Incorporation of C-reactive Protein Measurement into Alvarado Score Improves Model Performance


1. Department of Emergency Medicine, Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Taoyuan, Taiwan.
2. Department of Family Medicine, Chang Gung Memorial Hospital, Taoyuan, Taiwan.
3. Graduate Institute of Clinical Medical Sciences, College of Medicine, Chang Gung University, Taoyuan, Taiwan.
4. Department of Emergency Medicine, Chang Gung Memorial Hospital, Keelung, Taiwan and Chang Gung University College of Medicine, Taoyuan, Taiwan.
5. Department of Rehabilitation and Physical Medicine, Taipei Veterans General Hospital, Taipei, Taiwan.
6. Department of Rehabilitation and Physical Medicine, National Yang-Ming University, Taipei, Taiwan.
7. Department of Emergency Medicine, National Taiwan University Hospital Yunlin Branch, Douliou, Taiwan.
8. Department of Emergency Medicine, National Taiwan University Hospital, Taipei, Taiwan.

*Corresponding author: Dr. Chien-Chang Lee, Department of Emergency Medicine, National Taiwan University Hospital, No 7, Chung Shan South Road, Taipei 100, Taiwan. E-mail: cclee100@gmail.com
Dr. Shy-Shin Chang, Department of Family and Emergency Medicine, Chang Gung Memorial Hospital, No.5, Fusing St. Gueishan Township, Taoyuan County 333, Taiwan. E-mail: sschang0529@gmail.com

Abstract

Objectives

The aim of this study was to assess whether incorporation of C-reactive protein (CRP) measurement into Alvarado Score may improve the appendicitis diagnosis performance.

Methods

A prospective observational study was carried out in the emergency department (ED) of a university hospital between July 2006 and June 2007. Adult patients presented to the ED with a provisional diagnosis of appendicitis were enrolled. Each patient underwent CRP and Alvarado score evaluation on admission. We added the CRP variable to the original Alvarado score and compared c-statistics, model fit, and calibration of the modified and original score. We also performed classification and reclassification tree (CART) analysis to see if addition of CRP variable would improve discrimination.

Results

Of the 299 study patients, 198 (66.2%) had confirmed diagnosis of appendicitis. Controlling for the Alvarado score and CRP elevation > 50 mg/L, was independently associated with acute appendicitis, justifying the addition of the CRP variable to the Alvarado score. Based on the beta-coefficient in the regression model, the CRP variable had a weight of 5 points. The area under the receiver operating characteristic (ROC) curve was 0.73 for original Alvarado score and 0.80 for CRP-modified Alvarado score. At an optimal cutoff point determined by ROC analysis (6 points), the Alvarado score had a sensitivity of 72.7% and a specificity of 63.4%. At a cutoff point of 9 points, the modified score had a sensitivity of 94.1% and a specificity of 46.5%. CART analysis created 2 decision trees for diagnosis of appendicitis. The CRP-modified decision tree had an area under curve (AUC) of 0.817, higher than that of the Alvarado score decision tree (AUC: 0.806).

Conclusion

Incorporation of CRP into the Alvarado score can improve diagnostic performance of the Alvarado score. Future improvement of appendicitis diagnostic score should take into consideration the information of CRP. Limited by the small sample size, the results of this study require further external validation.

Cite this article: Chien C. Incorporation of C-reactive protein Measurement into Alvarado Score Improves Model Performance. J J Emergen Med. 1(1): 002.
Introduction

With a lifetime risk around 7%, acute appendicitis remains one of the most common surgical conditions that requires prompt diagnosis [1]. Despite recent advances in diagnostic medicine, negative appendectomy rate remains as high as 8%–15% [2-5]. Diagnosis of appendicitis and determination of the severity of appendicitis on clinical grounds, still remains a routine challenge for first-line clinicians [6,7]. The Alvarado scoring system, developed in 1984, incorporates 8 readily available types of information in the emergency department (ED) to aid clinicians to quantitatively assess the likelihood of appendicitis on the basis of clinical findings alone [8]. The reported sensitivity and specificity of Alvarado score has been variable, ranging from 60% to 80%, determined by the prevalence of appendicitis in the studied population [5,9-14]. Because it can be rapidly, repeatedly, and easily evaluated, once the patient’s condition changes, it is gradually accepted as an important diagnostic aid. Although the Alvarado score provides quantitative diagnostic information, several validation studies showed that it had insufficient positive predictive value (PPV) and cannot be used as a sole tool for determination of the need for surgery [3-5,12,15,16].

