

Are scoring systems detecting acute appendicitis reliable? a prospective clinical study

 Mustafa Taner Bostancı¹,  İbrahim Yılmaz¹,  Sanem Güler Cimen¹,  Koray Kosmaz²,
 Aysun Gökçe³,  Mehmet Alperen Avcı⁴

¹University of Health Sciences, Dışkapı Yıldırım Beyazıt Education and Research Hospital, Department of General Surgery, Ankara, Turkey

²University of Health Sciences, Ankara Training and Research Hospital, Department of Surgery, Ankara, Turkey

³University of Health Sciences, Dışkapı Yıldırım Beyazıt Education and Research Hospital, Department of Pathology, Ankara, Turkey

⁴Sorgun State Hospital, Department of General Surgery, Yozgat, Turkey

Cite this article as: Bostancı MT, Yılmaz İ, Güler Cimen S, Koşmaz K, Gökçe A, Avcı MA. Are scoring systems detecting acute appendicitis reliable? a prospective clinical study. J Health Sci Med 2022; 5(1): 167-172.

ABSTRACT

Aim: The diagnosis of acute appendicitis is still challenging. Negative appendectomy rate is reported as 20-30%. Therefore, various scoring systems have been developed to prevent unnecessary appendectomies. The aim of our study is to analyze the diagnostic value of different scoring systems in acute appendicitis.

Material and Method: The study conducted with 200 consecutive patients who were operated with the diagnosis of acute appendicitis. After the admission and deciding the operation via experienced general surgeon, the probability of acute appendicitis was calculated using via 6 different scoring systems: Alvarado, Eskelinen, Ohmann, Lintula, RIPASA, Fenyo, respectively, and results were analyzed statistically. $p < 0.05$ was considered statistically significant.

Results: RIPASA had the highest sensitivity (83.8%) and accuracy (78%) ratio in all patient groups; The Lintula scoring system had the highest specificity (66.7%). Sensitivity values were found to be higher in males, and the sensitivity of Lintula (44.6%/12.6%; $p: 0.001$) and Fenyo (67.6%/19.4%; $p < 0.001$) scores were significantly higher in males. The specificity value was significantly higher for Eskelinen (85.7%/35.7%; $p: 0.031$) and Lintula (100%/50%; $p: 0.022$) in females.

Conclusion: RIPASA is the scoring system with the highest sensitivity for both genders. Eskelinen and Lintula is noteworthy in females. We believe that the use of scoring tests with classifying according to gender may lead to a decrease in negative appendectomies.

Keywords: Acute appendicitis, scoring systems, appendectomy, clinical decision

INTRODUCTION

Acute appendicitis is one of the most common causes of abdominal pain in patients admitted to the emergency, and acute appendectomy operation is the major part of general surgery education (1). Diagnosis of acute appendicitis should be done as soon as possible to minimize morbidity, mortality and unnecessary surgical interventions. However, despite the improvements in the accuracy of diagnostic methods, the level of diagnostic failings reach 20-30% (2). The diagnosis depends on history, symptoms, physical examination, and laboratory findings. Routine radiological examinations are also used in the emergency department, especially in clinically ill-defined patients. However, radiological examinations are time-consuming methods and have not been found to be specific (3). These challenges in diagnosing acute appendicitis led researchers to develop some clinical scoring systems and

consequently, many scoring systems have been described from the 1980s (4-8). In our study, we aimed to analyze the effectiveness of different scoring systems in the diagnosis of acute appendicitis in randomized patients hospitalized for appendectomy operation.

MATERIAL AND METHOD

This prospective study was conducted with 200 consecutive patients who underwent appendectomy with the diagnosis of acute appendicitis between November 2017 and June 2018, after the approval of Clinical Research Ethics Committee of Health Sciences University Dışkapı Yıldırım Beyazıt Training and Research Hospital (Date: 08.10.2021, Decision No: 896), and under circumstances of Helsinki Declaration of Principles.

Each patient was hospitalized after examination by one same general surgeon with a diagnosis of acute appendicitis, and detailed anamnesis, physical examination findings and laboratory tests results (complete blood count, biochemistry, urine test) were enrolled. Informed consent was obtained from all individual participants included in the study. Preoperative abdominal ultrasonography (USG) was performed to 169 patients, and abdominal computed tomography (CT) was performed to 98 patients. A questionnaire consisting of the parameters of the whole scoring systems included in the study was filled out according to the symptom and test results determined by the general surgery residents for each patient. Then the acute appendicitis probability was calculated according to six different scoring systems and the obtained scores were recorded. The cut-off points as described in the original report of each scoring system were used. The assumed Cut-Off (COP) points were 7 for Alvarado, 55 for Eskelinen, 12 for Ohmann, 21 for Lintula, 7.5 for RIPASA, and -2 for Fenyo (**Table 1**) (4-9).

