Sawsan Hanafi et al. Saudi J Er Med. 2025;6(1):266-271
DOI: 10.24911/SJEMed.72-1715681107

ORIGINAL ARTICLE

Comparative diagnostic accuracy of appendicitis inflammatory response and Alvarado scores in adult appendicitis

Sawsan Hanafi[1,2,3], Noura Alshareef[1,2], Wesam Edrees[1,2], Jamil M. Baljoon[2,4*], Faris Salah Abunar[2,4]

ABSTRACT

Background: Acute appendicitis is a common surgical emergency, with diagnostic challenges that can lead to unnecessary surgeries. This study aims to compare the diagnostic accuracy of the appendicitis inflammatory response (AIR) score and the Alvarado score in a clinical setting, examining the influence of demographic variables on their performance.

Methods: A retrospective analysis was conducted on 166 patients at National Guard Health Affairs Hospital, Saudi Arabia, from May 2016 to April 2022. Patients were categorized into two groups: 83 with appendicitis and 83 with acute abdominal symptoms but without appendicitis. We evaluated the diagnostic accuracy of the AIR and Alvarado scores through risk probability of developing appendicitis, sensitivity, and specificity. We also examined the effects of age and gender on the scores' accuracy.

Results: The AIR score demonstrated higher overall sensitivity (95.7%) and specificity (90.5%) than the Alvarado score, which showed a sensitivity of 87.3% and a specificity of 52.4%. The AIR score was particularly effective in identifying patients with a high probability of appendicitis. Subgroup analysis revealed significant demographic influences, with men showing higher incidence rates of appendicitis and age impacting the scoring accuracy than women.

Conclusion: Our study's findings advocate for the AIR score's expanded use in clinical practice to improve diagnostic accuracy and patient management tailored to demographic variations. Further validation in varied clinical settings and larger cohort and case-control researches are recommended.

Keywords: Appendicitis, diagnosis, AIR score, Alvarado score.

Introduction

Acute appendicitis is one of the more common causes of acute abdominal pain in both adults and children who visit emergency departments (ED), with a lifetime risk of 8.6% for males and 6.7% for females [1,2]. It primarily occurs when patients are in adolescence and or their twenties, with a male-to-female ratio of 3:22 [3]. The most frequent symptom of acute appendicitis is abdominal pain, typically migrating from the periumbilical or epigastric area to the right iliac fossa in approximately 75% of patients [4]. This condition is often accompanied by abdominal rigidity, nausea, vomiting, and fever. Acute appendicitis remains a significant cause of morbidity, particularly when diagnosis is delayed, leading to advanced stages such as diffuse peritonitis [3].

Despite its prevalence, diagnosing acute appendicitis in the ED remains challenging [5,6], as evidenced by high rates of unnecessary appendectomies. Numerous clinical and imaging tools have been developed to assist in diagnosis [7]. Imaging techniques such as ultrasonography (US) and computed tomography (CT) are crucial for definitive diagnosis [8]. However, the operator dependency on the US and the radiation risks from CT are notable

Correspondence to: Jamil M. Baljoon
*King Abdullah International Medical Research Center, Jeddah, Saudi Arabia.
Email: jbaljoon@gmail.com
Full list of author information is available at the end of the article.
Received: 09 May 2024 | **Accepted**: 06 July 2024

drawbacks; additionally, delays in reporting can impede timely management [8,9]. Conversely, clinical decision rules and scoring systems offer objective, non-invasive, and straightforward assessment methods [10]. Prompt and accurate diagnosis of acute appendicitis is critical for deciding between early surgical intervention and conservative management to minimize complications and unnecessary surgeries [8,11].

Various clinical scoring systems have been developed to facilitate the diagnosis of acute appendicitis [5]. These systems leverage complete medical histories, thorough clinical examinations, and basic laboratory tests [4,9]. Although the Alvarado scoring system, introduced in 1986, has been widely used for decades, it has several drawbacks [8,12]. The newer appendicitis inflammatory response (AIR) scoring system addresses these limitations [5,12]. The AIR system places less emphasis on subjective symptoms and more on clinical signs. It includes the C-reactive protein (CRP) test, which has shown strong correlations with acute appendicitis in numerous studies [7]. This approach makes the AIR scoring system preferable to the Alvarado system, which relies heavily on subjective criteria and can be particularly challenging to apply to children, leading to potential misdiagnoses [4,8]. Therefore, clinical scoring systems are crucial for diagnosing acute appendicitis, especially in resource-limited settings [7].

