ORIGINAL ARTICLE

Evaluation of Diagnostic Scores for Acute Appendicitis

Wu Xingye, Li Yuqiang, Wang Rong and Zhang Hongyu

ABSTRACT

Objective: To assess the sensitivity of the Alvarado score (AS), modified Alvarado score (MAS), Fenyo-Lindberg score (FS), Lintula score (LS), Eskelinen score (ES), Teicher score (TS), and Christian score (CS) [seven scorings] in patients with acute appendicitis (AA).

Study Design: Analytical study.

Place and Duration of Study: The First Affiliated Hospital of Chongqing Medical University, China, from January 2012 to June 2015.

Methodology: Patients with diagnosis of AA were evaluated retrospectively to compare the scoring systems. The diagnostic sensitivity (the correct number of diagnoses divided by the total number) was compared. Data were analyzed using SPSS software.

Results: One hundred and seventy-nine patients were studied. The sensitivity of AS was 92.7%; It outperformed each of the other scores. The sensitivity of FS, LS, and TS in women was lower than that in men ($p=0.016$, $p<0.001$, and $p<0.001$, respectively). The sensitivity of the FS, ES, TS, and CS in patients with a duration of illness greater than 48 hours was lower than that in patients with a duration of illness less than 48 hours ($p<0.001$ for all).

Conclusion: AS is the most useful and sensitive diagnostic tool for AA. FS, LS, and TS had a lower diagnostic sensitivity in women; and FS, ES, TS, and CS had a low sensitivity in patients with a duration of illness greater than 48 hours.

Key Words: *Acute appendicitis. Diagnosis. Alvarado score. Sensitivity.*

INTRODUCTION

Acute appendicitis (AA) is the commonest cause of abdominal pain, which require surgery. Symptoms of AA overlap with various conditions, making it difficult to diagnose.[1] The diagnosis in the young, elderly, and female is more difficult to make than in others, because numerous other diseases may behave like AA in these patients.[2] The inability to make an early diagnosis is a critical reason for morbidity and mortality. Based on various reports, the negative appendectomy rate is about 13-40%.[3-5] Therefore, it is necessary to improve the methods for early diagnosis and intervention.

In recent years, ultrasonography and computerized tomography have been used to make a rapid and accurate diagnosis of AA. Subsequently, the negative laparotomy rate decreased to about 10%.[6,7] However, time laging, high expenses, and variable accessibility of imaging methods are the reasons owing to which the current diagnostic techniques do not adequately address the condition, particularly in developing countries where imaging is not widely used.[8] In such situations, the diagnosis of AA relies upon symptoms, physical signs, and laboratory tests.

The purpose of the clinical decision rules is to assist doctors in the evaluation of patients.[9] In 1986, based on symptoms, signs, and the results of diagnostic tests of patients with suspected AA, Alvarado constructed a 10-point clinical scoring system, called the Alvarado score, to assist in the diagnosis of AA.[10] Other scorings have also been developed, such as the modified Alvarado score (MAS), Fenyo-Lindberg score (FS), Lintula score (LS), Eskelinen score (ES), Teicher score (TS), and Christian score (CS).

Some scorings for the diagnosis of AA have been compared, but most of the analyses only focus on comparing between/among two or three scoring systems. However, to the authors' knowledge, the application and usefulness of seven scorings in the diagnosis of AA has not been evaluated in China. The objective of this study was to assess the sensitivity of the seven scorings in the diagnosis of AA.

METHODOLOGY

The present study involved the selection of consecutive patients with AA, who were undergoing appendectomy. Pathologists confirmed appendicitis by histopathological examination in the Department of General Surgery at the First Affiliated Hospital of Chongqing Medical University from January 2012 to June 2015. Information on the disease history, clinical findings, and results of laboratory tests were recorded. Patients in whom histopathological findings were negative, were excluded from the study. The Principal Committee of the First Affiliated Hospital of Chongqing Medical University authorized this research. The seven scoring systems are based on different variables, with different points assigned to each variable according to previous

Department of Gastrointestinal Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China.

Correspondence: Dr. Zhang Hongyu, Youyi Road No.1, Yuan Jiagang, Yuzhong District, Chongqing, China.
E-mail: cy_zhanghongyu@126.com

Received: April 10, 2015; Accepted: December 11, 2017.

reports.[10-16] For example, the following five criteria were used in the diagnosis of AA for the CS: i. Abdominal pain: defined for the study as abdominal pain (not right iliac fossa alone) occurring within 48 hours of presentation, ii. Vomiting: one or more episodes, iii. Right lower quadrant tenderness, iv. Low grade fever: defined as fever ≤38.8ºC, and v. Polymorphonuclear leukocytosis: defined as a total count ≥10,000, with polymorphs ≥75%. Table I and Table II show the criteria that were used in the diagnosis of AA for the other scores.