To enhance the PPV of the original Alvarado score, we hypothesized the inclusion of information of inflammatory biomarkers such as C-reactive protein (CRP) that may achieve this goal. CRP has not been proved a sensitive marker in diagnosing early appendicitis, but it is a highly specific marker for patients with abdominal symptoms for more than 24 hours or advanced appendicitis [17-19]. It has been reported to be 33% to 95% specific for appendicitis in patients with acute abdominal pain and some suggest that CRP may be more sensitive (83% to 90%) in detecting appendiceal perforation [17-19]. In this study, we seek to evaluate the contribution of CRP to the Alvarado score by comparison of the area under the ROC curve between original and modified logistic regression models. Besides, we further evaluated the additional value of CRP by comparison of the performance parameters of classification and regression tree (CART) analysis.

Methods

Patients and Clinical Data Collection

We performed a secondary analysis of data collected from a prospectively enrolled cohort of patients with suspected appendicitis between July 2007 and Dec 2008, in the ED of a university hospital. The study protocol was approved by the Institutional Review Board. The study included all patients aged 12 years and older who had a provisional diagnosis of acute appendicitis in the ED. Exclusion criteria were age less than 12 years, referred patients with unclear initial conditions, and discharge before the confirmation test for appendicitis. For those who did not undergo surgery, the diagnosis of appendicitis was confirmed by surgical histopathology or computed tomography. Blood samples were obtained for routine tests (hemogram, CRP, and, if necessary, blood culture). We recorded the following data for each patient: sex, age, subjective symptoms, abdominal physical findings, admission vital signs, hemogram, and CRP. CRP was measured using a nephelometric assay (Dade-Behring, SA, Paris, France) with a detection limit of 0.2 mg/L. Those performing case ascertainment were blinded to CRP results. Alvarado score was calculated for each patient. Image findings and final diagnoses were extracted from medical charts with structured data recording forms.

Statistical Methods

Logistic regression was performed considering the dichotomized predictive parameters as independent values and appendicitis as the outcome value. Parameters significantly associated with appendicitis were entered into a multiple regression model with Alvarado score forced into the model as a continuous variable. For ease of use in clinical setting, we then created a modified Alvarado score using the original Alvarado score variables plus CRP. The beta coefficients from the logistic regression model were used to determine the weight of the CRP variable. For the validation of the score, we performed internal model validation by using bootstrapping. Unlike the data-splitting approach, this approach permits the use of the entire dataset for model development and is now widely regarded as a better approach for validation than data splitting [20-22]. We performed bootstrapping 200 times, and the optimism-corrected C statistics were then reported. We compared the optimism-corrected C statistics between original and modified Alvarado score model. In addition, we also compared model fit by Bayesian information criteria (BIC) value. BIC are a likelihood-based statistics that penalize increased covariate number and can be used to help select a best-fit and parsimonious model among different extended models. Lastly, we performed Hosmer-Lemeshow test in the 2 models to see if there was significant departure between predicted and observed count in different appendicitis risk categories. We further performed CART analysis [23]. CART produces a decision tree by binary recursive partitioning method. We built 2 decision trees based on the original Alvarado score variables and additional CRP variable. The AUC of the resulting tree analysis was evaluated. All tests were 2-tailed, and P values <0.05 were considered statistically significant. Logistic regression analyses were performed using STATA ver. 11 (STATA Inc., College Station, TX) and CART analysis using “rpart” package in R.
Results

Patient Characteristics

During the study period, 369 patients had a provisional diagnosis of appendicitis in the ED. Thirty-eight patients were excluded because of lack of confirmed surgical pathology or CT. Another 32 patients were further excluded because of a confirmed infectious etiology such as urinary tract infection, pelvic inflammatory disease, or tubo-ovarian abscess. Finally, 299 patients were included for analysis. The mean age of the study sample was 35.6 ± 17.7 years and 155 patients (51.8%) were men. One hundred and ninety-eight patients (66.2%) had a confirmed diagnosis of acute appendicitis, of which 94 (47.5%) had perforated appendicitis.

