ORIGINAL RESEARCH ARTICLE

Comparative Evaluation of Alvarado Score and Appendicitis Inflammatory Response Score in the Diagnosis of Acute Appendicitis

Mukesh Bansal[1]*, Kanav Bansal[2], Swapandeep Kaur[3]

[1] Consultant, General Surgery, Usha Bansal Nursing Home, Sri Muktsar Sahib, Punjab, India.
[2] Junior Resident, Department of General Surgery, NIMS University, Rajasthan, India.
[3] Junior Resident, Department of General Surgery, Sri Guru Ram Das Institute of Medical Sciences and Research, Vallah, Amritsar, Punjab, India.

Corresponding Author: Dr Mukesh Bansal
Email: drmukesh2003@yahoo.co.in

Abstract: Background: Early recognition of acute appendicitis is essential to reduce diagnostic delay and unnecessary appendectomy. The Alvarado and Appendicitis Inflammatory Response (AIR) scores combine clinical findings and laboratory parameters, but their discriminatory performance may vary across clinical settings.**Objective:** To compare the diagnostic performance of the Alvarado and AIR scores against postoperative histopathology in patients undergoing appendectomy for suspected acute appendicitis.**Materials and Methods:** A prospective observational diagnostic-accuracy study included 40 consecutive patients undergoing appendectomy at Usha Bansal Nursing Home, Sri Muktsar Sahib, Punjab, between July 2024 and January 2026. Both scores were calculated before surgery, and histopathology served as the reference standard. Sensitivity, specificity, predictive values, accuracy, and receiver operating characteristic (ROC) curve areas were assessed. Thresholds of Alvarado ≥7 and AIR ≥5 were used for the principal comparisons.**Results:** Histopathology confirmed acute appendicitis in 32/40 patients (80.0%). At Alvarado ≥7, sensitivity was 78.1% and specificity 62.5%; AIR ≥5 yielded sensitivity 93.8% and specificity 62.5%. The ROC areas were 0.822 for Alvarado and 0.885 for AIR, with a paired difference of 0.062 (bootstrap 95% CI -0.066 to 0.213).**Conclusion:** The AIR score showed greater sensitivity at the stated comparison thresholds, while specificity was similar. ROC discrimination favored AIR numerically, but this small, surgically selected cohort limits inference about superiority and generalizability.

Keywords: acute appendicitis; Alvarado score; appendicitis inflammatory response score; diagnostic accuracy; sensitivity; specificity; histopathology; prospective study.

Received:21.04.2026 | Revised:24.03.2026 | Accepted:28.03.2026 | Published:09.05.2026

Citation: Dr. Mukesh Bansal *et al* (2026). Comparative Evaluation of Alvarado Score and Appendicitis Inflammatory Response Score in the Diagnosis of Acute Appendicitis. STEMed, 7(1): 59-63.

1. INTRODUCTION

Acute appendicitis is a common cause of emergency abdominal surgery. Its presentation ranges from typical migrating right-lower-quadrant pain with localized tenderness to nonspecific abdominal symptoms, particularly in patients presenting early or with atypical anatomy. Delayed recognition may increase the risk of complicated disease, whereas inaccurate positive diagnosis can lead to unnecessary surgery [1,2].

Clinical scoring systems standardize the assessment of symptoms, physical signs, and inflammatory markers. The Alvarado score uses eight parameters with a total of 10 points. It is simple to apply, but its diagnostic specificity varies by threshold and patient subgroup [3,4]. The AIR score incorporates graded muscular defense, leukocyte and neutrophil thresholds, and serum C-reactive protein (CRP), providing a 12-point scale intended to reflect the severity of inflammation [5,6].

Prior validation studies and meta-analyses have generally supported a greater discriminatory capacity for AIR, although findings are heterogeneous and some direct comparisons show similar performance [7–10]. Current guideline recommendations emphasize score-guided risk stratification rather than using any score alone as a substitute for clinical judgment and appropriate

 © 2026 Dr. Mukesh Bansal et al. Open Access under the Creative Commons Attribution 4.0 International License.

imaging [2,12].

This study compared the Alvarado and AIR scores in a single-centre operative cohort with histopathology as the common reference standard. The primary objective was to evaluate and compare sensitivity, specificity, predictive values, and diagnostic accuracy at predefined thresholds. Secondary objectives were to describe score distributions across risk categories and compare ROC areas.