Table I: Characteristics of the Alvarado score and the Modified Alvarado score, and the distribution of data for individual variables in each scoring system.

Feature	Alvarado score	The distribution of data for individual variables (n=179)	Modified Alvarado score	The distribution of data for individual variables (n=179)
Nausea-vomiting	1	37.99% (n=68)	1	37.99% (n=68)
Anorexia-acetone (in the urine)	1	92.74% (n=166)	1	92.74% (n=166)
Tenderness in right lower quadrant	2	100% (n=179)	2	100% (n=179)
Migration	1	94.97% (n=170)	1	94.97% (n=170)
Rebound pain	1	95.53% (n=171)	1	95.53% (n=171)
Elevation of temperature ≥37.3°C	1	45.25% (n=81)	1	45.25% (n=81)
Leukocytosis >10.0×10^9/L	2	88.27% (n=158)	2	88.27% (n=158)
Shift to the left (>75% neutrophils)	1	94.97% (n=170)		
The cut-off to predict acute appendicitis	≥7		≥7	

Table II: Characteristics of the Fenyo-Lindberg score and the Lintula score, and the distribution of data for individual variables in each scoring system.

Feature	Fenyo-Lindberg score	The distribution of data for individual variables (n=179)	Lintula score	The distribution of data for individual variables (n=179)
Sex				
Male	8	51.40% (n=92)	2	51.40% (n=92)
Female	-8	48.60% (n=87)	0	48.60% (n=87)
Pain duration				
<24h	3	39.66% (n=71)		
24h-48h	0	38.55% (n=69)		
>48h	-12	21.79% (n=39)		
Progression of pain				
Yes	3	98.32% (n=176)		
No	-4	1.68% (n=3)		
Severe pain			2	84.36% (n=151)
Relocation of pain				
Yes	7	94.97% (n=170)	4	94.97% (n=170)
No	-9	5.03% (n=9)		
Vomiting				
Yes	7	37.99% (n=68)	2	37.99% (n=68)
No	-5	62.01% (n=111)		
Aggravation with cough				
Yes	4	100% (n=179)		
No	-11	0% (n=0)		
Body temperature ≥37.5ºC			3	43.58% (n=78)
Pain in RLQ				
Yes	4	100% (n=179)	4	100% (n=179)
No	-6	0% (n=0)		
Rebound tenderness				
Yes	5	95.53% (n=171)	7	95.53% (n=171)
No	-10	4.47% (n=8)		
Rigidity				
Yes	15	17.32% (n=31)	4	17.32% (n=31)
No	-4	82.68% (n=148)		
Bowel sounds change			4	44.70% (n=80)
WBC				
<8.9×10^9/L	-15	10.61% (n=19)		
9-13.9×10^9/L	2	46.93% (n=84)		
>14×109/L	10	42.46% (n=76)		
The cut-off to predict acute appendicitis		Total score -10 ≥-2		≥21

RLQ = Right lower quadrant; WBC = White blood cell.

Statistical analysis was performed using SPSS version 21 for Windows. Ages were presented as mean ± standard deviation (SD). The ratio of male, female, and duration of illness was obtained by dividing the number of male, female, or duration of illness by the total number. The diagnostic sensitivity was defined as the correct number of diagnoses divided by the total number. The diagnostic sensitivity of each score was compared using the Chi-square test. All analyses were two-sided, with statistical significance set at $p<0.05$.

RESULTS

One hundred and seventy-nine patients were enrolled in this study. Their ages ranged from 13 to 87 years (mean 44.13 ±19.52 years). There were 92 (51.4%) males and 87 (48.6%) females (male:female=1.1:1). The variables of scoring systems and the distribution of data for individual variables in each scoring system was given in Tables I, II and III. Table IV summarizes the results of this study. Overall, AS had a sensitivity of 92.7%, while the sensitivity value of the MAS was 88.3%; FS, 87.7%; LS,

Table III: Characteristics of the Eskelinen score and the Teicher score, and the distribution of data for individual variables in each scoring system.