Acute appendicitis remains a challenging diagnosis in patients presenting to EDs with acute abdominal pain [8,13]. A rapid, reliable method for diagnosing acute appendicitis is useful to physicians, particularly in ED settings. An ideal clinical scoring system should expedite and enhance diagnostic accuracy while reducing the need for invasive tests [8]. Given the scarcity of studies led by emergency physicians on the diagnostic accuracy of clinical scoring systems for acute appendicitis globally, and particularly the lack of research in Saudi Arabia, this study aims to evaluate the diagnostic accuracy of the AIR scoring system compared to the Alvarado system against CT findings in patients presenting to ED with acute abdominal pain and suspected acute appendicitis [9].

Subjects and Methods

Study design and setting

We conducted a retrospective case control at National Guard Health Affairs (NGHAs), a tertiary hospital located at King Abdulaziz Medical City in Jeddah, Saudi Arabia, from May 2015 to April 2022.

Participant selection and inclusion/exclusion criteria

The inclusion criteria consisted of the following: Patients with suspected appendicitis with abdominal pain who had a CT scan or appendectomy were included in the study. Those with positive appendicitis were compared with those without appendicitis. Then, the test characteristics could be calculated for the two scores. Therefore, the study included two groups: patients diagnosed with appendicitis (83 patients) and a control group of patients who presented with acute abdomen but did not receive a diagnosis of appendicitis (83 patients)

We confirmed appendicitis diagnoses through CT or appendectomy as aforementioned. We excluded pediatric patients aged 14 years or younger and patients with incomplete data, such as missing CRP tests.

Study design, measurements, and outcome variables

Using the Roasoft, Inc. website, we calculated the required sample size with a 95% confidence interval (CI), a 5% margin of error, and an estimated population of 73 patients with suspected appendicitis, following the guidelines by Hassan et al. [14,15]. The initial calculation suggested a sample size of 62; however, we expanded this to 166 patients using a non-random consecutive sampling method to meet inclusion and exclusion criteria and to increase the representation of the sample. Data were collected from the hospital's electronic records using the Best Care 2.0 System, and data entry was restricted to authors and designated data collectors. The primary outcomes measured were the sensitivity and specificity of the AIR and Alvarado scoring systems in diagnosing

Table 1. *Alvarado score.*

Diagnosis		Alvarado score
Signs	RLQ tenderness	2
	Temperature > 37.3°C	1
	Rebound tenderness	1
Symptoms	Migration of pain to RLQ	1
	Nausea or vomiting	1
	Anorexia	1
Laboratory values	Leukocytosis left shift	1
	Leukocytosis > 10	2
	Total	10

Abbreviation: RLQ, right lower quadrant.

Table 2. *AIR score values.*

Diagnosis		AIR score
RLQ pain		1
Temperature ≥ 38.5°C		1
Rebound tenderness†		2
Vomiting		1
Polymorphonuclear leukocytes	70%-84%	1
	≥85%	2
WBC count	10-14.9 × 10.000	1
	≥ 15 × 10.000	2
CRP	10-49	
	≥ 50	

Abbreviation: CRP, C-reactive protein; RLQ, right lower quadrant.

†Rebound tenderness in AIR score is typically divided into light, medium, and strong with a score of 1, 2, and 3, respectively. However, this subdivision of rebound tenderness could not be obtained from the data. Thus, all patients with rebound tenderness were given a score of 2.

appendicitis. Recorded patient characteristics included age, gender, and scores from the AIR and Alvarado scoring systems, as detailed in Tables 1 and 2. All data were entered into an Excel file to ensure confidentiality.

Data Analysis

We analyzed the data using JMP Pro 15.2.0 software (SAS Institute Inc., Cary, NC, USA). Descriptive statistics included age and gender distributions. Inferential statistics for the Alvarado score included right lower quadrant (RLQ) tenderness, temperature ≥ 37.3°C, rebound tenderness, migration of pain to the RLQ, anorexia, nausea/vomiting, leukocytosis > 10,000 cells/µl, and leukocyte left shift. For the AIR, the components analyzed included vomiting, RLQ pain, rebound tenderness, temperature ≥ 38.5°C, polymorphonuclear leukocytes, white blood cell (WBC) count, and CRP count. We presented differences in scores between the groups using means with a 95% CI and standard deviation (SD). Parametric tests included chi-square and student *T*-test, while non-parametric tests included the Mann–Whitne test.