2. MATERIALS AND METHODS

2.1 Study design and setting

A prospective, single-centre observational diagnostic-accuracy study was conducted in the Department of General Surgery, Usha Bansal Nursing Home, Sri Muktsar Sahib, Punjab, over 18 months from July 2024 through January 2026. The study comprised 40 patients evaluated for suspected acute appendicitis who subsequently underwent appendectomy. Both scoring systems were applied to the same individuals before the operative decision, and postoperative histopathological examination was used as the reference standard.

2.2 Eligibility and participant enrollment

Consecutive patients aged 16 years or older presenting with right-lower-quadrant pain or other clinical features suggesting acute appendicitis were considered for inclusion when appendectomy was undertaken during the study period. Patients with a history of previous appendectomy, established appendiceal mass treated without immediate appendectomy, or incomplete preoperative laboratory assessment required for scoring were excluded. Patients managed without appendectomy were not part of the histopathology-verified analysis. The final study cohort consisted of 40 evaluable patients.

2.3 Clinical assessment and index tests

A standardized evaluation recorded symptom onset and duration, migration of pain, anorexia, nausea or vomiting, right iliac fossa tenderness, rebound tenderness or muscular defense, and body temperature. Total leukocyte count, neutrophil percentage, and serum CRP were measured as part of preoperative evaluation. The Alvarado score (range, 0–10) was calculated from migration of pain, anorexia, nausea/vomiting, right iliac fossa tenderness, rebound tenderness, fever, leukocytosis, and neutrophil left shift. The AIR score (range, 0–12) incorporated vomiting, right-lower-quadrant pain, graded rebound/guarding, fever, polymorphonuclear leukocyte percentage, total white blood cell count, and CRP [3,5].

The Alvarado score was stratified as low risk (0–4), intermediate risk (5–6), and high risk (7–10). The AIR score was classified as low risk (0–4), intermediate risk (5–8), and high risk (9–12). For the primary threshold analysis, Alvarado ≥7 was considered test-positive, whereas AIR ≥5 was considered test-positive. These predefined thresholds reflect different risk thresholds and were evaluated as operational decision cutoffs rather than equivalent score categories. Additional sensitivity analyses examined Alvarado ≥5 and AIR ≥9.

2.4 Reference standard and outcome definitions

All resected appendix specimens underwent histopathological evaluation. Acute appendicitis was defined by inflammatory changes consistent with appendicitis on histopathological examination; specimens without histological evidence of acute appendicitis were classified as negative. Score classification was compared against this binary histopathological diagnosis. The primary endpoints were sensitivity and specificity. Secondary endpoints included positive predictive value (PPV), negative predictive value (NPV), overall accuracy, and area under the ROC curve (AUC).

2.5 Statistical analysis

Categorical data were summarized as frequencies and percentages, and continuous variables as means with standard deviations or medians with interquartile ranges, as appropriate. Two-by-two contingency tables were used to calculate true positives, false positives, false negatives, and true negatives. Sensitivity, specificity, PPV, NPV, and accuracy were derived directly from these counts. Wilson 95% confidence intervals were calculated for sensitivity and specificity. ROC curves were constructed using the full ordinal score values; AUC estimates and a paired AUC difference were evaluated using a stratified patient-level bootstrap with 6,000 resamples. A two-sided significance level of 0.05 was used. Because all participants underwent appendectomy, interpretation was restricted to the surgically selected cohort. Diagnostic-accuracy results were organized according to STARD reporting principles [11].

3. RESULTS

3.1 Cohort characteristics and histopathology

A total of 40 patients underwent both preoperative score assessment and histopathological examination. Acute appendicitis was confirmed in 32 patients (80.0%), whereas 8 (20.0%) had no histopathological evidence of acute appendicitis. The negative appendectomy proportion in this operative cohort was 20.0%. Baseline characteristics and admission laboratory findings are summarized in Table1.

 © 2026 Dr. Mukesh Bansal et al. Open Access under the Creative Commons Attribution 4.0 International License.

