International Journal of Research in Medical Sciences
Qureshi AS et al. Int J Res Med Sci. 2026 Sep;14(9):4227-4229
www.msjonline.org
pISSN 2320-6071 | eISSN 2320-6012
DOI: https://dx.doi.org/10.18203/2320-6012.ijrms20263153

Commentary

Methodological considerations in the use of systemic inflammatory indices for risk stratification in acute appendicitis

Ataah S. Qureshi*, Darren L. Pereira

MGM Medical College and Hospital, Kamothe, Navi Mumbai, Maharashtra, India

Received: 15 July 2026
Revised: 17 August 2026
Accepted: 25 August 2026

***Correspondence:**
Ataah S. Qureshi,
E-mail: ataahqureshi513@gmail.com

ABSTRACT

Systemic inflammatory indices derived from the complete blood count, including the neutrophil-to-lymphocyte ratio, platelet-to-lymphocyte ratio, and systemic immune-inflammation index, have gained interest as low-cost adjuncts for diagnosing and risk-stratifying acute appendicitis. Using a recent retrospective study by Örün et al as an illustrative example, this commentary examines four recurring methodological issues in this literature: validation against a non-independent reference standard that shares laboratory input with the indices themselves; small subgroup sizes that yield imprecise discrimination estimates; substantial heterogeneity in proposed cut-off values across studies and populations; and incomplete reporting of effect sizes, such as regression coefficients presented without accompanying odds ratios or confidence intervals. These issues recur across much of the published literature, including existing systematic reviews and meta-analyses, and limit the extent to which encouraging single-study findings can be translated into validated bedside tools. We argue that future studies should validate these indices against independent reference standards such as histopathology, employ adequately powered analyses, pre-specify and externally validate cut-off values, and adhere to established reporting guidelines such as the TRIPOD statement. Addressing these structural weaknesses would strengthen the evidence base supporting the clinical adoption of systemic inflammatory indices in acute appendicitis.

Keywords: Acute appendicitis, Alvarado score, Neutrophil-to-lymphocyte ratio, Systemic immune-inflammation index, Risk stratification, Methodology

INTRODUCTION

Acute appendicitis remains one of the most common surgical emergencies of the abdomen, and its diagnosis continues to rely heavily on clinical scoring systems such as the Alvarado score, often supplemented by imaging.[1] Over the past decade, systemic inflammatory indices derived from the routine complete blood count, including the neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and systemic immune-inflammation index (SII), have attracted considerable interest as low-cost adjuncts for diagnosing and risk-stratifying appendicitis, and have now been evaluated in several systematic reviews and meta-analyses.[2-4] A recent retrospective study by Örün et al examined the association between these indices and Alvarado score-based risk categories in patients with computed tomography-confirmed appendicitis, reporting encouraging discriminative performance for several indices.[5] While individually informative, this growing body of literature illustrates several recurring methodological issues that, in our view, merit closer attention before these indices can be reliably translated into clinical practice. We discuss four such issues below, using the study by Örün et al as a recent and illustrative example.

Validation against a non-independent reference standard

Many studies in this area, including that of Örün et al, validate inflammatory indices against Alvarado score-based risk categories.[5] The Alvarado score itself

incorporates the leukocyte count as a scored component, and NLR and SII are both derived from the same complete blood count from which that leukocyte count is obtained.[1] Demonstrating that these indices correlate with Alvarado-defined categories may therefore partly reflect shared laboratory input rather than validation against a genuinely independent reference standard, such as histopathological confirmation of appendicitis and its severity (uncomplicated versus perforated/gangrenous).

Small subgroup sizes and unstable discrimination estimates

Cohort sizes in this literature are frequently small relative to the number of subgroup analyses performed. In the Örün et al. cohort of 69 patients, further stratification into three Alvarado-based risk categories, with an additional post-hoc analysis restricted to scores ≥5, produced area-under-the-curve (AUC) estimates with wide confidence intervals (e.g., NLR AUC 0.81, 95% CI: 0.71–0.91, in the high-risk subgroup).[5] AUC estimates derived from such small strata are inherently imprecise, a concern that persists even at the meta-analytic level, where pooled estimates for NLR across dozens of primary studies still show considerable between-study heterogeneity in reported sensitivity, specificity, and cut-off values.[2,3,6]

Heterogeneity of proposed cut-off values

Reported reference ranges and cut-offs for the same index vary considerably across studies, limiting cross-study comparability. The mean SII reported by Örün et al (11,131±1,543) is markedly higher than ranges reported in a pediatric cohort by Siki et al, where SII values were in the low hundreds to low thousands.[7] Similarly, meta-analyses of NLR in appendicitis have themselves highlighted substantial between-study heterogeneity in proposed cut-off values, with one adult meta-analysis proposing an NLR cut-off of 4.7 for diagnosis of appendicitis (AUC 0.96), while a pediatric-focused meta-analysis reported different pooled accuracy estimates across its included studies.[2,4] Some of this variability may reflect genuine population differences, but a recent methodological review has highlighted that thresholds derived and tested within the same dataset, without independent validation, are prone to "threshold overfitting," which can inflate apparent diagnostic performance and undermine cross-study comparability.[8]

Transparent reporting of effect sizes

Effect sizes are often reported in forms that are difficult for clinicians to interpret. Örün et al report raw logistic regression coefficients for NLR, PLR, and SII (e.g., SII coefficient=0.0016, p=0.008) without corresponding odds ratios or confidence intervals.[5] Adherence to established reporting standards for prediction and prognostic studies, such as the TRIPOD statement, alongside internal validation methods such as bootstrapping, would strengthen the evidentiary basis of this literature.[9] These four issues, along with their potential impact and suggested improvements, are summarized in Table 1.

