A prospective comparative study among Alvarado score, appendicitis inflammatory response (AIR) score, and adult appendicitis score (AAS) in the diagnosis of acute appendicitis in adults

Moataz T. Abouelnour, Tamer M. Elmahdy, Hamdy S. Abdallah, Osama H. Elkhadrawy

General Surgery Department, Faculty of Medicine, Tanta University, Tanta, Egypt

Correspondence to Moataz T. Abouelnour, BSc, Tanta main post office / 31511. Mobile: +0109 627 2172; Tel: 01096272172; e-mail: moataz.aboelnour@med.tanta.edu.eg

Received: 7 July 2023
Revised: 24 July 2023
Accepted: 1 August 2023
Published: 6 October 2023

The Egyptian Journal of Surgery 2023, 42:795–800

Background
Acute appendicitis (AA) is the most common condition presenting with acute abdomen requiring emergency surgery. Several clinical scores were developed to diagnose AA.
Aim
This study aimed to compare the diagnostic accuracy of three clinical scores (Alvarado score, appendicitis inflammatory response (AIR) score, and adult appendicitis score (AAS)) in the diagnosis of AA.
Methods
This prospective study was conducted on 100 adult patients (aged $\geq$ 18 years) presenting with acute abdomen suspicious of noncomplicated AA. Patients with complicated AA (abscess, mass, or diffuse peritonitis) and patients with uncomplicated AA subjected to nonoperative management (NOM) were excluded from our study. The histopathological results were used as the gold standard for diagnosis of AA to which the three clinical scores results were compared.
Results
Our study included 100 patients (60 males and 40 females), all of whom, had histopathologically proven AA. Due to this outcome, we could only conclude and compare the sensitivities of the three scores. Alvarado score had the highest overall sensitivity (91% at cut-off value $>$ 4 points), followed by AAS (80% at cut-off value $>$ 10 points), then AIR score (71% at cut-off value $>$ 4 points) however, the difference between Alvarado score and AAS was not statistically significant. In males, the Alvarado score had the highest sensitivity (88.3%), followed by AAS (86.7%), then the AIR score (71.7%) at the same cut-off values however, the difference between the Alvarado score and AAS was also not statistically significant. In females, Alvarado score had the highest sensitivity (95%) followed by AAS and AIR scores (70% for both) with a statistically significant difference. Obesity did not influence the outcome of the three scores.
Conclusions
Alvarado score and AAS are more sensitive than the AIR score. In relation to sex, Alvarado's score is the most sensitive in females.

Keywords:
adult appendicitis score, appendicitis inflammatory response, Alvarado, appendicitis

Egyptian J Surgery 42:795–800

1110-1121

Introduction

Acute appendicitis (AA) is the most common surgical condition presenting with acute abdomen requiring emergency surgery with a lifetime incidence of 10% in the general population [1].

Despite its common presentation, correct diagnosis of AA remains a clinical challenge with reported rates of misdiagnosis ranging between 20% and 40% and rate of negative appendectomy (NA) ranging between 7% and 12% [2,3].

A delay or misdiagnosis of AA can result in severe complications such as perforation, abscess formation, sepsis, and intraabdominal adhesions [4]. On the other hand, NA can be associated with a considerable increase in costs, length of stay, and morbidity [5].

Several clinical scoring systems were developed to diagnose AA. The most commonly used of these are the Alvarado score [6], the appendicitis inflammatory response (AIR) score [7], and the adult appendicitis score (AAS) [8].

Patients and methods

This prospective study was conducted in the gastrointestinal and laparoscopic surgery unit, General Surgery Department, Faculty of Medicine, Tanta University, Egypt during the period from June 2021 to July 2022 on 100 adult patients (aged ≥ 18 years) presenting to Tanta University Emergency Hospital with acute abdominal pain suspicious of noncomplicated AA.

Inclusion criteria

(1) Patients aged greater than or equal to 18 years.
(2) Patients with acute nontraumatic right lower quadrant (RLQ) pain suspicious of AA.

Exclusion criteria

(1) Patients with complicated AA (abscess, mass, or with diffuse peritonitis).
(2) Patients with uncomplicated AA subjected to nonoperative management (NOM).

After physical, laboratory and radiologic assessments, the three clinical scores (Alvarado score, AIR score, and AAS) were recorded for every patient using specific tables at their initial presentation by a surgeon who was not involved in decision-making (Tables 1–3).

Based on the discretion of a senior attending surgeon who was blinded to the patient's scores, appendectomy was performed when clinical, laboratory, and ultrasound examination met the diagnosis of AA.