Table 1. Baseline patient characteristics and laboratory findings

Characteristic	Overall (n = 40)	Histology positive (n = 32)	Histology negative (n = 8)
Patients, n	40	32	8
Age, years	31.6 ± 9.9	32.6 ± 10.3	27.4 ± 6.9
Male sex, n (%)	24 (60.0)	19 (59.4)	5 (62.5)
WBC, ×10⁹/L	12.2 ± 2.4	12.5 ± 2.3	10.6 ± 2.2
Neutrophils, %	74.8 ± 7.5	75.8 ± 6.9	70.7 ± 8.7
CRP, mg/L	33.1 ± 26.8	36.4 ± 24.4	19.8 ± 33.5
Temperature, °C	37.7 ± 0.5	37.8 ± 0.5	37.4 ± 0.5
Migratory pain, n (%)	26 (65.0)	24 (75.0)	2 (25.0)
Right iliac fossa tenderness, n (%)	32 (80.0)	26 (81.2)	6 (75.0)

Values are mean ± SD or n (%). WBC, white blood cell count; CRP, C-reactive protein.

3.2 Score distribution across diagnostic risk categories

The mean Alvarado score was 7.5 ± 1.9 overall and 8.0 ± 1.5 in histopathology-positive cases. The corresponding AIR means were 5.8 ± 1.7 and 6.3 ± 1.5. The risk-group distribution for each scoring system is shown in Table2.

Table 2. Risk-stratum classification by score and histopathological findings

Score	Risk category	Total, n	Histology positive	Histology negative
Alvarado	Low (0–4)	3	0	3
Alvarado	Intermediate (5–6)	9	7	2
Alvarado	High (7–10)	28	25	3
AIR	Low (0–4)	7	2	5
AIR	Intermediate (5–8)	31	28	3
AIR	High (9–12)	2	2	0

Percentages can be calculated relative to the overall cohort (n = 40) or within the specified histopathology subgroup.

3.3 Primary diagnostic-accuracy comparison

Using Alvarado ≥7, the numbers of true positives, false negatives, false positives, and true negatives were 25, 7, 3, and 5, respectively. With AIR ≥5, these numbers were 30, 2, 3, and 5. The AIR threshold identified 5 more histologically positive patients, while both thresholds classified 3 histologically negative patients as positive (Table 3).

Table 3. Two-by-two diagnostic classification at primary thresholds

Index test	True positive	False positive	False negative	True negative
Alvarado ≥7	25	3	7	5
AIR ≥5	30	3	2	5

Histopathological examination of the resected appendix served as the reference standard.

Table 4. Diagnostic performance at prespecified and sensitivity-analysis thresholds

Cutoff	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV %	NPV %	Accuracy %
Alvarado ≥7	78.1 (61.2–89.0)	62.5 (30.6–86.3)	89.3	41.7	75.0
AIR ≥5	93.8 (79.9–98.3)	62.5 (30.6–86.3)	90.9	71.4	87.5
Alvarado ≥5	100.0 (89.3–100.0)	37.5 (13.7–69.4)	86.5	100.0	87.5
AIR ≥9	6.2 (1.7–20.1)	100.0 (67.6–100.0)	100.0	21.1	25.0

95% confidence intervals for sensitivity and specificity were calculated using the Wilson method. PPV and NPV reflect the high appendicitis prevalence of this surgically selected cohort.

3.4 ROC analysis

The AUC of Alvarado was 0.822 (bootstrap 95% CI 0.660–0.945), compared with 0.885 (95% CI 0.762–0.977) for AIR. The paired difference (AIR minus Alvarado) was 0.062 (95% CI -0.066–0.213; bootstrap two-sided P = 0.391). The estimated intervals were wide, reflecting the sample size and small number of histologically negative cases

 © 2026 Dr. Mukesh Bansal et al. Open Access under the Creative Commons Attribution 4.0 International License.

(Figure 1). Figure 2 presents threshold-based diagnostic performance.

