RESEARCH PAPER

Comparative Diagnostic Accuracy of the Tzanakis and Alvarado Scoring Systems in Acute Appendicitis: A Retrospective Cohort Study

Dr Smruti Paramjyoti[1], Dr Nanditha Gudi[2], Dr M S Dwarakesh[3], Dr Sudarshan P B[4]

[1]Post Graduate, Department of General Surgery, Saveetha Medical College and Hospital, Chennai, India.
[2]Associate Professor, Department of General Surgery, Saveetha Medical College and Hospital, Chennai, India.
[3]Post Graduate, Department of General Surgery, Saveetha Medical College and Hospital, Chennai, India.
[4]Professor, Department of General Surgery, Saveetha Medical College and Hospital, Chennai, India.

Abstract

Background: Acute appendicitis remains a diagnostic challenge. The Alvarado score, based on clinical and laboratory parameters, is widely used but has shown inconsistent sensitivity. The Tzanakis score, which integrates ultrasonography (USG), has been proposed as a superior alternative. This study compared the diagnostic accuracy of these two scoring systems.

Methods: A retrospective cohort study was conducted from January 1, 2024 to December 31, 2025 at Saveetha Medical College, including 150 patients who underwent appendectomy for suspected acute appendicitis. Each patient was preoperatively scored using both the Alvarado and Tzanakis systems, with histopathological examination (HPE) as the diagnostic gold standard. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated.

Results: The Tzanakis score demonstrated a sensitivity of 80.6% and specificity of 100%. The Alvarado score showed a substantially lower sensitivity of 11.3%, with specificity of 100%. Right iliac fossa pain was universal (100%), followed by anorexia (61.3%) and nausea/vomiting (48.0%).

Conclusion: The Tzanakis scoring system exhibits significantly higher sensitivity than the Alvarado score while maintaining equally high specificity. Incorporation of ultrasound findings makes the Tzanakis score a more reliable and clinically useful tool for diagnosing acute appendicitis, with potential to reduce the rate of negative appendectomy.

Keywords: Appendicitis; Tzanakis score; Alvarado score; Diagnostic accuracy; Ultrasound; Sensitivity; Specificity; Negative appendectomy

How to cite this article: Paramjyoti S, Gudi N, Dwarakesh MS, Sudarshan PB. Comparative diagnostic accuracy of the Tzanakis and Alvarado scoring systems in acute appendicitis: a retrospective cohort study. Int J Drug Deliv Technol. 2026;16(74s): 795-799. DOI: 10.25258/ijddt.16.74s.96

Introduction

No abdominal surgical emergency is encountered more often than acute appendicitis, which carries an estimated lifetime risk of 7.0–8.0% [1]. Diagnosis still leans on a combination of history, examination, and supporting tests, yet real diagnostic uncertainty persists — especially among children, the elderly, and women — feeding into an error rate of 15.0–30.0% that surfaces either as unnecessary (negative) appendectomy or as delayed diagnosis culminating in perforation and peritonitis [2,3]. To bring order to this uncertainty, the Alvarado Score was introduced in 1986, converting symptoms (migratory pain, anorexia, nausea/vomiting), examination findings (right lower quadrant tenderness, rebound tenderness, fever), and laboratory markers (leukocytosis, neutrophilia) into a single numerical score [4].

Simple as it is to apply, its reported sensitivity swings widely across studies (55.0–96.0%) and often disappoints in day-to-day practice [5,6] — a shortfall generally attributed to its dependence on clinical variables alone, with no direct view of the appendix itself. Point-of-care ultrasound (USG) changed that equation by letting clinicians directly visualise an inflamed appendix. Building on this, the Tzanakis Score (2005) folds sonographic findings — a non-compressible, aperistaltic appendix over 6 mm across, an appendicolith, and/or peri-appendiceal fluid — into the same framework alongside clinical and inflammatory laboratory markers [7], with the aim of lifting diagnostic accuracy through this hybrid design.

Even so, the Tzanakis Score has not been applied uniformly across the literature, with comparative studies against the Alvarado Score varying by geography and clinical setting, and the degree to which radiologist skill shapes its performance remains a live practical question. We therefore undertook a direct, head-to-head comparison of the diagnostic performance of the Tzanakis and Alvarado

systems for acute appendicitis at our tertiary centre, anchoring both against histopathological examination as the definitive reference standard.

