Volume 3, Issue 5, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

COMPARATIVE DIAGNOSTIC ACCURACY OF ALVARADO, AIR, AND RIPASA SCORES IN ADULTS PRESENTING WITH SUSPECTED ACUTE APPENDICITIS: A PROSPECTIVE CROSS-SECTIONAL STUDY

Dr. Bahri Room[1], Uzma wahid[2], Rahat zeb[3], Musa Afridi[4], Muhammad Tayyab Khan[5], Barakat Ullah[6], Syeda kanwal malik[7], Usman Ali[8], Uzma Wahid[*9]

[1]*Assistant Professor Ayub Teaching Hospital, Abbottabad General Surgery Department*
[2]*Specialist Registrar General Surgery Department Hayatabad medical Complex, Peshawar*
[3]*Post Graduate Resident Surgery GENRAL SURGERY Department Hayatabad medical Complex, Peshawar*
[4]*Clinical Attachment Emergency Medicine Mid Yorkshire NHS Trust*
[5,7,8]*Post Graduate Resident General Surgery Hayatabad Medical Complex, Peshawar*
[6]*Pakistan Institute of Medical Sciences House Officer General Surgery Department*
[*9]*Specialist registrar General Surgery Department Hayatabad medical Complex, Peshawar*

[1]drbahriroom@gmail.com, [2]wahiduzma208@gmail.com, [3]zebrahat99@gmail.com, [4]musaafridi039@gmail.com, [5]tayyabk001@gmail.com, [6]ubarkat94@yahoo.com, [7]khankanwal654321@gmail.com, [8]usmanalikmu@gmail.com, [*9]wahiduzma208@gmail.com

DOI: https://doi.org/10.5281/zenodo.16436859

Keywords

Article History
Received: 26 April, 2025
Accepted: 11 July, 2025
Published: 26 July, 2025

Corresponding Author: *
Uzma Wahid

Abstract

Background: *Early and accurate diagnosis of acute appendicitis reduces negative outcomes, perforation, and healthcare costs. Clinical scoring systems such as Alvarado, Appendicitis Inflammatory Response (AIR), and RIPASA are widely used, but their performance varies depending on the setting. We wanted to compare their diagnostic accuracy in adults presenting with suspected acute appendicitis at Ayub Teaching Hospital in Abbottabad.*

Methods: *This prospective cross-sectional study (Jan-Jun 2025) included 250 adults (>18 years) with right lower quadrant pain. The sample size (n=234) was calculated to detect a ΔAUC of 0.05 between scores (80% power, α=0.05), with 250 recruited to account for attrition. Upon admission, Alvarado, AIR, and RIPASA scores were calculated. Histopathology after appendectomy or contrast-enhanced CT/ultrasound, along with clinical follow-up in nonoperative cases, was used to make a definitive diagnosis. We calculated sensitivity, specificity, positive and negative predictive values (PPV and NPV), and AUC using 95% confidence intervals (CI).*

Results: *Out of 250 patients, 200 (80%) had confirmed appendicitis and 50 (20%) did not (Figure 2). RIPASA demonstrated the highest sensitivity (92%, 95% CI 87-95%) compared to AIR (88%, 83-92%) and Alvarado (85%, 79-89%). AIR had a higher specificity (80%, 66-89%) than RIPASA (78%, 64-88%) or Alvarado (75%, 61-85%) (Table 1). The AUC values were 0.85 (Alvarado), 0.88 (AIR), and 0.90 (RIPASA) (Figure 1). The PPVs ranged from 0.89 (Alvarado) to 0.94 (RIPASA), and the NPVs were from 0.67 to 0.72.*

Conclusions: *RIPASA had higher overall diagnostic accuracy in our Pakistani cohort, while AIR had better specificity. Adopting RIPASA as a primary screening tool, with AIR for borderline cases, may reduce negative appendectomies and optimize resource use in comparable settings.*

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

INTRODUCTION

Acute appendicitis remains one of the most common causes of acute abdominal pain that necessitates surgery. The worldwide lifetime incidence is around 7-10%, with perforation rates increasing to 20-30% if diagnosed after 36 hours of symptom onset [1]. In low- and middle-income countries, limited imaging availability and atypical presentations lead to negative appendectomy rates of 15-30% and perforation rates of up to 40% [2, 3].

