-4, IL-6, IL-10, TNF-\( \alpha \), PCT, TBIL, DBIL, AST, and PA differed significantly between groups (all \( P < 0.05 \)). Both IL-6
and IL-10 were upregulated to higher levels in the G- group than the G+ group (more than tenfold from baseline vs less than tenfold from baseline (Table 1).

**Identification of optimal predictive variables by LASSO regression**

All 57 measured variables were included in the LASSO regression model, and those with non-zero coefficients (age, AST, and IL-6) were retained for further construction of a multivariate model (Figure 1).

**Construction of a multivariate logistic regression model**

The three variables identified by LASSO regression were then included as independent variables in a multivariate logistic regression model with type of bacterial infection (G- or G+) as the dependent variable (Table 2). However, the coefficient for AST was not significant, so AST was excluded. Therefore, the final multivariate model included age and plasma IL6 concentration.

**Validation of the prediction model**

The area under the ROC curve (AUC) was 0.761, and the prediction model demonstrated a specificity of 0.756 and sensitivity of 0.692 for distinguishing G- from G+ bacteremia (Figure 2A), generally considered moderate performance. The calibration curve of the model is shown in Figure 2B. Here, the diagonal dash line represents perfect prediction and the solid line represents the performance of the actual model, so the difference (distance) is indicative of prediction performance. The Hosmer-Lemeshow test yielded a P > 0.05, indicating good model fit.

**Decision curve analysis**

The DCA diagram indicated that when the risk threshold probability was between 0% and 68% (Figure 3), the model yielded higher accuracy for distinguishing G- from G+ bacteremia.
Nomogram for predicting septic infection type

Finally, a nomogram was constructed for individual clinical use (Figure 4). A typical case is used to explain nomogram performance. If the critically ill patient was 60 years old and plasma IL-6 Level was 4500 pg/mL, the nomogram yields a score of approximately 25 points (15 points plus 10 points), which is closer to the side of G- bacteremia and thus indicates that G- bacteremia is more likely.

DISCUSSION

Sepsis patients deteriorate rapidly, so timely identification of the causative pathogen and treatment initiation, including rational antibiotic therapy is essential[9,10]. The progression of bacterial infection into sepsis is strongly associated with the host cytokine profile. Multiple stressors, including infection, trauma and surgery can disrupt the Th1-Th2 balance[11,12]. In patients with sepsis, macrophages release large amounts of IL-1β, TNF-α, and prostaglandin E2 (PGE2) in response to stimulation by bacterial endotoxin or signals from damaged tissue. While IL-1β and TNF-α are generally proinflammatory, PGE2 inhibits IFN-γ release in response to endotoxin. Concomitantly, the hypothalamus-pituitary-adrenal axis can secrete a large quantity of glucocorticoids under stress, which in turn induces IkB synthesis, inhibits NF-kB signaling and the release of inflammatory cytokines, enhances the differentiation of Th2 cells, and promotes the release of multiple cytokines, such as IL-4, IL-10 and IL-6, that drive the transformation of Th1 cells into Th2 cells[13-15]. Consistent with Tang et al[7], cytokine levels were upregulated in both G- and G+ bacteremia patients. However, the increases in IL-6 and IL-10 relative to baseline were significantly greater in the G- group (P < 0.05), providing a metric for distinguishing G- from G+ bacteremia.

Among the 130 patients, age, IL-2, IL-4, IL-6, IL-10, TNF-α, PCT, TBIL, DBIL, AST and PA differed significantly between those infected with G+ bacteria and those infected with G- bacteria. These 11 variables were used to construct and validate a novel risk prediction model for G+ versus G- blood-borne infection. This multivariate analysis indicated that age and IL-6 were the key distinguishing variables. Introducing other
demographic and clinical parameters from biochemical and physical examinations into the risk nomogram may further enhance discrimination accuracy and allow the nomogram to be applied for larger clinical samples\textsuperscript{[16,17]}. However, the sample size was small, so important predictive variable may have been excluded.

The multivariate model indicated that age and plasma IL-6 alone were interchangeable predictors of G\textsuperscript{+} versus G\textsuperscript{-} bacteremia, while the regression coefficient for AST was not significant. Further, the Hosmer-Lemeshow test yielded a $P > 0.05$ for age and plasma, indicating goodness of fit. Specifically, older age was associated with higher serum IL-6 and greater incidence of G\textsuperscript{-} bacteremia. Nomograms are reliable and practical statistical algorithms that can rapidly predict the probability of clinical events by integrating multiple variables\textsuperscript{[18]} and risk factors\textsuperscript{[19]}. Further, these may be digitized for inclusion of additional factors for more accurate prediction and better clinical decision-making. A visual nomogram constructed in the R language considering these two variables predicted increased risk of G\textsuperscript{-} bacterial infection (e.g., if the IL-6 level is 4500 pg/mL in a 60-year-old patient, total score is close to the G\textsuperscript{-} bacteremia side on the nomogram). Thus, the older the patient and the higher the plasma IL-6 level, the greater the risk of G\textsuperscript{-} bacterial infection, while a younger patient with lower IL-6 is more likely to have a G\textsuperscript{+} infection. A reliable prediction model to distinguish G\textsuperscript{-} from G\textsuperscript{+} bacteremia based on age and plasma IL-6 could facilitate more rapid diagnosis and initiation of rational treatment, possibly because elderly patients suffer from multiple primary diseases and low immune function, conditions that increase susceptibility to G\textsuperscript{-} bacterial infection over G\textsuperscript{+} bacterial infection. Invasion by G\textsuperscript{-} bacteria evokes a cytokine imbalance after cytokine storm, including elevated IL-6. Other cytokines are also elevated but may not have reached significance for model inclusion due to the small sample size.

Sepsis is a progressive disease, so it is essential to control infection and prevent deterioration through rational use of antibiotics and control of cytokine storm. For this, accurate risk prediction tools are necessary to identify the pathogenic species at an early stage and implement the most effective therapeutic measures. In the present study, we
developed such a prediction tool to assist clinicians in prompt identification of bacteremia type.

CONCLUSION
In the present study, age and IL-6 concentration were found to distinguish G+ from G- bacteremia with high accuracy. Specifically, the probability of G- bacteremia is associated with older age and higher plasma IL-6 than G+ bacteremia. Plasma IL-6 and IL-10 Levels were increased tenfold from baseline in the G- group but by less than tenfold in the G+ group. While this prediction model may have excluded other significant distinguishing factors due to small sample size, it may still facilitate faster and more effective treatment.
<table>
<thead>
<tr>
<th>Primary Source</th>
<th>Number of Words</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>bmcinfectdis.biomedcentral.com</td>
<td>41</td>
<td>2%</td>
</tr>
<tr>
<td><a href="http://www.researchgate.net">www.researchgate.net</a></td>
<td>21</td>
<td>1%</td>
</tr>
<tr>
<td>academic.oup.com</td>
<td>14</td>
<td>1%</td>
</tr>
<tr>
<td>bmjopen.bmj.com</td>
<td>12</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td><a href="http://www.pathologyoutlines.com">www.pathologyoutlines.com</a></td>
<td>11</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>Ruohui Mo, Rong Shi, Yuhong Hu, Fan Hu.</td>
<td>10</td>
<td>&lt; 1%</td>
</tr>
</tbody>
</table>


Crossref

Exclude Quotes: On
Exclude Sources: < 10 words
Exclude Bibliography: On
Exclude Matches: < 10 words