Comparison of Model Performance, Discrimination, Model Fit and Calibration

On univariate analyses, most of the Alvarado score covariates were statistically significant predictors of acute appendicitis, except for anorexia, nausea/vomiting, or McBurney’s point tenderness. Serum levels of CRP were significantly higher in acute appendicitis than those in non-appendicitis (77.6 mg/L vs. 16.9 mg/L, P < 0.001). To determine the optimal cutoff for CRP, we performed an ROC analysis and found that a cutoff point of 50 mg/L could maximize sensitivity and specificity. On multivariate analysis, CRP elevation was independently associated with acute appendicitis after adjustment for Alvarado score. Results of univariate analysis are summarized in table 1.

Table 1. Characteristics of 299 patients and univariate correlation with the diagnosis of appendicitis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (N = 299)</th>
<th>Acute Appendicitis (N = 198)</th>
<th>Non-Appendicitis (N = 101)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td>35.6 ± 17.7</td>
<td>36.6 ± 19.0</td>
<td>34.8 ± 15.5</td>
<td>0.186</td>
</tr>
<tr>
<td>Sex, M</td>
<td>155 (51.8%)</td>
<td>124 (62.3%)</td>
<td>31 (29.8%)</td>
<td>0.001</td>
</tr>
<tr>
<td>Tenderness at McBurney’s point</td>
<td>232 (77.6%)</td>
<td>232 (77.9%)</td>
<td>0 (0%)</td>
<td>1.000</td>
</tr>
<tr>
<td>Initial white blood cell count, WBC</td>
<td>13,151.0 ± 6,631.5</td>
<td>13,096.5 ± 6,752.3</td>
<td>13,597.0 ± 5,450.9</td>
<td>0.831</td>
</tr>
<tr>
<td>CRP, mg/L</td>
<td>154 (51.8%)</td>
<td>154 (51.8%)</td>
<td>100 (99.0%)</td>
<td>0.018</td>
</tr>
</tbody>
</table>

On multivariate analysis, CRP > 50 mg/L is strongly and independently associated with acute appendicitis in the logistic regression model controlling for Alvarado score. The beta coefficients and standard errors for 2 different models are provided in Table 2. Based on the beta coefficient, CRP variable in the modified Alvarado score weighed 5 points. The modified Alvarado score was 10 points from the original 9 variables plus 5 points from CRP > 50 mg/L. We performed internal validation with 200 times bootstrapped samples and derived the optimism-corrected C statistics (area under the ROC curve). The modified score had better optimism-corrected C statistics than the original score (C statistics 0.80 vs. 0.73). The ROC curves of the 2 different models are shown in fig 1.

Table 2 also presents the comparisons of overall fit and calibration for discrimination of the 2 different models. For global model fit, the modified model had a lower value of BIC than original one, which means the new model was superior in terms of model fit and model simplicity. Finally, the Hosmer-Lemeshow tests were insignificant in the 2 models, which means there was no significant difference between predicted and observed count in different risk categories for the 2 models.

Table 3 summarizes the comparison of sensitivity, specificity, likelihood ratio positive, likelihood ratio negative, PPV, and negative predictive value between Alvarado and CRP-modified Alvarado score. The best cutoff values were determined by ROC analysis. Compared to the original score, CRP-modified score greatly enhanced the specificity, PPV, and likelihood ratio positive; although the sensitivity, negative predictive value, and likelihood ratio negative were slightly decreased.

CART Analysis

The CART procedure was carried out to determine the value of Alvarado score variables and CRP levels in the decision process of appendicitis. In the first CART analysis, we put in all Alvarado score variables for analysis.
The CART analysis created a decision tree that selected left-shift leukocytosis, presence of fever, rebound tenderness, anorexia, and pain relocation as significant decision nodes. These decision nodes identified 6 high probability groups and 3 low probability groups (Fig 2). The incidences of appendicitis in the high probability groups were 91%, 87.5%, 83.3%, 68.8%, 67.3%, and 61.5%. The incidences of appendicitis in the low probability groups were 16.7%, 25%, and 36.7%. The Alvarado score decision tree had an AUC of 0.806. In the second CART analysis, we incorporated the CRP information. CART created a decision tree that selected CRP greater than 50 mg/L, presence of fever, anorexia, pain relocation, and leukocytosis as significant decision nodes. These decision nodes identified 4 high probability groups and 2 low probability groups (Fig 3), greatly simplifying the algorithm. The incidences of appendicitis in the high probability groups were 91.8%, 78.6%, 72.2%, and 58.3%, and the incidences of appendicitis in the low probability groups were 23.4% and 23.1%. The CRP-modified decision tree had an AUC of 0.817, higher than that of the Alvarado score decision tree (AUC: 0.806).

