

# Bedside index for severity in acute pancreatitis (bisap)



**ASSIGN  
BUSTER**

## Abstract

*Objectives* The aim of this study was to evaluate the diagnostic performance of Bedside Index for Severity in Acute Pancreatitis (BISAP) score for predicting severe acute pancreatitis (SAP) in the early phase.

*Method* The PubMed, Cochrane library and EMBASE databases were searched until May 2014. The strict selection criteria and exclusion criteria were determined, and we applied hierarchic summary receiver operating characteristic (HSROC) model and bivariate random effects models to assess the diagnosability of the BISAP score for predicting SAP. We obtained pooled summary statistics for sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR) and calculated the area under the HSROC curve (AUC). The 95% confidence intervals (CI) for each diagnostic test measure were also calculated. Publication bias was assessed using Deek's funnel plot asymmetry test. Statistical analyses were performed using the STATA 12.0 software.

*Results* The pooled sensitivity, specificity, PLR, NLR, and DOR were 64.82%, 83.62%, 3.96, 0.42 and 9.41, respectively. The AUC was 0.77 and the HSROC curve for individual studies was generated and analyzed to explore the influence of threshold effects.

*Conclusion* We confirmed that BISAP score is an accurate means to predict SAP in the early phase.

**Keywords:** BISAP, HSROC curve, severe acute pancreatitis, acute pancreatitis

## 1. Introduction

<https://assignbuster.com/bedside-index-for-severity-in-acute-pancreatitis-bisap/>

Acute pancreatitis (AP) is an inflammatory condition of the pancreas with a clinical course that varies from mild to severe and characterized by activation of pancreatic enzymes to cause self-digestion of the pancreas <sup>1</sup>. Generally, AP is mild, self-limiting, and requires no special treatment and ranges about 80-90% of patients with only minimal or transitional systemic manifestations, but about 20%-30% of patients develop a severe disease that can progress to systemic inflammation and cause pancreatic necrosis, multi-organ failure, and potentially death <sup>1-4</sup>. So it is important to have an early, quick, and accurate risk stratification of AP patients, which would permit evidence-based early initiation of intensive care therapy for patients with severe AP (SAP) to prevent adverse outcomes and allow treatment of mild AP (MAP) on the common ward. Early identification of patients with SAP would allow the clinician to consider more aggressive interventions within a time frame that could prevent possible complications.

Currently, there are a variety of score systems developed for the early detection of SAP, such as Ranson's score <sup>5</sup>, acute physiology and chronic health examination (APACHE) II <sup>6, 7</sup> and computed tomography severity index (CTSI) <sup>8</sup>. Also there are many inflammation markers such as C-reactive protein (CRP), interleukin-6 (IL-6) and others <sup>9, 10</sup>. Several studies show that cytokines play an important role in the cascading inflammatory responses <sup>11</sup> and it may act as mediators of distant organ complications in SAP. So the levels of cytokine in serum may also reflect the degree of the inflammatory response <sup>12</sup>. In 2008, Wu *et al.* <sup>13</sup> proposed a new prognostic scoring system, the bedside index of severity in acute pancreatitis (BISAP), is a

simple and accurate method that can predict the clinical severity of AP within 24 h of presentation. BISAP incorporates five parameters: blood urea nitrogen > 25 mg/dL, impaired mental status, systemic inflammatory response syndrome (SIRS), age > 60 years, and detection of pleural effusion by imaging <sup>14</sup>.

Unfortunately there has been no systematic or meta-analytic review of cross-sectional studies of this scoring system. The purpose of this study was to aggregate the reported data across the different studies and to assess the ability of the BISAP score to predict SAP.

## 2. Materials and methods

### *2.1 Literature search*

The search was performed on three databases: PubMed, Cochrane library and EMBASE. These databases were searched from the first date available in each database up to May 2014, using the search terms 'acute pancreatitis' AND ('BISAP' OR 'bedside index of severity in acute pancreatitis'). Once articles had been collected, bibliographies were then hand-searched for additional references.

### *2.2 Inclusion and Exclusion Criteria*

To be included in this meta-analysis, studies must meet the following criteria: (1) studies evaluate the BISAP score for predicting SAP; (2) the subjects were diagnosed with AP; (3) prospective study; (4) the absolute numbers of true positive (TP), false negative (FN), false positive (FP), and

true negative (TN) test results were available or derivable from the article; (5) the clinical result of patients was indicated as SAP.

Studies were excluded if one of the following existed: (1) the numbers of TP, FN, FP, and TN test results were not derivable from the article; (2) cross-sectional study; (3) non-original articles, such as review, meeting abstract, case report and comment; (4) duplicate of previous publications and data description is not clear.