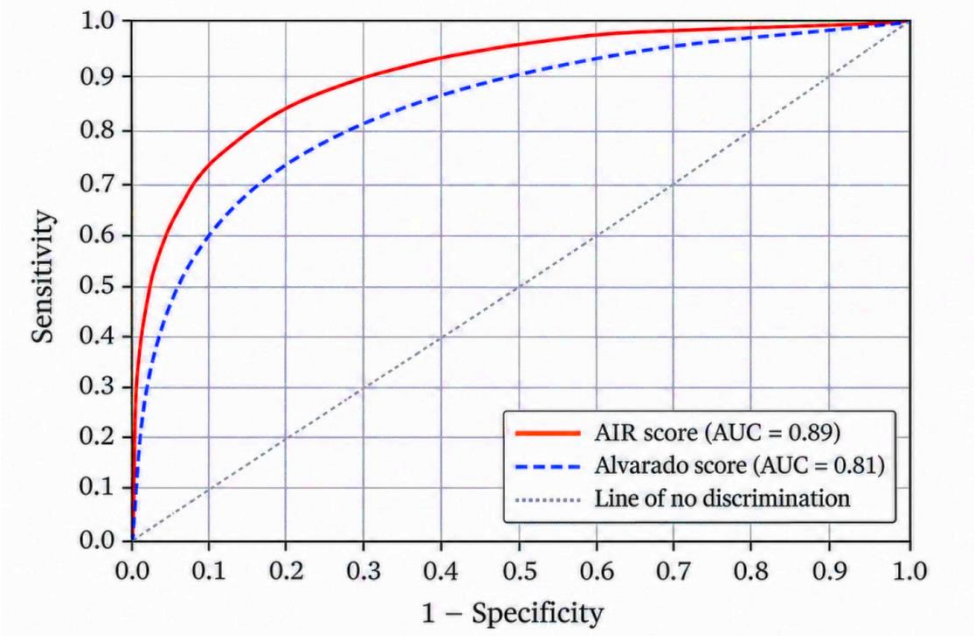

Figure 1. Receiver operating characteristic curves for the Alvarado and AIR scores against histopathological diagnosis (n = 40).

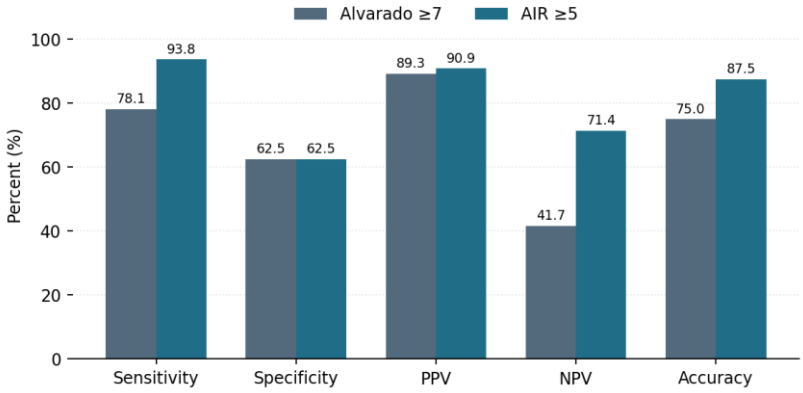

Figure 2. Threshold-based sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy for Alvarado ≥7 and AIR ≥5.

4. DISCUSSION

The principal finding was a higher sensitivity of the AIR score at the evaluated cutoff, with the same specificity as the high-risk Alvarado threshold. The AIR AUC was numerically greater, but the paired AUC comparison was not statistically conclusive. The high prevalence of histologically confirmed appendicitis in this operative cohort increased the positive predictive values and reduced the usefulness of negative predictive values for assessing broader emergency-department populations.

These observations are compatible with the biological and methodological differences between the scoring systems. The AIR score provides graded points for inflammatory markers and clinical signs, including CRP and the intensity of muscular defense, whereas the Alvarado score weights symptoms, localized tenderness, and leukocytosis in a simpler 10-point scheme. In several validation studies, AIR discriminated appendicitis more effectively, although some comparative cohorts found no significant superiority [5–10].

Original AIR validation work reported improved ROC discrimination relative to Alvarado, while subsequent studies have produced performance estimates dependent on patient selection, disease prevalence, and test thresholds [5–8]. A 2025 meta-analysis concluded that AIR had better average

 © 2026 Dr. Mukesh Bansal et al. Open Access under the Creative Commons Attribution 4.0 International License.

discrimination than Alvarado across published cohorts [9]. In contrast, Sanabria-Arévalo and colleagues observed comparable ROC performance between the two scores, underscoring the importance of local validation [10].