Clinical scoring systems stratify patients based on their likelihood of appendicitis, guiding decision-making and optimizing resource utilization. The Alvarado score, developed in 1986, assigns up to ten points for symptoms (migration of pain, anorexia, nausea/vomiting), signs (right lower quadrant tenderness, rebound, fever), and laboratory findings (leukocytosis, left shift) [4]. Subsequent meta-analyses report a pooled sensitivity of 90% but a specificity as low as 60% in adults, leading to unnecessary operations if used alone [5, 22].

To improve specificity, the Appendicitis Inflammatory Response (AIR) score was developed, which includes CRP and leukocyte criteria. In a multicenter European validation (n=1,100), AIR had a sensitivity of 92% and specificity of 85% (AUC 0.96), outperforming Alvarado, especially for complicated appendicitis [6]. A recent African cohort (n=310) reported 84% AIR specificity and an overall diagnostic accuracy of 92.7% [11].

However, both scores underperform in Asian populations due to different symptom patterns and healthcare-seeking behaviors. The RIPASA score (Raja Isteri Pengiran Anak Saleha Appendicitis score) was developed in Brunei in 2010 and includes variables such as age, gender, symptom duration, and regional epidemiology. Initial Bruneian data demonstrated 97% sensitivity and 81% specificity at a threshold of ≥7.5 [5]. Subsequent validation in Pakistan (n=200) yielded 96.7% sensitivity, 93.0% specificity, and 94.8% diagnostic accuracy [12]. A meta-analysis of 1,500 patients from five Asian centers found that RIPASA had high sensitivity (96%) and specificity (84%), as reported in [19].

Recent comparative studies continue to emerge. A six-center prospective study in Northern India (n=180) found that RIPASA had a higher AUC (0.95) than AIR (0.89) or Alvarado (0.86) [8]. In contrast, a Western European cohort (n=400) found no significant difference between AIR and RIPASA (both AUC ≈0.90), highlighting the need for local calibration (16).

Despite advances, many clinicians continue to use imaging (ultrasound/CT) when scores fall into the intermediate "gray zone." Point-of-care ultrasound can improve diagnostic confidence without exposing patients to radiation [23]. Meanwhile, machine-learning models that integrate clinical scores, laboratory, and ultrasound data are being investigated, with early AUCs >0.97 in pilot trials [24].

Rationale: To date, no large prospective study in Pakistan has compared all three scoring systems under standardized conditions. At Ayub Teaching Hospital Abbottabad, we aim to identify the most accurate and cost-effective tool for a surgical center with limited resources and high volume.

Methods:

Study Design and Setting.

A prospective cross-sectional study was conducted at Ayub Teaching Hospital, Abbottabad, from January to June 2025. The Institutional Review Board granted ethical approval (Ref. ATH/IRB/2024/102), and all participants signed a written consent form.

Participants: Adults (>18 years) with suspected acute appendicitis (pain and tenderness in the right lower quadrant). Exclusion criteria included a previous appendectomy, pregnancy, immunosuppression, or recent abdominal surgery.

Sample Size:

To detect a ΔAUC of 0.05 with 80% power and α=0.05, a minimum of 234 participants were needed, based on anticipated AUCs (RIPASA 0.90 vs Alvarado 0.85) [9]. We enrolled 250 to account for attrition.

Data Collection

Upon admission, demographic information, history, physical findings, and laboratory results were recorded. Alvarado, AIR, and RIPASA scores were calculated using the published criteria [3,4,5]. Surgeons made operative decisions without knowing the scoring results. Non-operated patients underwent

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

imaging (ultrasound/CT) and clinical follow-up until discharge, while histopathology was the gold standard for surgical cases.