Results

Demographic characteristics of study participants

Risk assessments using AIR and Alvarado scores

We assessed all 166 patients using the AIR score, finding that 59 patients (35.5%) had a low risk, 90 patients (54.2%) had an intermediate risk, and 17 patients (10.2%) had a high risk of developing appendicitis. Similarly, using the Alvarado score, 81 patients (48.8%) were at low risk, 33 patients (19.9%) were at moderate risk, and 52 patients (31.3%) were at high risk of developing appendicitis.

Table 3. *Comparison between mean age of patients who had appendicitis versus the control group.*

Variable	Patients with appendicitis (*n* = 83)	Control Group (*n* = 83)	*p*-value
Mean age (years)	29.3	46.1	< 0.0001
SD	15.7	18.6	
95% CI	25.6-33.0 years	42.42-49.8 years	

Abbreviations: SD, standard deviation; CI, confidence interval.

The statistical test was student *T*-test.

Table 4. *The means of AIR score and Alvarado score between patients with appendicitis versus the control group.*

Variable	Patients with Appendicitis (*n* = 83)	Control Group (*n* = 83)	*p*-value
Mean (SD) AIR score	6.36 (2.2)	3.1 (1.9)	< 0.0001
Mean (SD) Alvarado score	6.65 (2.1)	3.25 (1.9)	< 0.0001

Abbreviations: AIR, Appendicitis Inflammatory Response; SD, standard deviation.

The statistical test was student *T*-test.

Demographic characteristics and analysis of study participants

A total of 166 patients visiting NGHA hospital between May 2016 and April 2022 participated in the study. We divided these patients into two groups: 83 diagnosed with appendicitis and 83 presented with acute abdomen but were not diagnosed with appendicitis.

The study comprised 82 male participants (49.4%) and 84 female participants (50.6%), with an overall mean age of 37.7 years (95% CI = 34.8-40.6). Significant demographic differences existed between the groups: among those diagnosed with appendicitis, 53 (63.9%) were males and 30 (36.1%) were females, whereas in the control group, 29 (34.9%) were males and 54 (64.3%) were females. The mean age of patients diagnosed with appendicitis was 29.3 years (95% CI = 25.6-33.0), compared to 46.1 years (95% CI = 42.4-49.8) in the control group, as depicted in Table 3.

AIR Score Outcomes

The mean AIR score for patients diagnosed with appendicitis was 6.36 (95% CI = 5.92-6.81), while for those in the control group, it was 3.1 (95% CI = 2.65-3.54), as in Table 4. Significant differences were noted in categorical AIR scores: in the appendicitis group, 10 (12.1%) had a low probability, 56 (67.5%) had an intermediate probability, and 17 (20.5%) had a high probability of developing appendicitis. Conversely, in the control group, 49 (59%) had a low probability and 34 (41%) an intermediate probability; no control group patients had a high probability of developing appendicitis, as illustrated in Figure 1.

Sensitivity and Specificity of AIR and Alvarado Scores

When considering an AIR score of ≥6 as positive, sensitivity was 68.67% and specificity 85.54%. For an AIR score of ≥8, sensitivity decreased to 32.53%, but specificity reached 100%, as shown in Table 5. For the Alvarado score, when ≥6 was considered positive, sensitivity was 69.88% and specificity 87.95%. If ≥8 was considered positive, sensitivity was 39.76% and specificity 97.59%.

Alvarado Score Outcomes

The mean Alvarado score for appendicitis patients was 6.65 (95% CI = 6.21-7.09), while it was 3.25 (95% CI = 2.81-3.69) for the control group, as in Table 4. A significant difference was also observed in categorical Alvarado scores as depicted in Figure 2–4 when compared to AIR score: 46 (55.4%) appendicitis patients and only 6 (7.2%) control group patients had a high probability of developing appendicitis

Subgroup Analysis of AIR Score Components

Significant differences were found in vomiting, RLQ pain, rebound tenderness, polymorphonuclear leukocytes, and WBC count. RLQ pain was present in 82 (98.8%) appendicitis patients and 40 (48.2%) control