Materials and Methods

Setting and Research Design

This was a retrospective, observational cohort study performed at the Department of General Surgery, Saveetha Medical College and Hospital, Chennai, India, between January 1, 2024 and December 31, 2025. The Institutional Ethics Committee approved a waiver of informed consent given the retrospective study design

Study Population

150 consecutive adult patients (≥18 years) presenting with clinical signs suggestive of acute appendicitis, subsequently managed with laparoscopic or open appendectomy, were included. Patients with incomplete medical records, those not undergoing surgery, or those with an alternative final diagnosis (e.g., gynaecological pathology, diverticulitis) were excluded. All patients underwent a standardised preoperative abdominal ultrasound performed by a certified radiologist.

Data Collection

Data were obtained retrospectively from the hospital Medical Information Archival System (MIAS) and patient records using a pre-designed proforma, capturing:

- Demographics: age and sex.
- Clinical parameters: right iliac fossa pain, migratory pain, anorexia, nausea/vomiting; tenderness in the right iliac fossa, rebound tenderness, fever >37.3°C.
- Laboratory parameters: preoperative total white blood cell count and differential neutrophil count.
- Ultrasound findings: aperistaltic, non-compressible appendix >6 mm, appendicolith, peri-appendiceal fluid, and/or local fat stranding.
- Scoring systems: each patient assigned both Alvarado (1–10) and Tzanakis (1–18) scores. A Tzanakis score ≥8 was considered positive per the original validation publication [7]; Alvarado scores were categorised as 1–4 (unlikely), 5–6 (possible), and 7–10 (probable) [4].

The gold standard for definitive diagnosis was histopathological examination (HPE) of the resected appendix, reported by a qualified histopathologist blinded to clinical scoring.

Statistical Analysis

Data were recorded in Microsoft Excel and analysed using IBM SPSS for Windows, version 22.0. Continuous variables are reported as mean ± SD or median with interquartile range as appropriate; categorical variables as frequencies and percentages. Sensitivity, specificity, PPV, NPV, and overall accuracy were calculated against HPE as the reference standard. Agreement between HPE and each scoring system was evaluated, with $p<0.05$ considered statistically significant.

Results

Mean age of the study population was 32.5 ± 12.8 years, with a male-to-female ratio of 1.2:1. The most prevalent presenting symptom was right iliac fossa (RIF) pain, present in all 150 patients (100%), followed by anorexia in 92 patients (61.3%) and nausea/vomiting in 72 patients (48.0%). Migratory pain was reported by 57 patients (38.0%), rebound tenderness in 48 patients (32.0%), and fever >37.3°C in 64 patients (42.7%) (Table 1). Score-category distribution differed markedly between systems: using the Alvarado score, the majority of patients (89, 59.3%) fell into the "possible" category (5–8), with only 15 (10.0%) in the high-probability "probable" category (>8). In contrast, the Tzanakis score classified 107 patients (71.3%) into the positive diagnostic category (≥8) — indicating a substantially higher index of suspicion generated by the ultrasound-integrated system (Table 2).

Histopathological examination confirmed acute appendicitis in 124 of 150 patients (82.7%). The Tzanakis score (cut-off ≥8) demonstrated a sensitivity of 80.6% (100/124 true positives) and specificity of 100% (26/26 true negatives), with PPV of 100% and NPV of 44.8%. The Alvarado score, at the standard cut-off of >7 for a "probable" diagnosis, showed a markedly lower sensitivity of 11.3% (14/124 true positives) but an equally perfect specificity of 100% (26/26 true negatives). Lowering the Alvarado cut-off to >4 (including "possible" and "probable" categories) increased sensitivity to 69.4% but reduced specificity to 30.8% (Table 3).