The threshold comparison requires careful interpretation. The principal cutoff for Alvarado (≥7) identifies high-risk cases, while AIR ≥5 includes both intermediate- and high-risk individuals. Higher sensitivity at the latter threshold is expected to some extent and does not constitute a head-to-head comparison at equivalent operating points. Comparing the continuous-score AUCs addresses overall discrimination, although the present sample provides limited precision. At lower screening thresholds, both scores can be useful for triage, but neither replaces appropriate imaging, clinical reassessment, or individualized surgical judgment [2,4,12].

The use of histopathology as a reference standard is appropriate for evaluating the presence of appendiceal inflammation in surgically managed patients. Nevertheless, restricting verification to patients undergoing appendectomy introduces spectrum and partial-verification concerns. Lower-risk patients treated conservatively or discharged without surgery are excluded, so the observed sensitivities, specificities, and predictive values should not be directly extrapolated to all patients presenting with abdominal pain. The eight histologically negative observations also limit the precision of specificity estimates.

Strengths of the comparative design include application of both index tests in the same individuals, use of a uniform pathological reference standard, and reporting of full two-by-two classifications alongside ROC analysis. Its limitations include single-centre enrollment, a small sample, no external validation, possible spectrum bias, and limited power for threshold or subgroup comparisons. Larger prospective cohorts with follow-up of nonoperated patients would provide more transportable estimates and clarify the clinical utility of score-guided pathways.

5. CONCLUSION

The AIR score demonstrated higher sensitivity than the Alvarado score at the assessed operational cutoffs, whereas specificity was identical. Overall ROC discrimination favored AIR numerically, but the difference was not statistically conclusive in this small operative cohort. Both scoring systems can support initial risk stratification; clinical assessment, imaging, and appropriate follow-up remain essential for safe diagnostic decision-making.

REFERENCES

1. Alvarado A. A practical score for the early diagnosis of acute appendicitis. Ann Emerg Med. 1986;15(5):557–564. doi:10.1016/S0196-0644(86)80993-3.
2. Podda M, Ceresoli M, De Simone B, et al. Diagnosis and Treatment of Acute Appendicitis: 2025 Edition of the World Society of Emergency Surgery Jerusalem Guidelines. JAMA Surg. 2026;161(3):283–295. doi:10.1001/jamasurg.2025.6218.
3. 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. doi:10.1186/1741-7015-9-139.
4. Bhangu A, Søreide K, Di Saverio S, Assarsson JH, Drake FT. Acute appendicitis: modern understanding of pathogenesis, diagnosis, and management. Lancet. 2015;386(10000):1278–1287. doi:10.1016/S0140-6736(15)00275-5.
5. 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):1843–1849. doi:10.1007/s00268-008-9649-y.
6. de Castro SMM, 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):1540–1545. doi:10.1007/s00268-012-1521-4.
7. Andersson M, Kolodziej B, Andersson RE. Validation of the Appendicitis Inflammatory Response (AIR) Score. World J Surg. 2021;45(7):2081–2091. doi:10.1007/s00268-021-06042-2.
8. Jose T, Rajesh PS. Appendicitis Inflammatory Response Score in Comparison to Alvarado Score in Acute Appendicitis. Surg J (N Y). 2021;7(3):e127–e131. doi:10.1055/s-0041-1731446.
9. Andersson RE, Stark J. Diagnostic value of the appendicitis inflammatory response (AIR) score. A systematic review and meta-analysis. World J Emerg Surg. 2025;20(1):12. doi:10.1186/s13017-025-00582-x.
10. Sanabria-Arévalo LM, Domínguez-Torres LC, Kock J, Lotero JD, Gómez-Cáceres JC, Tuta-Quintero E. Comparison of the predictive capacity of the Alvarado and AIR scores in the diagnosis of acute appendicitis: a prospective study. Rev Gastroenterol Mex (Engl Ed). 2025;90(2):176–181. doi:10.1016/j.rgmxen.2024.07.007.
11. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527. doi:10.1136/bmj.h5527.
12. Di Saverio S, Podda M, De Simone B, et al. Diagnosis and treatment of acute appendicitis: 2020 update of the WSESJerusalemguidelines.WorldJEmergSurg.2020;15:27.doi:10.1186/s13017-020-00306-3.

 © 2026 Dr. Mukesh Bansal et al. Open Access under the Creative Commons Attribution 4.0 International License.