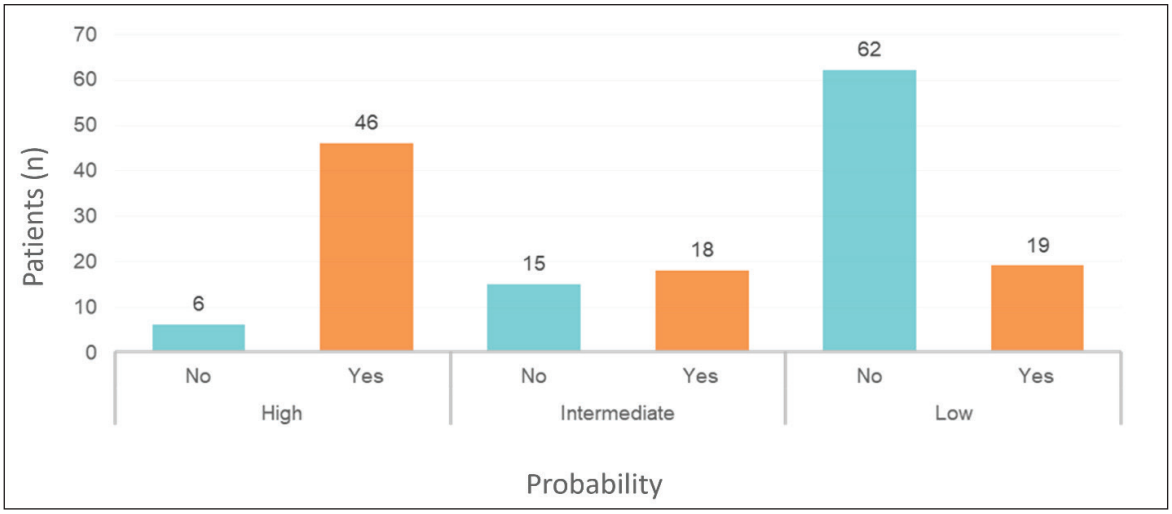

Figure 1. *Low probability, intermediate probability, and high probability of developing appendicitis for Alvarado score ($p < 0.0001$).*
Low risk of appendicitis: 0-4. Intermediate risk of appendicitis: 5-6. High risk of appendicitis: 7-10. The statistical test used was chi-square test.

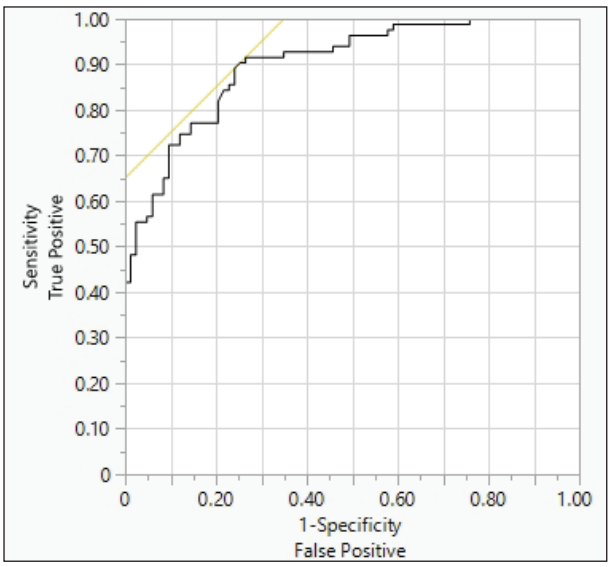

Figure 2. *Receiver operating characteristic (ROC) curve for Alvarado score with the following factors: Age, gender, and Alvarado score.*
AUC (Area Under Curve) = 0.9002

group patients. Rebound tenderness occurred in 63 (75.9%) appendicitis patients and 10 (12.1%) control group patients. Sixty (72.3%) appendicitis patients and 40 (48.2%) control group patients experienced vomiting. For WBC counts, 28 (33.7%) appendicitis patients had counts ≥ 15 × 10^9/l, compared to 6 (7.2%) in the control group.

Subgroup Analysis of Alvarado Components

Significant differences were found in RLQ tenderness, WBC count > 10 × 10^9/l, rebound tenderness, migration of pain to RLQ, vomiting, and leukocyte left shift among the appendicitis patients. RLQ tenderness was observed in 82 (98.8%), WBC > 10 × 10^9/l in 55 (66.7%), and rebound tenderness in 62 (74.7%) of those diagnosed with appendicitis. The two groups had no significant differences concerning fever and anorexia factors.