Table 1. Prevalence of Clinical Signs and Symptoms in the Study Cohort (N=150)

Sign/Symptom	Number (n)	Percentage (%)
Pain in right iliac fossa	150	100.00
Anorexia	92	61.33
Nausea/vomiting	72	48.00
Migratory pain	57	38.00
Rebound tenderness	48	32.00
Temperature >37.3°C	64	42.67

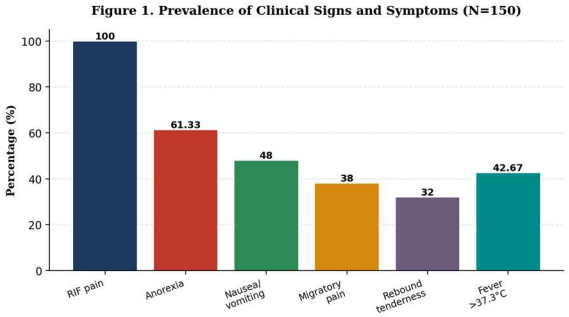

Figure 1: Prevalence of key clinical signs and symptoms among the study cohort at presentation (N=150).

Table 2. Distribution of Study Population by Score Category

Score Category	Alvarado Score, n (%)	Tzanakis Score, n (%)
Low probability (≤4)	46 (30.7%)	29 (19.3%)
Intermediate (5–8)	89 (59.3%)	14 (9.3%)
High probability (>8)	15 (10.0%)	107 (71.3%)
Total	150 (100%)	150 (100%)

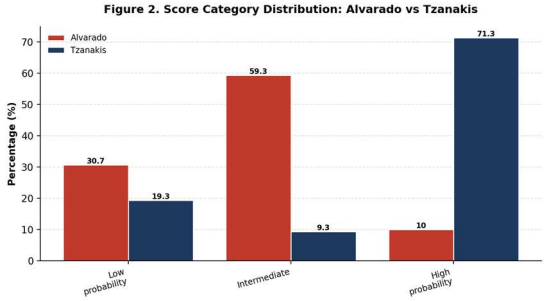

Figure 2: Distribution of patients across diagnostic-probability categories, contrasting the Alvarado and Tzanakis scoring systems. The Tzanakis score classified a substantially larger proportion of patients into the high-probability category.

Table 3. Diagnostic Accuracy of Alvarado and Tzanakis Scoring Systems

Parameter	Tzanakis Score (≥8)	Alvarado Score (>7)
Sensitivity	80.6%	11.3%
Specificity	100%	100%
Positive predictive value (PPV)	100%	100%
Negative predictive value (NPV)	44.8%	19.1%
Overall accuracy	83.3%	26.7%
p-value (vs. HPE)	<0.001	<0.001

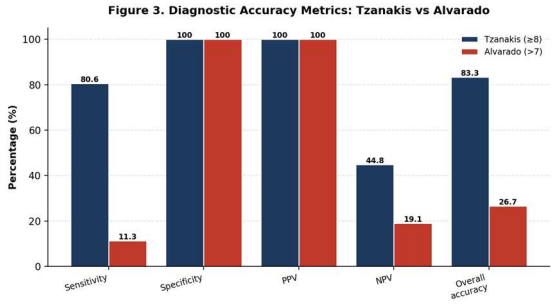

Figure 3: Comparative diagnostic accuracy metrics for the Tzanakis and Alvarado scoring systems against histopathological gold standard.

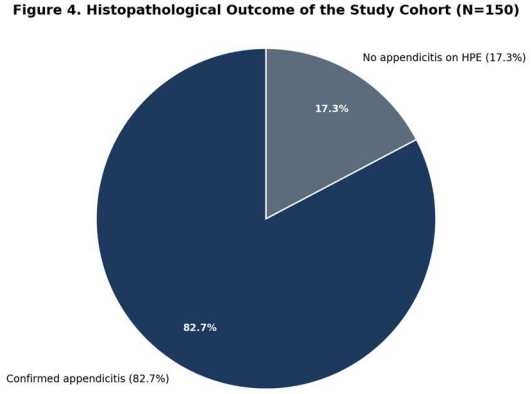

Figure 4: Proportion of patients with histopathologically confirmed acute appendicitis (gold standard) within the overall study cohort.

Discussion

Our retrospective cohort study demonstrates that the Tzanakis scoring system provides substantially better diagnostic performance than the traditional Alvarado score for detecting appendicitis. The dramatic sensitivity difference (80.6% vs 11.3%), with both scores maintaining 100% specificity, has significant clinical implications: the Tzanakis score correctly identifies over 80% of patients who truly have appendicitis, whereas the Alvarado score at its traditional high-probability cut-off would miss the substantial majority (88.7%) of true appendicitis cases. The strikingly low sensitivity of the Alvarado score observed in our study, while notable, is not unprecedented — Castro et al. reported a similarly low sensitivity (19.4%) for the Alvarado score in a population where imaging played an integral diagnostic role [8].