Table 1: Summary of methodological issues in the use of systemic inflammatory indices for risk stratification in acute appendicitis, illustrated using Örün et al.

Methodological issue	Illustrative example from Örün et al	Potential implication	Suggested methodological improvement
Validation against a non-independent reference standard	Systemic inflammatory indices were evaluated against Alvarado score-based risk categories, although the Alvarado score itself incorporates leukocyte count.	The observed association may partly reflect shared laboratory components rather than true independent validation, potentially overestimating diagnostic performance.	Validate inflammatory indices against independent reference standards, such as histopathological diagnosis and disease severity (uncomplicated versus complicated appendicitis).
Small subgroup sizes and unstable discrimination estimates	A cohort of 69 patients was further divided into three Alvarado score categories, with additional subgroup analyses.	Small subgroup sizes produce wider confidence intervals and less precise estimates of diagnostic accuracy, limiting the robustness of conclusions.	Conduct adequately powered studies, minimise unnecessary subgroup analyses, and externally validate diagnostic models in larger cohorts.
Heterogeneity of proposed cut-off values	Considerable variation in reported NLR, PLR, and SII cut-off values exists across studies and patient populations.	Poor comparability between studies reduces external validity and limits routine clinical implementation.	Use pre-specified cut-off values, validate them in independent external cohorts, and work towards standardised reporting across studies.
Incomplete reporting of effect sizes	Logistic regression coefficients were presented without corresponding odds ratios or 95% confidence intervals.	Clinicians may find it difficult to interpret the magnitude and clinical relevance of reported associations.	Report odds ratios with 95% confidence intervals and adhere to established reporting guidelines, including the TRIPOD statement.

DISCUSSION

Taken together, these four issues, namely non-independent validation, small subgroup sizes, cut-off heterogeneity, and inconsistent effect-size reporting, are not unique to any single study; they recur across much of the literature on inflammatory indices in appendicitis, including within existing systematic reviews and meta-analyses.[2-4] This does not diminish the biological plausibility or clinical promise of these indices. Systemic inflammatory indices are inexpensive, rapidly available from a test ordered on virtually every patient with suspected appendicitis, and grounded in a coherent pathophysiological rationale linking neutrophilia and lymphopenia to the intensity of the systemic inflammatory response. However, the field's ability to translate this promise into a validated bedside tool depends on addressing these structural weaknesses. Future primary studies would benefit from validation against independent reference standards, adequately powered subgroup analyses (or a single well-powered stratum), pre-registered cut-offs validated in separate cohorts, and effect-size reporting consistent with the TRIPOD statement.[9]

CONCLUSION

Inflammatory indices hold genuine promise as accessible tools for risk stratification in acute appendicitis. Addressing the recurring issues of validation circularity, subgroup sample size, cut-off standardization, and transparent effect-size reporting, illustrated here using Örün et al alongside the broader literature, would substantially strengthen the evidence base and facilitate the eventual clinical adoption of these tools.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Alvarado A. A practical score for the early diagnosis of acute appendicitis. Ann Emerg Med. 1986;15:557-64.
2. Hajibandeh S, Hajibandeh S, Hobbs N, Mansour M. Neutrophil-to-lymphocyte ratio predicts acute appendicitis and distinguishes between complicated and uncomplicated appendicitis: a systematic review and meta-analysis. Am J Surg. 2020;219(1):154-63.
3. Yang J, Liu C, He Y, Cai G. The role of neutrophil-to-lymphocyte ratio (NLR) and platelet-to-lymphocyte ratio (PLR) in the diagnosis of acute appendicitis: a systematic review and meta-analysis. Front Pediatr. 2022;10:1012962.
4. Eun S, Ho IG, Bae GE, Kim H, Koo CM, Kim MK, et al. Neutrophil-to-lymphocyte ratio for the diagnosis of pediatric acute appendicitis: a systematic review and meta-analysis. Eur Rev Med Pharmacol Sci. 2021;25(22):7097-107.
5. Örün S, Demirel AT, Çağman U, Yıldırım B, Balkaya M, Çelik M, et al. The Association of Systemic Inflammatory Indices with Alvarado Score-based Risk Stratification in Acute Appendicitis: A Retrospective Study. Eurasian J Emerg Med. 2026;25(1):275-80.
6. Hajian-Tilaki K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. Caspian J Intern Med. 2013;4:627-35.
7. Siki FO, Sarıkaya M, Gündüz M, Sekmenli T. Evaluation of the systemic immune inflammation index and the systemic inflammatory response index as new markers for the diagnosis of acute appendicitis in children. Ann Saudi Med. 2023;43(5):329-38.
8. Arredondo Montero J. A tailored fit that doesn't fit all: the problem of threshold overfitting in diagnostic studies. Diagnosis (Berl). 2025;16.
9. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.

Cite this article as: Qureshi AS, Pereira DL. Methodological considerations in the use of systemic inflammatory indices for risk stratification in acute appendicitis. Int J Res Med Sci 2026;14:4227-9.