Discussion

The results of this investigation have shown that the incorporation of the CRP information may improve the diagnostic performance of the Alvarado score for appendicitis. At the best cutoff point determined by ROC analysis, the CRP-modified score had greatly enhanced the PPV than the original score, at the cost of a slight decrease in the negative predictive value. To reduce the negative appendectomy rate, PPV may be a more important parameter than negative predictive value. In addition to Alvarado score modification, we also performed CART analysis to establish a clinical decision algorithm based on CRP and Alvarado score variables. Compared to the CART analysis with Alvarado score variables alone, we showed that incorporation of CRP information may lead to a simplified decision tree with improved diagnostic performance.

Several clinical scoring systems have been proposed with sensitivity ranging from 48% to 77% and specificity from 73% to 87% [10]. Surgical decision based on Alvarado score alone has been reported to result in negative appendectomy in 12% to 33% cases [3, 9-11]. It is obvious that the score alone is not sufficient to make a surgical decision. Recent efforts have been made to design an algorithm that uses Alvarado score as an initial screening tool followed by diagnostic imaging in doubtful cases [13,24]. This approach has greatly enhanced the diagnostic accuracy, but patients with high Alvarado score may not undergo diagnostic imaging examination. The relatively high false positive rate for patients with high Alvarado scores is still a problem. Our approach may further refine this approach by reducing the false positive rate in the initial score-screening stage and possibly reduce the need of imaging examination in many patients. This may be very important in today’s ED environment, as ED crowding is severe and patients may have a long waiting time for CT or ultrasound. Several studies have shown that delayed surgery is one of the most important factors for appendiceal perforation.

Another unique feature in our work is the application of CART analysis to facilitate appendicitis diagnosis. CART analysis is based on the nonparametric recursive partitioning method and can build a decision tree structure to classify subjects into high- and low-risk groups. Unlike the conventional logistic regression method, CART analysis does not assume a parametric probability risk function form and is not affected by outlying observations. In this study, CART analysis was used to create a decision tree that can assist clinicians in making an appendectomy decision. Several clinical decision tools have been proposed to aid the diagnosis of acute appendicitis, such as Naïve Bayes method, neural network, and expert system [10,25]. Unlike CART analysis that uses a simple and transparent reasoning method, the decision process of the aforementioned methods is a black box for most clinicians who know little about statistics and decision theory. The Alvarado score or CART decision tree provided in this work is readily understood by most health professionals and can be readily applied to clinical setting without additional software or computational resources.

Results of this study should be interpreted with its several limitations. First, this a single center studies with limited sample size, and the reliability or external validity of our modified score or algorithm awaits further larger sample validation. Second, as CRP is a late marker for appendicitis, the accuracy of the results of this study may be subjected to the different EDs with different timings and severity of appendicitis in the patients. The proposed modified score or decision algorithm may have better performance in a referral center, where patients are in a more severe state and present at a later stage of the disease. Third, McBurney’s point tenderness was not a significant finding in our study, because our study population included patients with a
Conclusions

Our work showed that inclusion of CRP information may potentially improve the clinical decision rules for acute appendicitis. With significantly better specificity and lower false-positive rates than the original Alvarado score, CRP-modified score is potentially useful in confirming a diagnosis of acute appendicitis, but not in ruling it out. Whether the use of CRP-modified clinical decision rule can have a positive impact on patient outcomes remains to be answered by further research.

1. What is current knowledge?
   • Alvarado score can had insufficient positive predictive value (PPV) and cannot be used as a sole tool for determination of the need for surgery.

2. What is new here?
   • CRP-modified Alvarado score can significantly enhance PPV and may be useful in confirming a diagnosis of acute appendicitis.
   • The low negative predictive value of the CRP-modified Alvarado score makes it not a suitable rule-out diagnostic tool.

References


Cite this article: Chien C. Incorporation of C-reactive protein Measurement into Alvarado Score Improves Model Performance. J J Emergen Med. 1(1): 002.