The excised appendices were then, sent for histopathological examination. The diagnosis of AA was confirmed when transmural infiltration of the appendiceal wall by neutrophils was found. AA was further classified into phlegmonous appendicitis (acute suppurative appendicitis) defined as neutrophilic infiltration of the appendiceal wall with associated inflamed and ulcerated mucosa and often crypt abscesses or advanced (gangrenous, perforated) appendicitis.

The results of the three clinical scores were compared with the histopathological results for every patient.

Table 1 Alvarado score parameters and algorithm [6]

Symptoms	
Migration	1
Anorexia-acetone	1
Nausea-vomiting	1
Signs	
Tenderness in the RLQ	2
Rebound pain	1
Elevation of temperature greater than 37.5°C	1
Laboratory	
Leukocytosis	2
Shift to the left	1
Total score 10	
1–4 unlikely, 5–6 possible, 7–8 present and 9–10 definite acute appendicitis	

Table 2 AIR score parameters and algorithm [7]

Vomiting	1
Pain in right inferior fossa	1
Rebound tenderness or muscular defense	
Light	1
Medium	2
Strong	3
Body temperature greater than 38.5°C	1
Polymorphonuclear leukocytes	
70–84%	1
≥85%	2
WBC count	
10-14.9×10^9/l	1
≥15×10^9/l	2
CRP concentration	
10-49 g/l	1
≥50 g/l	2
Sum (0–12)	
0-4 low probability, 5-8 intermediate probability and 9–12 high probability	

Table 3 AAS parameters and algorithm [8]

Symptoms and findings		Score
Pain in RLQ		2
Pain relocation		2
RLQ tenderness		3/1*
Guarding	Mild	2
	Moderate/ Severe	4
Laboratory tests		**Score**
Blood leukocyte count (x10^9)	≥7.2 and < 10.9	1
	≥10.9 and < 14	2
	≥14	3
Proportion of neutrophils (%)	≥62 and < 75	2
	≥75 and < 83	3
	≥83	4
CRP (mg/l), symptoms less than 24 h	≥4 and < 11	2
	≥11 and < 25	3
	≥25 and < 83	5
	≥83	1
CRP (mg/l), symptoms greater than 24 h	≥12 and < 53	2
	≥53 and < 152	2
	≥152	1

RLQ, right lower abdominal quadrant. * Men and women aged 50 +/women 16–49. 0-10 low probability, 11–15 intermediate probability, greater than or equal to 16 high probability,greater than or equal to 18 extra high probability.

Data were fed to the computer and analyzed using IBM SPSS software package version 20.0. (Armonk, NY: IBM Corp) Qualitative data were described using number and percent. The Kolmogorov-Smirnov test was used to verify the normality of distribution Quantitative data were described using range (minimum and maximum), mean, standard deviation, median and interquartile range. Significance of the obtained results was judged at the 5% level.

As a patient requiring surgery, informed consent from every patient was obtained before surgery and full explanation of the study and its purposes were given.

Each patient's file was given a code number and his/her name was obscured from the investigations.

Our study was approved by the Research Ethics Committee of Faculty of Medicine, Tanta University in 28/6/2021 (Approval Code: 34768/6/21).

Results

Our study included 100 patients all of whom, had histologic evidence of AA. The demographic characteristics of the studied population and the histopathological results are illustrated in (Table 4).

Due to the histopathological confirmation of AA in 100% of the studied patients, only the sensitivity of each score for the diagnosis of AA could be calculated. Specificity, positive predictive value, negative predictive value, diagnostic accuracy, and receiver operating characteristic curve could not be calculated.

Alvarado score had a higher overall sensitivity and lower false negative rate (FNR) (91% and 9%, respectively at cut-off value $>$ 4 points) than AAS (80% and 20%, respectively at cut-off value $>$ 10 points) but without statistically significant difference. AIR score had the lowest sensitivity and the highest FNR (71% and 29%, respectively at cut-off value $>$ 4 points) which achieved statistically significant difference in relation to both Alvarado score and AAS (Table 5).

In males, Alvarado score had a higher sensitivity and lower FNR (88.3% and 11.7% respectively) than AAS (86.7% and 13.3%, respectively) but without statistically significant difference. AIR score had the lowest sensitivity and the highest FNR (71.7% and 28.3%, respectively) which achieved statistically significant difference in relation to both Alvarado score and AAS (Table 5).

In females, Alvarado score's sensitivity (95% at cut-off value $>$ 4 points) was statistically significantly higher than both AIR score and AAS sensitivities (70% for both at cut-off values $>$4 points and $>$10 points, respectively) (Table 5).