The same general surgeon who decided to perform the operation did not know the results of the scoring before the operation. All of the pathologic materials were worked on and analyzed histopathologically and the results confirmed via the same pathologist after the operation. Neutrophilic infiltration of the appendiceal wall, mucosal ulceration, cryptitis, crypt abscesses, necrosis

were determined as appendicitis histopathologically, and without each of these findings, cases were accepted as “disease absent category”. Inclusion criterias were all patients with phlegmonous, gangrenous or perforated appendicitis, and had suspended or ill-defined clinics without age or gender limitation. All patients were informed about the study and written informed consent was obtained from them. Patients who did not consent the study or operation, used antibiotics and referred from another hospital to our hospital were excluded from study.

Statistical Analysis

The data were analyzed by the Statistical Package for Social Sciences 20.0 (SPSS). In order to determine the sensitivity, specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), accuracy values of scoring systems; a cross-table and the Mc Nemar test, Chi-square test were used. $p < 0.05$ was considered statistically significant.

RESULTS

The mean age was 31 (19-88), and 162 patients (81%) were male. Pathology reports were found to be acute appendicitis in 179 patients, while the remaining 21 pathologic material were evaluated as innocent appendix. A negative appendectomy rate was 10.5%. Thirty-six of

Table1. Clinical parameters and cut-off points of the scoring tests						
Clinical signs	Alvarado	Eskelinen	Ohman	Lintula	RİPASA	Fenyo
Age			+		+	
Gender				+	+	+
Duration of pain (<48sa)		+			+	+
Increase of pain level						+
Pain level				+		
Continuous pain			+			
Pain localized in right lower quadrant		+	+	+	+	
Migration of pain	+		+	+	+	+
Loss of appetite	+				+	
Nausea & vomiting	+			+	+	+
Difficulty in miction						
Increased pain when cough						+
Right lower quadrant tenderness	+	+			+	
Rigidity		+				+
Defense			+	+	+	
Rebound	+	+	+	+	+	+
Bowel sounds				+		
Rovsing's sign					+	
Fever (>37.3°C)	+			+	+	
Leucocyte (>10000)	+	+	+		+	+
Neutrophilia (>% 75)	+					
No dysuric symptoms			+			
Negative urine test					+	
Foreigner					+	
Cut-off	7	55	12	21	7.5	-2

the patients (18%) were undergone laparoscopic surgery, whereas the others (82%) were undergone conventional surgery. Sensitivity and specificity rates of the clinical parameters were summarized in **Table 2**. The comparison of scoring systems with the pathology results is shown in **Table 3**. The data obtained from the comparison of scoring systems for male and female genders are shown in **Table 4**, and **Table 5**. The sensitivity of Lintula (p=0.001)

and Fenyo (p<0.001) scoring systems were significantly higher in males. The specificity of Eskelinen (p=0.031) and Lintula (p=0.022) scoring systems were significantly higher in females. We revealed that abdominal CT was superior to abdominal USG in terms of sensitivity (p<0.001), specificity (p=0.013) and accuracy rate (p<0.001) in the evaluation of the radiologically obtained data compared to histopathology results (**Table 6**).

Table 2. Comparison of clinical signs in terms of sensitivity, specificity, PPV*, NPV°

Clinical signs	Sensitivity	Specificity	PPV*	NPV°	Accuracy
Duration of pain (<48sa)	73.2	28.6	89.7	11.1	68.5
Increase of pain level	69.3	52.4	92.5	16.7	67.5
Pain level	58.7	33.3	88.2	8.6	56
Continuous pain	92.7	28.6	91.7	31.6	86
Pain localized in right lower quadrant	59.8	52.4	91.5	13.3	59
Loss of appetite	70.9	61.9	94.1	20.0	70
Nausea & vomiting	49.7	47.6	89.0	10.0	49.5
Difficulty in miction	9.5	100	100	11.5	19
Increased pain when cough	56.4	61.9	92.7	14.3	57
Right lower quadrant tenderness	98.3	0	89.3	0	88
Rigidity	35.2	66.7	90.0	10.8	38.5
Defense	64.2	38.1	89.8	11.1	61.5
Rebound	72.6	47.6	92.2	16.9	70
Bowel sounds	7.3	95.2	92.9	10.8	16.5
Rovsing's sign	17.3	81.0	88.6	10.3	24
Fever (>37.3°c)	79.9	9.5	88.3	5.3	72.5
Leucocyte (>10000)	19.0	76.2	87.2	9.9	26
Neutrophilia (>% 75)	52.0	28.6	86.1	6.5	49.5
Negative urine test	78.2	19.0	89.2	9.3	72