This illustrates how diagnostic paradigms have shifted with the advancement of imaging technology: the Alvarado score was developed and validated prior

to widespread routine preoperative ultrasound or CT, and in the present-day clinical setting where equivocal cases are commonly evaluated by imaging, its purely clinical criteria appear insufficient alone for high-sensitivity detection. This is illustrated in our cohort, where 100% of patients reported RIF pain but only 32% reported rebound tenderness and 38% reported migratory pain — clinical features the Alvarado score weights heavily but which are frequently absent or atypical. By contrast, the strong sensitivity of the Tzanakis system is consistent with its original validation study, which reported a sensitivity of 88.4% and specificity of 81.3% [7]. More recent comparative data support our findings: Malla et al. reported a direct comparison in which Tzanakis sensitivity was 84.5% versus 64.2% for Alvarado [9].

The hybrid nature of the Tzanakis score — combining objective ultrasonographic findings (non-compressibility, diameter >6 mm, peri-appendiceal fluid) with clinical findings and leukocytosis — underlies its diagnostic strength, and is particularly valuable in patients with atypical presentations, such as pregnant women and elderly patients, where clinical diagnosis alone is often unreliable. These findings carry meaningful clinical practice implications. Minimising the negative appendectomy rate (NAR) remains a key management goal, avoiding unnecessary surgical risk and cost; current literature consistently shows that incorporating ultrasound into the diagnostic pathway reduces NAR [3].

Our finding of 100% PPV for Tzanakis scores ≥8 indicates that positive scores confer very high diagnostic confidence, potentially reducing surgical hesitation and the risk of perforation from diagnostic delay. Conversely, the low sensitivity of the Alvarado score risks inappropriate deferral of surgery in true appendicitis cases, though its high specificity retains value as a rule-in screening tool when strongly positive. Important limitations of the Tzanakis score include its dependence on high-quality ultrasound and an experienced operator; published sensitivity of ultrasound for appendicitis diagnosis ranges from 71% to 97% depending on operator expertise. Our study was conducted at a tertiary centre with dedicated radiologists, which may have contributed to the strong observed performance; in settings with limited or inconsistent ultrasound access, the practical utility of the Tzanakis score may be correspondingly diminished.

The 100% specificity observed in our study is also likely influenced by selection bias inherent to the design — as only patients who proceeded to surgery were included, our cohort had a higher pre-test probability of appendicitis than an unselected emergency department population. Our findings align with the broader literature trend favouring combined clinical-and-imaging diagnostic strategies. Gorter et al., in a systematic review and meta-analysis of diagnostic strategies for appendicitis, concluded that combining clinical scoring with imaging yields the best diagnostic accuracy, even in adult populations [12]. Other composite scoring systems, such as the Appendicitis Inflammatory Response (AIR) score, have likewise shown improved sensitivity over Alvarado but still benefit from additional imaging or laboratory correlation [13].

Future research priorities include prospective, multicentre validation of the Tzanakis score across healthcare settings with varying radiological resources; development of simplified, expedited ultrasound protocols deployable by emergency physicians or surgeons to broaden access to Tzanakis-based assessment; and formal cost-effectiveness analysis comparing Tzanakis-based diagnostic pathways against traditional clinical evaluation with selective imaging.

The strengths of our study includes, both scoring systems were applied prospectively to every patient in the cohort at the time of clinical assessment, avoiding the retrospective score-reconstruction bias that affects some comparative scoring-system studies; histopathological examination — an objective, blinded reference standard — was used uniformly as the diagnostic gold standard; and all ultrasound examinations were performed by certified radiologists using a standardised protocol, reducing operator-related variability in Tzanakis score assignment.

The Limitations of our study includes as a surgical cohort restricted to patients who proceeded to appendectomy, our population inherently carried a higher pre-test probability of appendicitis than an unselected emergency department population, which likely inflates the specificity and PPV figures for both scoring systems relative to what would be observed in a broader, undifferentiated presentation cohort; this selection effect should be borne in mind when extrapolating our findings to primary triage settings.