Feature	Eskelinen score	The distribution of data for individual variables (n=179)	Teicher score	The distribution of data for individual variables (n=179)
Sex				
Male			2	51.40% (n=92)
Female			-1	48.60% (n=87)
Age				
20 years			-1	37.43% (n=67)
-39 years				
>50 years			3	32.96% (n=59)
Pain at presentation				
RLQ	7.02	97.21% (n=174)		
No in RLQ	3.51	2.79% (n=5)		
Pain duration				
<48h	4.26	78.21% (n=140)		
≥48h	2.13	21.79% (n=39)		
Duration 1½ days			2	74.30% (n=133)
Duration 2 days			1	3.91% (n=7)
Duration 3 days			-3	21.79% (n=39)
GU symptoms			-3	6.15% (n=11)
Pain in RLQ				
Yes	22.81	100% (n=179)		
No	11.41	0% (n=0)		
Rebound tenderness				
Yes	8.5	95.53% (n=171)		
No	4.25	4.47% (n=8)		
Rigidity				
Yes	13.24	16.76% (n=30)	3	16.76% (n=30)
No	6.62	83.24% (n=149)	-3	83.24% (n=149)
Right-sided rectal mass			-3	1.12% (n=2)
WBC				
$<10\times10^9$/L	5.88	13.41% (n=24)	-3	13.41% (n=24)
$\geq10\times10^9$/L	11.76	86.59% (n=155)		
$>13\times10^9$/L			2	47.49% (n=85)
The cut-off to predict acute appendicitis	≥55		≥-3	

GU = Gastric ulcer; RLQ = Right lower quadrant; WBC = White blood cell.

Table IV: Discriminating capacity of the scores.

	Number	AS	MAS	FS	LS	ES	TS	CS
Overall (sensitivity)	179	166 (92.7%)	158 (88.3%)	157 (87.7%)	142 (79.3%)	164 (91.6%)	157 (87.7%)	139 (77.7%)
Gender								
Men (sensitivity)	92	86 (93.5%)	78 (84.8%)	86 (93.5%)	83 (90.2%)	88 (95.7%)	89 (96.7%)	66 (71.7%)
Women (sensitivity)	87	80 (92.0%)	80 (92.0%)	71 (81.6%)	59 (67.8%)	76 (87.4%)	68 (78.2%)	73 (83.9%)
P		P=0.695	P=0.136	P=0.016	P<0.001	P=0.059	P<0.001	P=0.051
Times								
<48h (sensitivity)	132	125 (94.7%)	119 (90.2%)	124 (93.9%)	104 (78.8%)	127 (96.2%)	128 (97.0%)	118 (89.4%)
≥48h (sensitivity)	47	41 (87.2%)	39 (83.0%)	33 (70.2%)	38 (80.9%)	37 (78.7%)	29 (61.7%)	21 (44.7%)
P		P=0.090	P=0.189	P<0.001	P=0.764	P<0.001	P<0.001	P<0.001

AS = Alvarado score; MAS = Modified Alvarado score; FS = Fenyo-Lindberg score; LS = Lintula score; ES = Eskelinen score; TS = Teicher score; CS = Christian score.

79.3%; ES, 91.6%; TS, 87.7%; and CS, 77.7%. There was a statistically significant difference observed when AS, MAS, FS, ES, and TS were compared with LS ($p<0.001$, $p=0.022$, $p=0.033$, $p=0.001$, and $p=0.033$, respectively) and CS ($p<0.001$, $p=0.008$, $p=0.012$, $p<0.001$, and $p=0.012$, respectively).

For men, the sensitivities of the scores evaluated in this study were 93.5% (AS), 84.8% (MAS), 93.5% (FS), 90.2% (LS), 95.7% (ES), 96.7% (TS), and 71.7% (CS). There was a statistically significant difference between MAS and ES ($p=0.013$), TS ($p=0.005$). And there was a statistically significant difference between CS and the other scores. For women, sensitivities of the scores evaluated in this study were 92.0% (AS), 92.0% (MAS), 81.6% (FS), 67.8% (LS), 87.4% (ES), 78.2% (TS), and 83.9% (CS). There was a statistically significant difference between the LS and the following scores: the AS ($p<0.001$), MAS ($p<0.001$), ES ($p=0.002$), CS ($p=0.013$), and FS ($p=0.036$). Additionally, there was a statistically significant difference between the FS and the AS ($p=0.044$), MAS ($p=0.044$), between the TS and AS ($p=0.011$), MAS ($p=0.011$). The accuracy of the FS, LS, and TS in women was lower than it was in men ($p=0.016$, $p<0.001$, and $p<0.001$, respectively).