### *2.3 Data extraction and quality assessment*

All data were extracted independently by two authors according to the inclusion criteria listed above. Disagreements were resolved by discussion or solved by consultation of a third reviewer. The following characteristics were collected from each study: the first author, year of publication, source, experiment design, sample size, the reference standard (gold standard), the numbers of TP, FN, FP, and TN and others. The QADAS (Quality Assessment of Diagnostic Accuracy Studies) criteria were used to assess the quality of diagnostic accuracy studies included in this meta-analysis<sup>15</sup>.

### *Statistical analyses*

Hierarchical summary receiver operating characteristic (HSROC) model and bivariate random effects model were performed in STATA 12.0 (StataCorp, College Station, TX, USA) software using the program 'metandi' to generate pooled accuracy estimates of sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR) and calculated the area under the HSROC curve (AUC)<sup>16</sup>. The HSROC curve for <https://assignbuster.com/bedside-index-for-severity-in-acute-pancreatitis-bisap/>

individual studies was generated and analyzed to explore the influence of threshold effects. The 95% confidence intervals (CI) for each diagnostic test measure were also calculated. Publication bias was assessed using Deek's funnel plot asymmetry test <sup>17</sup> .

### 3. Results

#### 3. 1 Eligible Studies

The process of selecting studies for the meta-analysis was shown in Fig. 1. There were 32 studies potentially eligible studies identified. Of these, 14 studies were excluded after screening based on abstracts or titles to avoid obvious irrelevance. Finally, 9 studies <sup>14 , 18-25</sup> met the inclusion criteria and were included in the meta-analysis. The data collected from the related studies was summarized in Table 1. Among these studies, kim *et al.* <sup>20</sup> reported the results of the meta-analysis with the cutoff values set at 2 and 3, respectively. All patients were recruited within 24 h from the time of admission or transfer and used for the calculation of the BISAP scores. All included citations were prospective cohort studies. The absolute numbers of TP, FN, FP, and TN were calculated by sample size and the degree of sensitivity and specificity.

A summary of the quality of the studies was displayed in table 2. The included studies were not described the tenth quality indicator (were the index test results interpreted without knowledge of the results of the reference?) and the eleventh quality indicator (were the reference standard results interpreted without knowledge of the results of the index test?) <sup>15</sup> . At the

same time, there are some studies not described in detail for eliminate and exit objects.

### *3. 2 Meta-analysis*

The results of the HSROC model were show in Table 3. The pooled sensitivity of BISAP testing for the diagnosis of SAP was 64. 82% (95% CI: 54. 47%-73. 74%), and the specificity was 83. 62% (95%CI: 70. 03%-91. 77%). The pooled DOR was 9. 41 (95%CI: 5. 38-16. 45), the PLR was 3. 96 (95%CI: 2. 27-6. 89), and the NLR was 0. 42 (95%CI: 0. 34-0. 52). The AUC of the HSROC was 0. 77 (95%CI: 0. 73-0. 80) (Fig. 2). The  $I^2$  index of heterogeneity was 95% (95% CI, 91%-99%).

### *3. 3 Subgroup Analyses*

There was a negative correlation between the logits of sensitivity and specificity (Spearman correlation coefficient, 20. 09), indicating the present of an important effect of the diagnostic threshold (cutoff level) on the performance of BISAP score. The following cutoffs were selected for subgroups analysis (Table 4).

Analysis of studies that set the BISAP cutoff point at 2, the pooled sensitivity, specificity, PLR, NLR, and DOR were 67. 30% (95%CI: 60. 53%-73. 42%), 78. 28% (95%CI: 68. 86%-85. 46%), 3. 10 (95%CI: 2. 12-4. 52), 0. 42 (95%CI: 0. 34-0. 51) and 7. 42 (95%CI: 4. 39-12. 54), respectively. The AUC of the HSROC was 0. 70(95%CI: 0. 66-0. 74).

Analysis of studies that set the BISAP cutoff point at 3, the pooled sensitivity, specificity, PLR, NLR, and DOR were 61.18% (95%CI: 41.20%-78.00%), 88.64% (95%CI: 88%-97.18%), 5.39 (95%CI: 1.80-16.12), 0.44 (95%CI: 0.30-0.64), and (95%CI: 4.44-34.03), respectively. The AUC of the HSROC was 0.78 (95%CI: 0.75-0.82).

### *3.4 Publication Bias*

Deeks' Funnel Plot Asymmetry Test for the overall analysis showed that no significant publication bias was found ( $P = 0.359$ , Fig. 3).