PPV*=Positive predictive value, NPV°=Negative predictive value

Table 3. Comparison scoring tests in terms of sensitivity, specificity*, PPV, and NPV°

Scoring tests	Sensitivity	Specificity	PPV*	NPV°	Accuracy
Alvarado	67	61.9	93.8	18.1	66.5
Eskelinen	64.8	52.4	92.1	14.9	63.5
Ohmann	76.5	42.9	91.9	17.6	73
Lintula	39.1	66.7	90.9	11.4	42
Fenyo	59.2	57.1	92.2	14.1	59
RIPASA	83.8	28.6	90.9	17.1	78

PPV*=Positive predictive value, NPV°=Negative predictive value

Table 4. Comparison of sensitivity of scoring tests according to gender

Gender	n	Sensitivity					
		Alvarado	Eskelinen	Ohmann	Lintula	Fenyo	RIPASA
Male	162	68.9%	66.9%	77.7%	44.6%	67.6%	83.8%
Female	38	58.1%	54.8%	71.0%	12.9%	19.4%	83.9%
p		0.242	0.131	0.822	0.001	<0.001	0.99

Table 5. Comparison of specificity of scoring tests according to gender

Gender	n	Specificity					
		Alvarado	Eskelinen	Ohmann	Lintula	Fenyo	RIPASA
Male	162	57.1%	35.7%	28.6%	50.0%	42.9%	21.4%
Female	38	71.4%	85.7%	71.4%	100%	85.7%	42.9%
p		0.525	0.031	0.061	0.022	0.061	0.306

Table 6. Comparison of imaging methods with pathological results

Screening Tests	Sensitivity	Specificity	PPV*	NPV°	Accuracy
Abdominal USG	78.1	11.1	87.3	4.6	66.2
Abdominal CT	95.1	50	90.6	66.6	87.7
p	<0.001	0.013	0.444	<0.001	<0.001

USG=Ultrasonography, CT=Computed tomography, PPV*=Positive predictive value, NPV°=Negative predictive value

DISCUSSION

Acute appendicitis is the most common cause of abdominal pain in emergency clinics worldwide. As a matter of fact the imminent appendectomy prevents many complications, and the quick and accurate diagnosis of acute appendicitis must be performed in time (10). The fact that the clinical entity may show different clinical manifestations in early period and lack of a fully reliable diagnostic test may leave surgeons unstable in the diagnosis, and in addition this situation may lead them to perform unnecessary appendectomy operation for some cases.

To date, the most commonly used imaging modality for the diagnosis of acute appendicitis is abdominal USG. Although abdominal USG, abdominal CT and laparoscopy are used effectively in the diagnosis, the rate of negative appendectomy is 20-30% and the rate of non-diagnosed perforated appendicitis is 4% (11,12). It is known that the technical factors such as the experience of the radiologist who examined the diagnostic accuracy of abdominal USG and the quality of the device and the patient's factors such as body mass index, localization of the appendix, and density of intestinal gases affect the diagnostic accuracy of abdominal USG. A range of sensitivity and specificity varies 44-100%, 47-99%, respectively (13). Abdominal CT is a more reliable method for the diagnosis of acute appendicitis and it is reported that the sensitivity and specificity rates are 94%, 95%, respectively (14). However, abdominal CT have some disadvantages such as exposure to ionizing radiation, long preparation time and higher cost (15). Wani et al. (16) stated that the indiscriminate use of abdominal CT could cause ambiguous results especially in patients with early appendicitis and may lead to unnecessary appendectomy in patients with spontaneous regression. In our study, abdominal CT was superior to abdominal USG in terms of sensitivity, specificity and accuracy rates in the diagnosis of acute appendicitis correlated with the literature.