Discussion

This study examined the diagnostic accuracy of the AIR score and the Alvarado score in a cohort of 166 patients with acute appendicitis. Our results demonstrate the AIR score's superior diagnostic precision over the Alvarado score, echoing findings from previous studies [4,13,15,16]. Although the sensitivity and specificity of both scores were similar, significant differences emerged when categorizing the scores into low, intermediate, and high probabilities of developing appendicitis.

For instance, Madasi's 2016 study highlighted the AIR score's superior predictive validity, with an area under the ROC curve of 0.95 compared to 0.74 for the Alvarado score [13]. This was accompanied by a sensitivity of 95.7% and a specificity of 90.5% for the AIR score, markedly higher than the Alvarado score's sensitivity of 87.3% and specificity of 52.4%, reinforcing the AIR score's diagnostic capabilities. Our study aligns with these observations, confirming the AIR score's higher sensitivity and specificity in accurately diagnosing acute appendicitis since the ROC curve for AIR and Alvarado scores were 0.905 and 0.9002, respectively.

Additionally, Hassan et al. [15] reported sensitivities of 77.97% and specificities of 85.71% for the AIR score, outperforming the Alvarado score, which supports our findings of the AIR score's effectiveness in reducing unnecessary appendectomies. Noor et al. [16] reported that CT scans showed higher sensitivity and specificity in patients with equivocal clinical scores (≥4-≤6) [16]. This supports our results, highlighting the AIR score's critical role in diagnosing ambiguous cases. Karki and Hazra's research supported the AIR score as a superior diagnostic tool, with sensitivity levels closely matching our findings, further validating its efficacy in clinical settings [4].

Table 5. *Sensitivity, specificity, and value for AIR and Alvarado scores (prevalence of 6.7% in the general population).*

Diagnostic value	AIR score		Alvarado score	
	≥ 6	≥ 8	≥ 6	≥ 8
Sensitivity	68.67%	32.53%	69.88%	39.76%
Specificity	85.54%	100.00%	87.95%	97.59%

Abbreviations: AIR, Appendicitis Inflammatory Response.

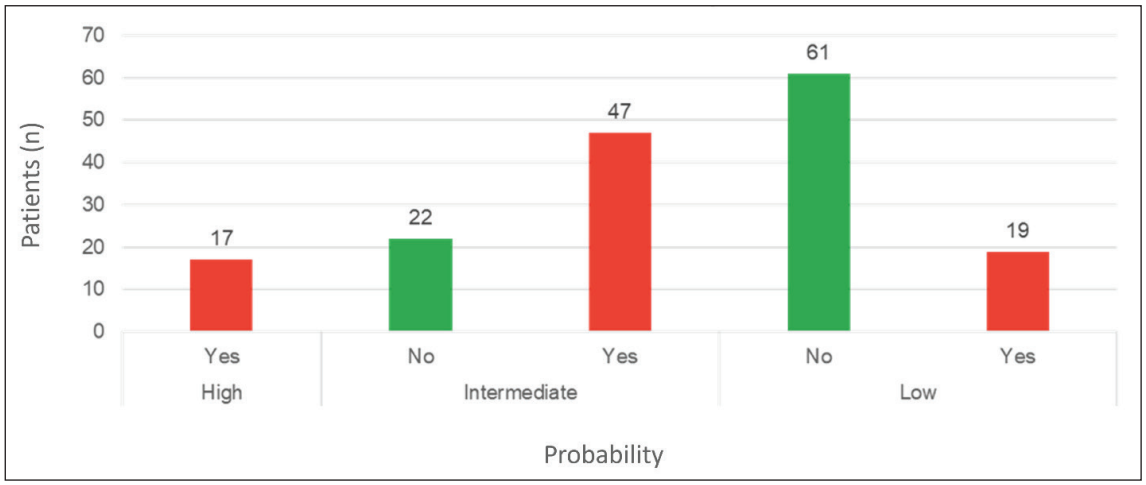

Figure 3. *AIR score and low, intermediate, and high probability of developing appendicitis (p < 0.0001)*
Abbreviation: AIR, Appendicitis Inflammatory Response. Low risk of appendicitis: 0-4. Intermediate risk of appendicitis: 5-8. High risk of appendicitis: 9-12. The statistical test used was chi-square test.