For patients with a duration of illness less than 48 hours, the highest sensitivity was 97.0% (TS). There was a statistically significant difference between the MAS and TS ($p=0.024$), between CS and ES ($p=0.032$), TS ($p=0.015$). Additionally, there was a statistically significant difference between LS and the other scores. For patients with a duration of illness greater than 48 hours, the highest sensitivity was 87.2% (AS). As the results differed significantly between FS and AS ($P=0.044$), there was significant difference between the TS and the following other scores: the AS ($p=0.005$), MAS ($p=0.021$), LS ($p=0.040$). Additionally, there was a statistically significant difference between the CS and the other scores, except the TS. The sensitivity of FS, ES, TS, and CS in patients with a duration of illness greater than 48 hours was lower than that in patients with a duration of illness less than 48 hours ($p<0.001$ for all).

DISCUSSION

Even today, this common condition continues to be difficult to diagnose. Scoring systems have been developed to help evaluate patients with AA in the clinical setting. In this study, we sought to assess the sensitivity of the seven scoring systems applied to patients with AA at the First Affiliated Hospital of Chongqing Medical University.

In all the 179 patients with AA, the sensitivity of AS, MAS, FS, ES, and TS was higher than that of the LS and CS. This result partially shows similarity with those of other studies. Ohmann *et al*. measured the performance of 10 scores using one database in a large multicenter trial, using standardized criteria and contrasted the outcomes with published data. Re-examination of the published data showed that AS exceeded each of the other scores, whereas the FS and CS followed.[17] In this study, AS demonstrated the highest sensitivity (92.7%), while LS and CS had low sensitivity. LS was initially created for the pediatric patients; therefore, this may be the reason for the lower accuracy observed in the adults. The aim of CS was to decrease the negative appendectomy rate. This origin may lead to some missed diagnoses of AA.[16] By contrast, if three out of five criteria were present in CS, the process of active observation was initiated for the patient. In this process, no antibiotics were given. If the fourth criterion was met, surgery was performed immediately. However, in this retrospective review, it was difficult to ensure that active observation was performed carefully in such a way that no antibiotics were given. These factors may have affected the accuracy of CS.

Some studies have found that the sensitivity of a scoring system for appendicitis is influenced by gender, so the data was analysed according to gender. FS, LS, and TS exhibited better performance in terms of sensitivity in the subgroup analysis of men than in women. Female patients with tenderness in the right lower quadrant can potentially have a variety of conditions, including pelvic infectious disease and other gynecological pathologies. Therefore, AA may be misdiagnosed in female patients. Depending on the gender of patients, FS and LS have different numerical values for the same symptoms. Gender is also one of the factors considered in TS. However, in this study, these considerations did not improve the sensitivity of these scoring systems in women, which is consistent with other reports.[17]

In this study, it was observed that the longer the duration of illness, the poorer the sensitivity of FS, ES, TS, and CS. There are a number of other reasons for this. FS, ES, and TS include the duration of pain as one of their components, and have different numerical values for symptoms depending on the length of time. The longer course was associated with lower scores. In addition, in cases involving a longer course of the disease, patients are more likely to be treated with analgesics and antibiotics that may mask the symptoms. These factors may have affected the accuracy of the scores.

Many scoring systems for AA have been created. AS is the most widely used system, and demonstrates best performance in validation studies. AS is the only scoring system presented in the document published by the American College of Emergency Physicians to guide decision-making for AA. Professionals within this organisation believe that combining different symptoms and signs into a scoring system may be more valuable in the prediction of appendicitis.[18] Some scholars think that the scoring weights of AS may be biased for patients

who are suspected of having appendicitis, because the system was created retrospectively, involving patients who had all been operated upon, with the suspicion of appendicitis.[19] In this study, the sensitivity of AS was up to 92.7% and it outperformed the other scores. A systematic review found that the AS is a useful diagnostic method for all patient groups when a cut-off score of 5 is used. Using this cut-off score, only two people missed being diagnosed in this study. However, Gwynn found that about 8.4% of patients with appendicitis had an AS below 5 in a Class III study. Therefore, the cut-off score of AS requires further optimisation in the future.[20]

Today, AA continues to be a troublesome disease. Scoring systems are intended to help the clinical diagnosis of patients with AA. To the authors' knowledge, this is the first time that seven scoring systems for appendicitis have been evaluated and compared in China. However, the sample population needs to be expanded, and the specificity of each score should be analysed. A prospective randomised controlled trial is to be conducted to evaluate the effect of those scores in China in the near future.

CONCLUSION

FS, LS, and TS have a low diagnostic sensitivity in women, while FS, ES, TS, and CS have a low sensitivity in patients with a duration of illness greater than 48 hours. Alvarado score remained the most sensitive test in all comparisons.