The pain that begins in the epigastric region and is located in the right lower quadrant is typical for acute appendicitis. In our study, patients who were admitted to the emergency department with right lower quadrant pain and were operated with a preliminary diagnosis of acute appendicitis were expected to have a high sensitivity in the "right lower quadrant tenderness" parameter when evaluated with histopathological findings. In addition, "acute continuous abdominal pain" parameter had high sensitivity value and the accuracy rate was remarkable. Moreover the parameters of "no difficulty in urination", "normoactive bowel sounds", and "positive Rovsing sign" showed high specificity values.

Consequently, the difficulties in diagnosing acute appendicitis led clinicians define various scoring systems in order to reduce the frequency of negative surgery. One of them, Alvarado score, is the first known for the diagnosis of acute appendicitis and is the most well known in the general surgery society. The score is calculated over 10 points and a score of ≥ 7 is considered to be an acute appendicitis marker. The sensitivity rates in the literature are reported to be 39-100% and the specificity rates are between 41-98% (17). Eskelinen score is developed to be more specific to female gender. When a score of ≥ 55 was determined as the cut off value, it was found that the sensitivity of 79% and the specificity of 85% in literature (5). The decimal value of the Eskelinen score is the disadvantage of its practical application (5). The Ohmann score was developed to reduce the number of negative appendectomies and ≥ 12 score was accepted as a marker of acute appendicitis. Ohmann stated that he reduced negative appendectomy rates to 2% (6). The Lintula score is another scoring system with a specificity of 87% of its sensitivity when it's score is ≥ 21 (98%) (7). The RIPASA score, which was defined in recent years, is another scoring system, and it is reported that sensitivity is 87%, and specificity is 67% when the score is ≥ 7.5 (8). It includes some parameters such as age, gender and duration of symptoms which are lack in the Alvarado score (8). Another scoring system, the Fenyo-Lindberg score, was reported to have a sensitivity of 73% and a specificity of 73% when the score is ≥ -2 (9).

Almost all scoring systems concerning about diagnosis of appendicitis varies in different countries depending on the differing nutritional, working and cultural conditions. In our country, there are some studies about the accuracy of scoring systems in the diagnosis of acute appendicitis. Erdem et al. (12) found that the specificity of Ohmann and RIPASA tests was higher to the Alvarado and Eskelinen tests for the diagnosis of acute appendicitis in their study which is conducted with 113 patients. Kırkıl et al. (18) reported that the sensitivity of the Lintula test was higher than the Alvarado test in a study of 114 patients. Yılmaz et al. (19) reported that Alvarado test was more sensitive than Ohmann test in the diagnosis of acute appendicitis. Yoldas et al. (20) reported that, when evaluating the accuracy of the Lintula score in the diagnosis of acute appendicitis, it was an appropriate method for Turkish society.

Currently, the studies considering two or three scoring systems are quite common compared to studies checked more than three scoring system in the literature. Besides, it is accepted as highly valuable when the sensitivity and specificity values are above 80% (21). In our study, when we compared the results of frequently used six different scoring systems statistically, even its specificity was not

too high, the 83.8% sensitivity ratio obtained for RIPASA scoring system was remarkable. When we compared the scoring systems in terms of gender variable, different results were obtained. In fact, the sensitivity values of scoring systems were generally higher in males and also this difference was statistically significant for Fenyo, and Lintula in our study. When we analyzed in terms of specificity, the difference was statistically significant for Eskelinen, and Lintula which were higher in females compared to males. RIPASA was the scoring system with a sensitivity ratio of over 80% for both sexes according to the current study, and these findings were correlated with the literature (21).

We did not include possible appendicitis cases; our population is compromised of diagnosed appendicitis by general surgeon with clinical findings, physical examination, laboratory and imaging findings without scoring systems results. We could have missed some true negative and false negative data which could affect the NPV, PPV, sensitivity, specificity rates. We focused on only clinically diagnosed appendicitis cases.

CONCLUSION

RIPASA had the highest sensitivity for general population and the genders according to the study. Eskelinen and Lintula were noteworthy in females with high specificity values. Moreover, Fenyo and Lintula had considerably high sensitivity values in males. We concluded that the use of these tests in question with classifying distinctively according to gender may reduce the rate of negative appendectomy especially in females and in the patients not in the most common seen age group and may help clinicians in diagnosis.

ETHICAL DECLARATIONS

Ethics Committee Approval: The study was carried out with the permission of Clinical Research Ethics Committee of Health Sciences University Dışkapı Yıldırım Beyazıt Training and Research Hospital (Date: 08.10.2021, Decision No: 896).

Informed Consent: All patients signed the free and informed consent form.