The retrospective, single-centre design, combined with the concentration of ultrasound expertise at a tertiary referral radiology department, may limit generalisability to centres with less consistent access to experienced sonographers — a caveat we have already flagged in the Discussion but which bears restating as a formal limitation. Inter-observer variability in ultrasound interpretation between

different radiologists was not formally assessed (e.g., via a kappa statistic on a subset of paired readings), which would have strengthened confidence in the reproducibility of Tzanakis scoring at our centre.

Triage pathway design: incorporating point-of-care or radiology-performed ultrasound into the initial assessment pathway for suspected appendicitis, rather than relying on clinical scoring alone, is likely to meaningfully reduce both negative appendectomy and diagnostic-delay-related perforation, based on the magnitude of sensitivity gain demonstrated here. Resource allocation: since Tzanakis scoring is contingent on ultrasound quality and operator experience, departments adopting this approach should invest in radiologist training and image-quality assurance, rather than treating ultrasound access alone as sufficient for reliable score application. Special populations: the hybrid, imaging-inclusive nature of the Tzanakis score may offer particular value in atypical presentations — pregnancy, older age, and populations where clinical signs are frequently equivocal — where purely clinical scores such as Alvarado are already known to under-perform.

Institutional audit: given the low sensitivity of the Alvarado score demonstrated in this and comparable series, institutions relying on Alvarado alone for operative decision-making may wish to audit their own negative appendectomy rate, as a low-sensitivity screening tool used in isolation risks systematically deferring surgery in patients who would benefit from it.

Conclusion

Ultrasonographic findings, incorporated via the Tzanakis score, provide a substantially more accurate method of diagnosing acute appendicitis than the Alvarado score alone, owing to superior sensitivity with equally high specificity. The Alvarado score, in its traditional form, demonstrates limited reliability as a stand-alone diagnostic tool in the current era of widely available diagnostic ultrasound. Hospitals with adequate access to ultrasonography should consider adopting the Tzanakis score as a clinical decision aid for patients with suspected acute appendicitis. Wider adoption of ultrasound-integrated scoring has the potential to improve diagnostic accuracy, reduce unnecessary surgical procedures, ensure timely treatment for true appendicitis cases, and optimise utilisation of hospital resources.

References

1. 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.
2. Flum DR, Koepsell T. The clinical and economic correlates of misdiagnosed appendicitis: nationwide analysis. Arch Surg. 2002;137(7):799-804.
3. Seetahal SA, Bolorunduro OB, Sookdeo TC, Oyetunji TA, Greene WR, Frederick W, et al. Negative appendectomy: a 10-year review of a nationally representative sample. Am J Surg. 2011;201(4):433-7.
4. Alvarado A. A practical score for the early diagnosis of acute appendicitis. Ann Emerg Med. 1986;15(5):557-64.
5. 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.
6. Kalan M, Talbot D, Cunliffe WJ, Rich AJ. Evaluation of the modified Alvarado score in the diagnosis of acute appendicitis: a prospective study. Ann R Coll Surg Engl. 1994;76(6):418-9.
7. Tzanakis NE, Efstathiou SP, Danulidis K, Rallis GE, Tsioulos DI, Chatzivasiliou A, et al. A new approach to accurate diagnosis of acute appendicitis. World J Surg. 2005;29(9):1151-6.
8. 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):1540-5.
9. Malla BR, Bhandari D, Oli P. Comparison of Alvarado and Tzanakis scoring systems in the diagnosis of acute appendicitis. J Nepal Health Res Counc. 2019;17(4):450-4.
10. 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(10):2278-85.
11. van Randen A, Bipat S, Zwinderman AH, Ubbink DT, Stoker J, Boermeester MA. Acute appendicitis: meta-analysis of diagnostic performance of CT and graded compression US related to prevalence of disease. Radiology. 2008;249(1):97-106.
12. Gorter RR, Eker HH, Gorter-Stam MA, Abis GS, Acharya A, Ankersmit M, et al. Diagnosis and management of acute appendicitis. EAES consensus development conference 2015. Surg Endosc. 2016;30(11):4668-90.
13. 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.