REFERENCES

1. Bundy DG, Byerley JS, Liles EA, Perrin EM, Katznelson J, Rice HE. Does this child have appendicitis? *JAMA* 2007; **298**: 438-51.
2. Gilmore OJ, Browett JP, Griffin PH, Ross IK, Brodribb AJ, Cooke TJ, *et al*. Appendicitis and mimicking conditions. A prospective study. *Lancet* 1975; **2**:421-4.
3. Kalan M, Talbot D, Cunliffe WJ, Rich AJ. Evaluation of the modified Alvarado score in the diagnosis of acute appendicitis: a prospective study. *Ann R Coll Surg Engl* 1994; **76**:418-9.
4. Andersson RE, Hugander A, Thulin AJ. Diagnostic accuracy and perforation rate in appendicitis: association with age and sex of the patient and with appendicectomy rate. *Eur J Surg* 1992; **158**:37-41.
5. Hale DA, Molloy M, Pearl RH, Schutt DC, Jaques DP. Appendectomy: a contemporary appraisal. *Ann Surg* 1997; **225**:252-61.
6. SCOAP Collaborative, Cuschieri J, Florence M, Flum DR, Jurkovich GJ, Lin P, *et al*. Negative appendectomy and imaging accuracy in the Washington State Surgical Care and Outcomes Assessment Program. *Ann Surg* 2008; **248**:557-63.
7. Wagner PL, Eachempati SR, Soe K, Pieracci FM, Shou J, Barie PS. Defining the current negative appendectomy rate: for whom is preoperative computed tomography making an impact? *Surgery* 2008; **144**:276-82.
8. Ohle R, O'Reilly F, O'Brien KK, Fahey T, Dimitrov BD. The Alvarado score for predicting acute appendicitis: a systematic review. *BMC Med* 2011; **9**:139.
9. Meltzer AC, Baumann BM, Chen EH, Shofer FS, Mills AM. Poor sensitivity of a modified Alvarado score in adults with suspected appendicitis. *Ann Emerg Med* 2013; **62**:126-31.
10. Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med* 1986; **15**:557-64.
11. Kanumba ES, Mabula JB, Rambau P, Chalya PL. Modified alvarado scoring system as a diagnostic tool for acute appendicitis at Bugando Medical Centre, Mwanza, Tanzania. *BMC Surg* 2011; **11**:4.
12. Enochsson L, Gudbjartsson T, Hellberg A, Rudberg C, Wenner J, Ringqvist I, *et al*. The Fenyö-Lindberg scoring system for appendicitis increases positive predictive value in fertile women – a prospective study in 455 patients randomized to either laparoscopic or open appendectomy. *Surg Endosc* 2004; **18**:1509-13.
13. Lintula H, Kokki H, Pulkkinen J, Kettunen R, Gröhn O, Eskelinen M. Diagnostic score in acute appendicitis. Validation of a diagnostic score (Lintula score) for adults with suspected appendicitis. *Langenbecks Arch Surg* 2010; **395**:495-500.
14. Sitter H, Hoffmann S, Hassan I, Zielke A. Diagnostic score in appendicitis. Validation of a diagnostic score (Eskelinen score) in patients in whom acute appendicitis is suspected. *Langenbecks Arch Surg* 2004; **389**:213-8.
15. Teicher I, Landa B, Cohen M, Kabnick LS, Wise L. Scoring system to aid in diagnoses of appendicitis. *Ann Surg* 1983; **198**:753-9.
16. Christian F, Christian GP. A simple scoring system to reduce the negative appendicectomy rate. *Ann R Coll Surg Engl* 1992; **74**:281-5.
17. Ohmann C, Yang Q, Franke C. Diagnostic scores for acute appendicitis. Abdominal Pain Study Group. *Eur J Surg* 1995; **161**:273-81.
18. Howell JM, Eddy OL, Lukens TW, Thiessen ME, Weingart SD, Decker WW, *et al*. Clinical policy: Critical issues in the evaluation and management of emergency department patients with suspected appendicitis. *Ann Emerg Med* 2010; **55**:71-116.
19. Andersson M, Andersson RE. The appendicitis inflammatory response score: a tool for the diagnosis of acute appendicitis that outperforms the Alvarado score. *World J Surg* 2008; **32**:1843-9.
20. Gwynn LK. The diagnosis of acute appendicitis: clinical assessment versus computed tomography evaluation. *J Emerg Med* 2001; **21**:119-23.