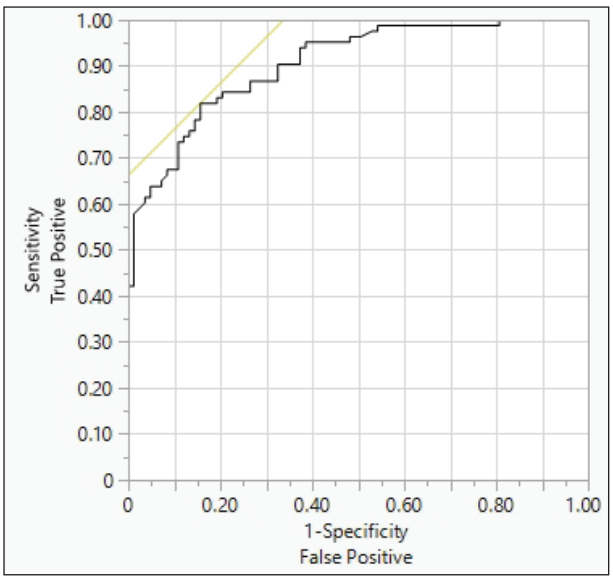

Figure 4. *ROC curve for AIR score with the following factors: Age, gender, and AIR score.*
AUC (Area Under Curve) = 0.9045.

Our study explored age and gender differences in the diagnostic accuracies of the AIR and Alvarado scores. We found a notable gender disparity, with male patients showing a higher incidence of acute appendicitis than female patients. This suggests a need for potential gender-specific adjustments when applying these scores. Moreover, age emerged as a significant factor affecting the scores' sensitivity and specificity, indicating that younger patients might exhibit different symptoms or response patterns, impacting score accuracy.

Our study confirms that the AIR score has higher sensitivity and specificity than the Alvarado score. It also suggests that age and gender affect diagnostic accuracy. Supported by evidence from Madasi [13], Hassan et al. [15], Noori et al. [16], and Karki and Hazra [4], we recommend using the AIR score more broadly in clinical practice to improve diagnosis and patient care. Our findings highlight the need for diagnostic strategies that consider demographic factors in clinical assessments and decision-making.

This study has several limitations that should be considered when interpreting the results. First, its retrospective design may introduce biases related to data collection and selection, as it relies on the accuracy and completeness of medical records. Second, the study was conducted at a single tertiary hospital, which might limit the generalizability of the findings to other settings with different patient populations or healthcare practices. Additionally, the equal division of participants into groups with and without appendicitis may not reflect the true prevalence of the condition, potentially skewing the diagnostic performance metrics of the scoring systems. Lastly, while the study examined demographic factors such as age and gender, other potential confounders like underlying comorbidities and the severity of presentation were not thoroughly analyzed, which could influence the diagnostic accuracy of the AIR and Alvarado scores.

In conclusion, our study and the literature suggest that the AIR score is highly effective for diagnosing appendicitis and highlights how age and gender influence diagnostic outcomes. We recommend a diagnostic strategy that

combines clinical assessments with selective imaging to improve the understanding and use of these tools. This approach is supported by evidence from various studies [1-4,13,15,16], which emphasize the AIR score's role in enhancing diagnostic accuracy and reducing unnecessary procedures. Our findings suggest that further research is needed to confirm these tools' effectiveness in different clinical environments.

List of Abbreviations

AIR	Appendicitis inflammatory response
CI	Confidence interval
CRP	C-reactive protein
CT	Computed tomography
ED	Emergency department
IRB	Institutional review board
NGHA	National Guard Health Affairs
RLQ	Right lower quadrant
ROC	Receiver operating characteristic
US	Ultrasound
WBC	White blood cells

Conflict of interests

The authors declare no conflicts of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Consent to participate

The study design of this paper was a retrospective case control study. Thus, no patient consent was required.

Consent for publication

All authors consent to the publication of this manuscript.

Ethical approval

The Institutional Review Board (IRB) of King Abdullah International Medical Research Center approved this research under IRB approval number IRB/2911/23 and study number NRJ23J/227/09.