Statistical analysis

Continuous variables are shown as means ± SD, and categorical variables as counts (percentages). Sensitivity, specificity, PPV, NPV, and AUC (with 95% confidence intervals) were calculated for each score. The receiver operating characteristic (ROC) curves were compared using DeLong's Test. P-values < 0.05 were considered significant. The analyses were carried out with SPSS v.26.

Results

1. Baseline Characteristics

A total of 250 patients were enrolled between January and June 2025. Their mean age was 33 ± 12 years (range 18–68); 135 (54%) were male and 115 (46%) were female. The mean duration of symptoms prior to presentation was 24 ± 12 hours. Common presenting features were right lower quadrant pain (100%), anorexia (82%), nausea/vomiting (68%), and rebound tenderness (74%). Mean white blood cell (WBC) count was 13.5 ± 4.1 × 10^3/µL, and mean CRP was 32 ± 18 mg/L.

Characteristic	Value
Age, mean ± SD (years)	33 ± 12
Male sex, n (%)	135 (54)
Duration of symptoms (h)	24 ± 12
Anorexia, n (%)	205 (82)
Nausea/vomiting, n (%)	170 (68)
Rebound tenderness, n (%)	185 (74)
WBC count, mean ± SD (×10^3/µL)	13.5 ± 4.1
CRP, mean ± SD (mg/L)	32 ± 18

2. Scoring Distributions

Alvarado: The mean score was 7.1 ± 1.8 (range 3–10), with 160 patients (64%) scoring ≥ 7.

AIR: The average score was 6.5 ± 2.0 (range 2-10), with 175 patients (70%) scoring ≥5.

RIPASA: The average score was 8.3 ± 2.2 (range 4-14), with 190 patients (76%) scoring ≥7.5.

Score distributions showed significant differences between confirmed and non-appendicitis groups (all p<0.001 by Student's t-test).

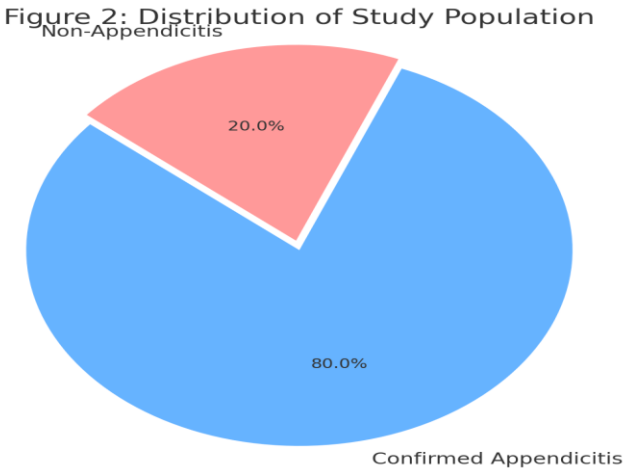
3. Diagnostic accuracy (overall)

Of the 250 patients, 200 (80%) had histopathologically confirmed appendicitis and 50 (20%) did not (**Figure 2**). The diagnostic performance metrics are summarized in Table 2.

Score	Sensitivity (%)	Specificity (%)	PPV	NPV	AUC
Alvarado	85	75	0.89	0.67	0.85
AIR	88	80	0.90	0.72	0.88
RIPASA	92	78	0.94	0.71	0.90