Discussion

AA is one of the most common surgical causes of acute abdomen. The importance of accurately diagnosing AA lies in reducing the cost related to NA and the morbidity and mortality related to missed diagnosis of AA [2,3]. Individual clinical symptoms and signs and laboratory investigations, specifically inflammatory markers, are sometimes not conclusive for definite diagnosis of AA. So, to increase the accuracy of diagnosis of AA, clinical scores can be helpful to triage patients into groups, those who can be safely discharged from hospital, those who require admission for follow up or further investigations, and those who

Table 4 Demographic data of the studied patients and histopathological results

Demographic data	*N* (%)	Phlegmonous AA *N*=82	Advanced AA *N*=18	*P* value
Age				
Range	18–80	-	-	-
Mean±SD	29.49±11.45	-	-	-
Median	26.50 (22.0–32.50)	-	-	-
Sex				
Male	60 (60%)	50 (83.3%)	10 (16.7%)	0.671
Female	40 (40%)	32 (80%)	8 (20%)	
BMI (kg/m^2)				
Range	20.0–39.0	-	-	-
Mean±SD	26.73±5.56	-	-	-
Median	26.0 (22.0–31.0)	-	-	-
$<$30	72 (72%)	62 (86.1%)	10 (13.9%)	0.086
$\geq$30	28 (28%)	20 (71.4%)	8 (28.6%)	

SD, Standard deviation; BMI, body mass index, statistically significant *P* less than or equal to 0.05.

Table 5 The relationship among the three different scores.

	Alvarado No. (%)	AIR No. (%)	AAS No. (%)	*P*
Total patients (*n*=100)				
Unlikely (Low)	9 (9.0)	29 (29.0)	20 (20.0)	<0.001*
Other	91 (91.0)	71 (71.0)	80 (80.0)	
Significance among scores		P_1<0.001*,P_2=0.016*,P_3=0.048*		
Male (*n*=60)				
Unlikely (Low)	7 (11.7)	17 (28.3)	8 (13.3)	<0.001*
Other	53 (88.3)	43 (71.7)	52 (86.7)	
Significance among scores		P_1<0.001*,P_2=0.712,P_3=0.001*		
Female (*n*=40)				
Unlikely (Low)	2 (5.0)	12 (30.0)	12 (30.0)	0.007*
Other	38 (95.0)	28 (70.0)	28 (70.0)	
Significance among scores		P_1=0.006*,P_2=0.006*,P_3=1.000		
BMI (<30) (*n*=72)				
Unlikely (Low)	9 (12.5)	25 (34.7)	16 (22.2)	<0.001*
Other	63 (87.5)	47 (65.3)	56 (77.8)	
Significance among scores		P_1<0.001*,P_2=0.086, P_3=0.027*		
BMI (≥30) (*n*=28)				
Unlikely (Low)	0	4 (14.3)	4 (14.3)	0.069
Other	28 (100.0)	24 (85.7)	24 (85.7)	

P: *P* value for associated among different scores. *P*1: *P* value for associated between **Alvarado and AIR scores**. *P*2: *P* value for associated between **Alvarado score and AAS**.*P*3: *P* value for associated between **AIR score and AAS**.* Statistically significant *P* less than or equal to 0.05.

can be moved to operating room immediately for urgent appendectomy [9].

Our study aimed to compare the diagnostic accuracy of three scores namely, Alvarado score, AIR score, and AAS in 100 adult patients with provisional diagnosis of AA by comparing the scores' results in every patient with the histopathological results of the excised appendices.

In our study, Alvarado score had a sensitivity of 91% at cut-off value greater than 4 value and 48% at cut-off value greater than 6 points. At a cut-off greater than 4, the sensitivity was higher in females compared with males (91% versus 88.3%, *P* value less than 0.027). These findings are consistent with the findings reported by Kim and colleagues and Bouali and colleagues who reported sensitivity of 88% and 94.9% at a cut-off value greater than 4, respectively and Sanjive and Ramaiah who reported sensitivity of 52.9% at cut-off value greater than 6 [10–12]. However, our results disagreed with that of Noor and colleagues who reported higher sensitivity of 68.15% at the cut-off value of greater than 6 points and Umar and colleagues who reported much higher sensitivity of 90.3% at cut-off value greater than 6 [13,14]. This discrepancy among our results and the results reported by latter authors can be explained by the inclusion of many subjective variables as anorexia, nausea/vomiting, and tenderness in Alvarado score.