Author details

Sawsan Hanafi[1,2,3], Noura Alshareef[1,2], Wesam Edrees[1,2], Jamil M. Baljoon[2,4], Faris Salah Abunar[2,4]

1. Emergency Department, Ministry of National Guard-Health Affairs, Jeddah, Saudi Arabia
2. King Abdullah International Medical Research Center, Jeddah, Saudi Arabia
3. Quality Management Department, Ministry of National Guard-Health Affairs, Riyadh, Saudi Arabia
4. College of Medicine, King Saud Bin Abdulaziz University for Health Sciences, Jeddah, Saudi Arabia

References

1. Karami MY, Niakan H, Zadebagheri N, Mardani P, Shayan Z, Deilami I. Which one is better? comparison of the acute inflammatory response, Raja Isteri Pengiran Anak Saleha Appendicitis and Alvarado scoring systems. Ann Coloproctol. 2017;33(6):227–31. https://doi.org/10.3393/ac.2017.33.6.227
2. Von-Mühlen B, Franzon O, Beduschi MG, Kruel N, Lupselo D. AIR score assessment for acute appendicitis. Arq Bras Cir Dig. 2015;28(3):171–3. https://doi.org/10.1590/S0102-67202015000300006
3. Snyder MJ, Guthrie M, Cagle S. Acute appendicitis: efficient diagnosis and management. Am Fam Physician. 2018;98(1):25–33.
4. Karki OB, Hazra NK. Evaluation of the appendicitis inflammatory response score against Alvarado score in diagnosis of acute appendicitis. Kathmandu Univ Med J. 2020;18(70):171–5.
5. Kollár D, McCartan DP, Bourke M, Cross KS, Dowdall J. Predicting acute appendicitis? A comparison of the Alvarado score, the appendicitis inflammatory response score and clinical assessment. World J Surg. 2015;39(1):104–9. https://doi.org/10.1007/s00268-014-2794-6
6. Jose T, Rajesh PS. Appendicitis inflammatory response score in comparison to Alvarado score in acute appendicitis. Surg J (N Y). 2021;7(3):e127–31. https://doi.org/10.1055/s-0041-1731446
7. Chisthi MM, Surendran A, Narayanan JT. RIPASA and air scoring systems are superior to alvarado scoring in acute appendicitis: diagnostic accuracy study. Ann Med Surg (Lond). 2020;59:138–42. https://doi.org/10.1016/j.amsu.2020.09.029
8. Althoubaity FK. Suspected acute appendicitis in female patients. Trends in diagnosis in emergency department in a University Hospital in Western region of Saudi Arabia. Saudi Med J. 2006;27(11):1667–73.
9. Díaz-Barrientos CZ, Aquino-González A, Heredia-Montaño M, Navarro-Tovar F, Pineda-Espinosa MA, Espinosa de Santillana IA. The RIPASA score for the diagnosis of acute appendicitis: a comparison with the modified Alvarado score. Escala RIPASA para el diagnóstico de apendicitis aguda: comparación con la escala de Alvarado modificada. Rev Gastroenterol Mex (Engl Ed). 2018;83(2):112–6. https://doi.org/10.1016/j.rgmx.2017.06.002
10. Awayshih MMA, Nofal MN, Yousef AJ. Evaluation of Alvarado score in diagnosing acute appendicitis. Pan Afr Med J. 2019;34:15. https://doi.org/10.11604/pamj.2019.34.15.17803
11. de Castro SM, Ünlü C, Steller EP, van Wagensveld BA, Vrouenraets BC. Evaluation of the appendicitis inflammatory response score for patients with acute appendicitis. World J Surg. 2012;36(7):1540–5. https://doi.org/10.1007/s00268-012-1521-4
12. 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(1):38–41.
13. Madasi V. Comparison of predictive validity of Alvarado score and appendicitis inflammatory response (AIR) score, a hospital based observational study. Int J surg Orthopedics. 2016;2(3):29–34.
14. Sample Size Calculator by Raosoft, Inc. [Internet]. 2024 [cited 17 April 2024]. Available from: http://www.raosoft.com/samplesize.html
15. 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 Nov 28;9(12):1937–42.
16. Noori IF, Jabbar AS, Noori AF. Clinical scores (Alvarado and AIR scores) versus imaging (ultrasound and CT scan) in the diagnosis of equivocal cases of acute appendicitis: a randomized controlled study. Ann Med Surg (Lond). 2023;85(4):676–83.