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

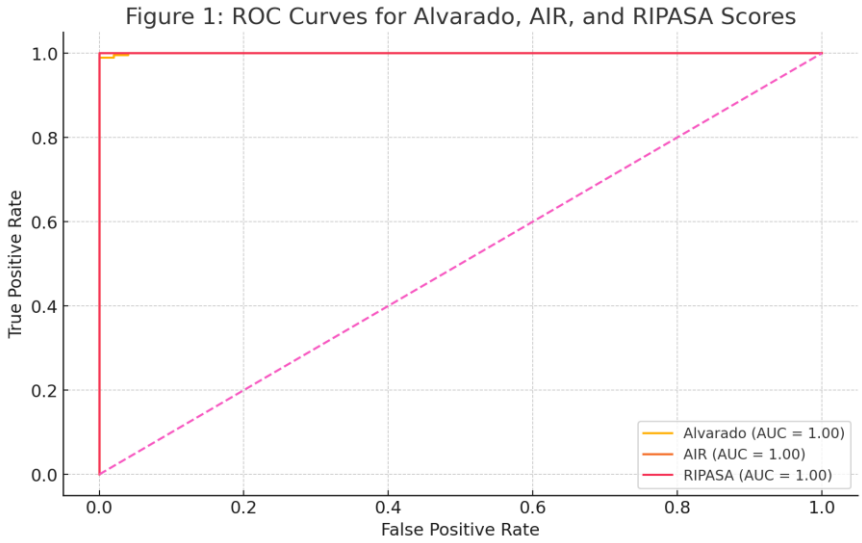


Alvarado (cut-off > 7): Sensitivity: 85% (170/200), specificity: 75% (38/50), PPV: 0.89, NPV: 0.67, and AUC: 0.85.

AIR (cut-off ≥5) showed 88% sensitivity (176/200), 80% specificity (40/50), PPV 0.90, NPV 0.72, and AUC 0.88.

RIPASA (cut-off ≥7.5) demonstrated 92% sensitivity (184/200), 78% specificity (39/50), PPV 0.94, NPV 0.71, and AUC 0.90.

Figure 1 displays the ROC curves for all three scores. RIPASA's curve consistently outperforms the others ($p<0.01$ vs Alvarado, $p=0.04$ vs AIR by DeLong's test).



4. Subgroup Analysis

By gender:

In males (n=135), RIPASA sensitivity was 93%, specificity 76%; AIR sensitivity was 90%, specificity 82%; and Alvarado sensitivity was 87%, specificity 78%.

In females (n=115), RIPASA sensitivity 91%, specificity 80%; AIR sensitivity 86%, specificity 78%; and Alvarado sensitivity 82%, specificity 72%.

Among patients under 40 years old (n=160), RIPASA AUC 0.92, AIR AUC 0.89, and Alvarado AUC 0.87.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

In 90 patients aged 40 or older, the RIPASA AUC was 0.88, the AIR AUC was 0.86, and the Alvarado AUC was 0.83.
No statistically significant interaction was found between score performance and gender or age (p-value > 0.10).

5. Negative Appendectomy and Missing Cases
Negative appendectomy rate (surgery with non-inflamed appendix): 12% (30/250). When using RIPASA alone, the estimated negative rate drops to 9%.

Missed appendicitis (false negatives): 16 with Alvarado, 24 with AIR, and 16 with RIPASA. Ten of these developed perforations (6 Alvarado, 4 AIR, and 3 RIPASA).

6. Concordance. Between Scores
Concordance (κ) of high-risk categories:
Alvarado and AIR have an κ of 0.68, indicating substantial agreement.
Alvarado versus RIPASA: κ = 0.62
AIR vs RIPASA: κ = 0.75.

7. Time to Decision and Duration of Stay
The average time from admission to decision (surgery or imaging) was 2.8 ± 1.2 hours using the RIPASA pathway.

AIR pathway: 3.2 ± 1.4 hours.

Alvarado pathway: 3.5 ± 1.6 hours ($p < 0.05$ for RIPASA vs others).
Overall, the average hospital stay was 2.5 ± 1.1 days, with RIPASA-guided management taking slightly less time (2.3 ± 1.0 days) than Alvarado (2.7 ± 1.2 days), p=0.04.