Referee Evaluation Process: Externally peer-reviewed.

Conflict of Interest Statement: The authors have no conflicts of interest to declare.

Financial Disclosure: The authors declared that this study has received no financial support.

Author Contributions: All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

- Rathod S, Ali I, Bawa AP, Singh G, Mishra S, Nongmaithem M. Development of the RIPASA score: a new appendicitis scoring system for the diagnosis of acute appendicitis. *Singapore Med J* 2010; 51: 220-5.
- Horzić M, Salamon A, Kopljar M, Skupnjak M, Cupurdija K, Vanjak D. Analysis of scores in diagnosis of acute appendicitis in women. *Coll Antropol* 2005; 1: 133-8.
- Escriba A, Gamell AM, Fernandez Y, Quintilla JM, Cubells CL. Prospective validation of two systems of classification for the diagnosis of acute appendicitis. *Pediatr Emerg Care* 2011; 27: 165-9.
- Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med* 1986; 15: 557-64.
- Eskelinen M, Ikonen J, Lipponen P. A computer based diagnostic score to aid in diagnosis of acute appendicitis: a prospective study of 1333 patients with acute abdominal pain. *Theor Surg* 1992; 7: 86-90.
- Ohmann C, Yang Q, Franke C. Diagnostic scores for acute appendicitis. *Abdominal Pain Study Group. Eur J Surg* 1995; 161: 273-81.
- 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.
- Chong CF, Adi MI, Thien A, et al. Development of the RIPASA score: a new appendicitis scoring system for the diagnosis of acute appendicitis. *Singapore Med J* 2010; 51: 220-5.
- Fenyo G, Lindberg G, Blind P, Enochsson L, Oberg A. Diagnostic decision support in suspected acute appendicitis: validation of a simplified scoring system. *Eur J Surg* 1997; 163: 831-8.
- 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.
- Toorenvliet BR, Wiersma F, Bakker RF, Merkus JW, Breslau PJ, Hamming JF. Routine ultrasound and limited computed tomography for the diagnosis of acute appendicitis. *World J Surg* 2010; 34: 2278-85.
- Erdem H, Çetinküner S, Daş K, et al. Alvarado, Eskelinen, Ohhmann and Raja Isteri Pengiran Anak Saleha Appendicitis scores for diagnosis of acute appendicitis. *World J Gastroenterol* 2013; 19: 9057-62.
- Pinto F, Pinto A, Russo A, et al. Accuracy of ultrasonography in the diagnosis of acute appendicitis in adult patients: review of the literature. *Crit Ultrasound J* 2013; 5: 1-3.
- Terasawa T, Blackmore CC, Bent S, Kohlwes RJ. Systematic review: Computed tomography and ultrasonography to detect acute appendicitis in adults and adolescents. *Ann Intern Med* 2004; 141: 537-46.
- Ozkan S, Duman A, Durukan P, Yildirim A, Ozbakan O. The accuracy rate of Alvarado score, ultrasonography, and computerized tomography scan in the diagnosis of acute appendicitis in our center. *Niger J Clin Pract* 2014; 17: 413-8.
- Wani MM, Yousaf MN, Khan MA, et al. Usefulness of the Alvarado scoring system with respect to age, sex and time of presentation, with regression analysis of individual parameters. *Internet J Surg* 2007; 11: 1-5.
- Nirajlal B, Gabriel R, Anand R, Sohil AK. Evaluation of Alvarado score in acute appendicitis: a prospective study. *Internet J Surg* 2011; 9: 1-5.
- Kırkil C, Karabulut K, Aygen E, et al. Appendicitis scores may be useful in reducing the costs of treatment for right lower quadrant pain. *Ulus Travma Acil Cerrahi Derg* 2013; 19: 13-9.

19. Yılmaz EM, Kapçı M, Çelik S, Manoğlu B, Avcil M, Karacan E. Should Alvarado and Ohmann scores be real indicators for diagnosis of appendicitis and severity of inflammation? *Ulus Travma Acil Cerrahi Derg* 2017; 23: 29-33.
20. Yoldas O, Karaca T, Tez M. External validation of Lintula score in Turkish acute appendicitis patients. *Int J Surg* 2012; 10: 25-7.
21. Konan A, Hayran M, Kılıç YA, Karakoç D, Kaynaroğlu V. Scoring systems in the diagnosis of acute appendicitis in the elderly. *Ulus Travma Acil Cerrahi Derg* 2011; 17: 396-400.