## 4. Discussion

Generally, Ranson, APACHE II, and CTSI scoring systems have been used to evaluate the severity of AP<sup>22, 23</sup>. However, these techniques all have their inherent strengths and weaknesses. For example, the Ranson's score<sup>5</sup> is relatively accurate at classifying the severity of AP, but the evaluation cannot be completed until 48 h, which will miss the potential for early treatment and increase mortality. The APACHE II system<sup>6, 7</sup> allows the determination of disease on the first day of admission and is more accurate than Ranson's score, but complexity is its major drawback. CTSI<sup>26, 27</sup> is calculated based on CT findings of some local complications and cannot reflect the systemic inflammatory response. Recently, the BISAP score has been proposed as an accurate method for early identification of patients at risk for in hospital mortality<sup>13</sup>. Several studies showed that BISAP score is a reliable and accurate means for predicting the severity of AP in the early phase<sup>18, 22, 23</sup>. But these studies are not systematic, so we collect the



reported data across the different studies and apply HSROC model and bivariate random effects model to assess the ability of the BISAP score to predict SAP. The pooled sensitivity, specificity, PLR, NLR, and DOR were 64.82%, 83.62%, 3.96, 0.42 and 9.41, respectively. The AUC of the HSROC was 0.77. Our meta-analysis indicated that BISAP score is a reliable and accurate means to predict SAP.

This meta-analysis assessed the diagnostic performance of BISAP in 1972 individuals from 9 research studies<sup>14, 18-25</sup>. The results show that the performance of BISAP to predict the severity of AP has a good specificity, but moderate sensitivity in predicting SAP. In addition, compared with other scoring systems in predicting SAP, BISAP has a higher specificity but a lower sensitivity<sup>21-23, 28</sup>. The low sensitivity may be caused by these factors. First, the characteristics of study participants are differences (cultural and geographical differences), such as lifestyle, race, and genetic basis. Second, etiologic distribution may also explain the noted differences. Third, the different definitions of SAP may also be a reason for these variations.

The HSROC curve presents a global summary of test performance and shows the tradeoff between sensitivity and specificity. The summary DOR and the AUC of the HSROC were 9.41 and 0.77, respectively. The predictive accuracy of BISAP scoring system was measured by AUC. An AUC of 1.0 represents a perfect test, whereas an AUC of 0.5 represents a test that performs no better than chance<sup>29</sup>. The result revealed that the discrimination of disease severity was good in our study, which is similar to other reports. DOR is a single indicator of test accuracy that combines the

sensitivity and specificity data into a single number. The DOR of a test is the ratio of the odds of positive test results in the patient with disease relative to the odds of positive test results in the patient without disease. The value of a DOR ranges from 0 to infinity, with higher values indicating better discriminatory test performance (higher accuracy). A DOR of 1.0 indicates that a test does not discriminate between patients with the disorder and those without it<sup>30</sup>. In the present meta-analysis, we found that the pooled DOR was 9.41, also indicated a high level of overall accuracy.

Since the HSROC curve and the DOR are not easy to interpret or use in clinical practice, and likelihood ratios are considered to be more clinically meaningful, we also presented both PLR and NLR as our measures of diagnostic accuracy. Likelihood ratios of  $> 10$  or  $< 0.1$  generate large and often conclusive shifts from pre-test to posttest probability (indicating high accuracy)<sup>31</sup>. The PLR and NLR value were 3.96 and 0.42, respectively. This result performed similar to traditional scoring systems in predicting SAP and suggested that the accuracy of still need to improve. But BISAP is relatively simple and had greater accuracy than other scoring systems, making it a promising method of predicting SAP<sup>14, 19, 21, 28</sup>. Furthermore, it may be combined in medical decision-making at the extreme of the prediction range, such as enrollment criteria for clinical trials, and as triaging intensive care unit admission<sup>32, 33</sup>.

We also explored systematically the issue of heterogeneity by use of subgroup analysis. In our analysis, the diagnostic threshold presented an important effect on the performance of BISAP score. The results

demonstrated that a BISAP score of 3 had greater accuracy and high predictive value than a score of 2 for predicting SAP.

Our meta-analysis had several limitations. First, when the BISAP scoring system converts continuous variables into binary values of equal weight, it fails to capture synergistic or multiplicative effects based on the interactions of interdependent systems<sup>21</sup>. Future research could focus on comprehensive reassessment of the pathologic mechanisms of AP with attention to the effects of preexisting risk factors (e. g. age, obesity, genetic) and well-defined end points, identification of accurate biomarkers to assess activity on these pathways, and mathematical models that have strong predictive accuracy.

Second, the exclusion of conference abstracts, letters to the editor, and non-English-language studies might have led to publication bias, which was not found in the present review. However, a review of these abstracts and letters suggested that the overall results were similar to the results in the English language studies included. Third, there is a risk for publication bias in which positive results or results with ‘expected’ findings are more likely to be published. We made every possible effort to minimize this type of bias by contacting investigators in the field of BISAP. If editors were more likely to publish manuscripts showing the ‘expected’ results of a good diagnostic performance for BISAP, then our results may be overestimating the real diagnostic performance of BISAP.

In conclusion, we confirmed that BISAP score is an accurate means to predict SAP in the early phase. Due to simplicity and easily obtained parameters,

<https://assignbuster.com/bedside-index-for-severity-in-acute-pancreatitis-bisap/>

BISAP score should gain broad acceptance in routine use not by replacing clinical assessment, but rather by complementing and objectifying it.