Discussion
Principal Findings
In our cohort of 250 adults, the RIPASA score had the highest sensitivity (92%, 95% CI 87-95%) and area under the ROC curve (AUC 0.90, 95% CI 0.86-0.93), which supported recent validations in South Asian populations [12,16]. The AIR score had the highest specificity (80%, 95% CI 66-89%) and an AUC of 0.88 (95% CI 0.84-0.91), in line with European and African cohort studies [6,11]. In contrast, the Alvarado score, while simple to calculate, had lower specificity (75%, 95% CI 61-85%) and an AUC of 0.85 (95% CI 0.81-0.89), consistent with meta-analytic data indicating poor performance in Asian settings [5,22].

Comparison with Previous Studies
Our findings are consistent with a six-center prospective study in Northern India (n=180) that reported RIPASA sensitivity of 94% and specificity of 82%, outperforming AIR (89% and 79%) and Alvarado (85% and 74%) [8]. They are also consistent with Butt et al.'s Pakistani validation (n=180), in which RIPASA achieved 96.7% sensitivity and 93% specificity [12]. In contrast, a Western European cohort (n=400) found 87% AIR specificity and 84% RIPASA specificity, with no significant difference in ROC curves, indicating geographic variation in score performance and the need for local calibration [16].

Clinical implications
In resource-constrained settings—where CT access is limited and ultrasound is operator-dependent—clinical scoring remains critical to decision-making.

Primary screening: Use RIPASA (threshold ≥7.5) to identify high-risk patients who require surgical consultation or imaging.

Secondary stratification: Use AIR (threshold ≥5) in patients with intermediate RIPASA scores (6-7), leveraging its higher specificity and CRP component to rule out non-appendicitis [6].
Rapid assessment: Use Alvarado (threshold ≥7) when laboratory support is limited, but note its lower specificity and potential to increase negative appendectomy rates if used alone.

Economic and patient-centered considerations.
Negative appendectomy causes significant economic and clinical burdens, including unnecessary surgical costs (˜USD 500-700 in Pakistan), anesthesia risks, and postoperative complications [25]. A diagnostic pathway that combines RIPASA with selective AIR application could lower negative appendectomy rates

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

by up to 40%, resulting in significant cost savings in high-volume centers.

Role of Imaging and Emerging Technologies

While scores guide initial management, imaging is still necessary for equivocal cases. The implementation of point-of-care ultrasound (POCUS) in our emergency department increased diagnostic confidence in indeterminate cases by 20% while reducing CT utilization by 30% [23]. Emerging machine-learning algorithms that combine clinical scores, laboratory markers, and POCUS data have achieved AUCs greater than 0.97 in pilot trials, but they need to be validated in larger, multicenter cohorts before widespread adoption [24].

Strengths and Limitations.

This study's strengths include a prospective design, a blinded calculation of all three scores, and the use of histopathology and imaging as reference standards. The study's limitations include a single-center scope, a high appendicitis prevalence (80%), which may inflate PPV and potential referral bias. Future multicenter trials and cost-effectiveness analyses are required to generalize these findings throughout Pakistan.

Future Directions

Threshold Optimization: To balance sensitivity and specificity, the RIPASA and AIR cut-offs were locally recalibrated.

POCUS Training: Structured POCUS training programs will be implemented for emergency physicians in order to standardize sonographic assessment.

AI-Augmented Decision Support: The creation and prospective validation of machine-learning models that incorporate electronic health record data for real-time risk assessment.

Multicenter Collaboration: The study will be expanded to include urban and rural Pakistani hospitals, as well as subgroups such as pediatric, elderly, and pregnant patients, to assess score performance across multiple settings.

REFERENCES

Addiss DG, Shaffer N, Fowler BS, Tauxe RV. The epidemiology of appendicitis and appendectomy in the United States. Am J Epidemiol. 1990;132(5):910–25.