When it comes to AIR score, there is no literature studied the validity of AIR score alone but in comparison with other scores. Only the authors who suggested this score studied its validity alone [15]. Our study had AIR sensitivity of 71% at a cut-off value greater than 4 points, which was equal in both sexes ($P < 0.099$) and that was quite lower than their reported sensitivity (96%) at the same cut-off value. This variation is probably due to many factors as including pediatric patients, much larger population number and finally, the different methodology (low probability patients were not operated upon). This methodology was also the same in their original description of the score with similar outcome to their validation study [7].

Concerning comparison between Alvarado score and AIR score, our study found that Alvarado score had higher sensitivity than AIR score (91% vs. 71%, respectively). This finding was consistent with that of Farahbakhsh and colleagues who reported sensitivity of 92% versus 77%, Gope and colleagues (98.33% vs. 81%), and Sartelli and colleagues (89.8% vs. 83.4%) [16–18]. While Hassan and colleagues reported higher sensitivity for AIR score than Alvarado score (77.97% vs. 67.8%, respectively), because they calculated the Alvarado sensitivity at a higher cut-off value (>5) than that used in our study [19]. Also, the inclusion of subjective variables by both scores namely right iliac fossa pain and tenderness, anorexia, nausea/vomiting may explain this difference.

Moreover, in AIR score, rebound tenderness is evaluated from 1 to 3 according to its severity which could be a subject to variation among different patients and its interpretation could vary among different surgeons.

Regarding AAS in our study, its sensitivity was 80% at a cut-off value greater than 10 points, 42% at a cut-off value greater than 15 points and 20% false negative rate. AAS was more sensitive in males (86.3%) compared with females (70%) and the difference was statistically significant ($P = 0.041$). The original authors of this score, Sammalkorpi and colleagues [8] reported a sensitivity of 95.9% and 58% for AAS at cut-off values of greater than 10 and greater than 15 points, respectively. In another validation study on a larger population, the authors reported 94.7% and 49.4% sensitivity at cut-off values greater than 10 and greater than 15, respectively which was nearby their previous results [20]. In both studies, the sensitivity of the tests was to somewhat higher than ours at both cut-off values. AAS, like the two other scores, includes many subjective variables that could vary among different patients and different assessors namely pain, migration, tenderness, and guard. The last variable also is awarded 2 or 4 points according to the severity of guarding which clearly increases the subjectivity of the score. Other drawbacks of their studies were that the decision maker was not blinded to score results and low probability patients were not operated upon. Finally, their higher sample size could contribute to the variable outcomes among the studies.

In terms of overall sensitivity and false negative rate in our study, Alvarado score had higher sensitivity than AAS (91% versus 80%) but without statistically significant difference. AIR score had the lowest sensitivity (71%) which achieved statistically significant difference in relation to both Alvarado score and AAS. In a similar study comparing the three scores, Elshakhs and colleagues [21] reported a sensitivity of 97.1%, 100%, and 94.1% for Alvarado score, AIR score, and AAS, respectively. Compared with our findings, the sensitivity values of their scores were higher than ours. Again, Elshakhs and colleagues did not operate mild cases of AA who responded to nonoperative management [22]. Moreover, the subjective nature of some variables in the three clinical scores could have also led to different outcomes.

In the current study, the diagnosis of AA was histologically proven in 100% of the study population. It is clear that this outcome is an odd one that, to the best of our knowledge, had never been reported in any other similar study. This outcome means that there was no negative appendectomy. Unfortunately, we could not calculate some outcomes namely, the specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy and the discriminatory value of the scores using the ROC curve. Indeed, these shortcomings had greatly undermined the usefulness of this study.

Conclusion

We concluded that Alvarado score was the best performing score regarding overall sensitivity and false negative rate. At a cut-off value greater than 4, Alvarado's score had higher sensitivity than AAS at a cut-off value greater than 10 (91% vs. 80%) but without statistically significant difference. At a cut-off value greater than 4, the AIR score had the lowest sensitivity (71%) which achieved statistically significant difference in relation to both the Alvarado score and AAS (P value < 0.001).

Alvarado score was more sensitive when used for females compared with males (91% vs. 88.3%, P value < 0.027), and AAS was more sensitive when used for males compared with females (86.3% vs. 70%, P value $= 0.041$) while AIR score had equal sensitivity in both sexes.

We recommended that a standardized method is needed to validate risk scores properly with establishment of optimal cut-off points for each score. Clinical scores evaluation should also focus on specific age, gender, and BMI groups as many literatures included both pediatric and adult patients. Also, scoring systems utilizing only objective variables are needed to avoid personal variations and potential differences inherent to subjective variables.

Financial support and sponsorship

Nil.

Conflicts of interest

No conflicts of interest to declare.

References

1 Ferri FF. Appendicitis, Acute. In: XX XX. Ferri's Clinical Advisor: Instant diagnosis and treatment by Ferri FF. 1st ed. XX: Elsevier Health Sciences; 2009. 284–287

2 Flum DR, Morris A, Koepsell T, Dellinger EP. Has misdiagnosis of appendicitis decreased over time?: a population-based analysis. JAMA 2001; 286:1748–1753.

3 Jones K, Peña AA, Dunn EL, Nadalo L, Mangram AJ. Are negative appendectomies still acceptable?. Am J Surg 2004; 188:748–754.

4 Brown TW, McCarthy ML, Kelen GD, Levy F. An epidemiologic study of closed emergency department malpractice claims in a national database of physician malpractice insurers. Acad Emerg Med 2010; 17:553–560.

5 Mock K, Lu Y, Friedlander S, Kim DY, Lee SL. Misdiagnosing adult appendicitis: clinical, cost, and socioeconomic implications of negative appendectomy. Am J Surg 2016; 212:1076–1082.

6 Alvarado A. A practical score for the early diagnosis of acute appendicitis. Ann Emerg Med 1986; 15:557–564.

7 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; 32:1843–1849.

8 Sammalkorpi HE, Mentula P, Leppäniemi A. A new adult appendicitis score improves diagnostic accuracy of acute appendicitis-a prospective study. BMC Gastroenterol 2014; 14:1–7.

9 Di Saverio S, Podda M, De Simone B, Ceresoli M, Augustin G, Gori A, *et al.* Diagnosis and treatment of acute appendicitis: 2020 update of the WSES Jerusalem guidelines. World J Emerg Med 2020; 15:1–42.

10 Kim K, Rhee JE, Lee CC, Kim KS, Shin JH, Kwak MJ, *et al.* Impact of helical computed tomography in clinically evident appendicitis. Emerg Med J 2008; 25:477–481.

11 Bouali M, El Berni Y, Moufakkir A, El Bakouri A, El Hattabi K, Bensardi F, *et al.* Value of Alvarado scoring system in diagnosis of acute appendicitis. Ann Med Surg 2022; 77:1–3.

12 Sanjive JG, Ramaiah RH. Comparison of RIPASA and Alvarado scoring in the diagnosis of acute appendicitis and validation of RIPASA scoring. Int Surg J 2019; 6:935–939.

13 Noor S, Wahab A, Afridi G, Ullah K. Comparing RIPASA score and Alvarado score in an accurate diagnosis of acute appendicitis. J Ayub Med Coll Abbottabad 2020; 32:38–41.

14 Umar MM, Abubakar IU, Agbo SP. Comparative study of alvarado score and its modifications in the preoperative diagnosis of acute appendicitis at a tertiary center in Sokoto, Nigeria. Nigerian J Surg 2020; 26:16–21.

15 Andersson M, Kolodziej B, Andersson RE. Validation of the Appendicitis Inflammatory Response (AIR) score. World J Surg 2021; 45:2081–2091.

16 Farahbakhsh F, Torabi M, Mirzaee M. A comparative study on the diagnostic validity of three scoring systems in the diagnosis of acute appendicitis in emergency centres. African J Emerg Med 2020; 10: 132–135.

17 Gope D, Dnayanmote AS, Thakkar SM, Tulsian AR, Kutty SA, Ranka M, *et al.* Comparison between AIR score and Alvarado score in cases of non-perforated and perforated acute appendicitis. Int Surg J 2019; 6:1108–1114.

18 Sartelli M, Baiocchi GL, Di Saverio S, Ferrara F, Labricciosa FM, Ansaloni L, *et al.* Prospective observational study on acute appendicitis worldwide (POSAW). World J Emerg Surg 2018; 13:1–10.

19 Hassan M, Jeilani M, Saad AA, Iqbal S, Boshnaq M. Evaluation of Alvarado score and appendicitis inflammatory response score as diagnostic tools for acute appendicitis. Int Surg J 2022; 9:1937–1942.

20 Sammalkorpi H, Mentula P, Savolainen H, Leppäniemi A. The introduction of adult appendicitis score reduced negative appendectomy rate. Scand J Surg 2017; 106:196–201.

21 Elshakhs S, Abdelsamie M, Fareed A, Abuomar MH. Reliability of the adult appendicitis score in diagnosing acute appendicitis. Menofia Med J 2019; 32:544–548.