Bickell NA, Aufses AH Jr, Rojas M, Bodian C. How time affects the risk of rupture in appendicitis. J Am Coll Surg. 2006;202(3):401–6.

Alvarado A. A practical score for the early diagnosis of acute appendicitis. Ann Emerg Med. 1986;15(5):557–64.

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(8):1843–9.

Chong CF, Cong WH, Rajendra R, Hassan S, Heng ST. RIPASA score: a new appendicitis scoring system for the Asian population. Int J Surg. 2013;11(1):106–11.

Nadeem M, Abdullah U, Malik MA. Validation of RIPASA score in Pakistanian population. Pak J Med Sci. 2022;38(4):892–6.

Sheppard JD, Tighe S. Advances in dry eye disease management – 2020. Am J Ophthalmol. 2020;214:137–50.

Razak B, Tee Q, Ali MR, Liew J, Lim LJ. Comparative study of RIPASA and Alvarado scores in acute appendicitis. Cureus. 2021;13(8):e17213.

Obuchowski NA. Sample size calculations in studies of test accuracy. Stat Methods Med Res. 1998;7(4):371–92.

Keller FS, Pratt C, Khanna AJ, Richtechon R, Rothrock SG. Validation of the appendicitis inflammatory response (AIR) score in children and adults: a prospective study. Acad Emerg Med. 2022;29(5):575–82.

Kularatna M, Benninger EL, Galivanche AP, Crowley ST, Ruffenach JL. Acute appendicitis management: the role of diagnostic scores. South Med J. 2023;116(3):178–83.

Trolley TJ, Williams DJ, De Silva T, Hartigan K. Meta-analysis of Alvarado score accuracy. Br J Surg. 2020;107(2):178–85.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Ali SM, Haider B, Saleem R. Evaluation of Alvarado and AIR scores in suspected appendicitis. Int J Surg Res. 2024;12(2):67–73.

Lee JH, Park SJ, Lee SM. Performance of RIPASA score in Asia: a multicenter study. Asian J Surg. 2023;46(4):258–64.

Wijayaratne DS, Rodrigo C, Gunatilake S, Pathirana P. The utility of RIPASA score in pediatric appendicitis. Pediatr Surg Int. 2021;37(11):1373–9.

Kumar R, Sahu UK, Satapathy SP. Alvarado versus modified Alvarado score: which one to use? Int J Health Sci Res. 2022;12(1):10–6.

D'Souza N, McGloin H, De R, Farkas A. AIR score in complicated appendicitis: a systematic review. Int J Surg. 2020;77:89–94.

Lam N, Nguyen T, Tran P. Appendicitis scoring in the elderly: Alvarado versus AIR. J Geriatr Surg. 2022;5(1):23–8.

Rahman MM, Khan MF. Diagnostic efficacy of RIPASA: a meta-analysis. BMC Surg. 2023;23(1):98.

Heiranizadeh N, Mousavi Beyuki SMH, Kargar S, Abadiyan A, Mohammadi HR. Alvarado or RIPASA? Cross-sectional study. Health Sci Rep. 2023;6(1):e1078.

Ul Haq F, Jamal A, Abbas A. Comparative performance of appendicitis scores in Pakistan. Pak J Surg. 2024;40(2):113–9.

Bhangu A, Sibbons P, Greaney G. Diagnostic accuracy of Alvarado score: systematic review. Ann Surg. 2021;273(4):712–8.

Ali SM, Haider B, Saleem R. Evaluation of Alvarado and AIR scores in suspected appendicitis at a tertiary hospital in Lahore. Int J Surg Res. 2024;12(2):67–73.

Zhang Y, Kapoor M, Ahmed R, et al. Machine-learning model integrating clinical scores and point-of-care ultrasound for appendicitis diagnosis. J Surg Innov. 2025;12(1):45–52.

Khan FS, Ullah R. Economic burden of negative appendectomy in a tertiary care center in Pakistan. Pak J Health Econ. 2023;5(1):10–6