Annals of Gastroenterology (2026) **39,** 379-382

Comments on modeling strategy and data handling in a meta-analysis of anti-integrin αvβ6 for primary sclerosing cholangitis

Javier Arredondo Montero

Complejo Asistencial Universitario de León, León, Spain

The meta-analysis by Papadakos *et al* [1], although clinically relevant, raises several methodological issues. Despite reporting a bivariate random-effects approach, the forest plots and summary receiver operating characteristic (SROC) curve are consistent with non-hierarchical univariate pooling. The SROC is symmetric, lacks confidence or prediction regions, and omits variance components and correlation parameters. Additionally, the pooled sensitivity (62.3%, 95% confidence interval [CI] 59.6-65.0%) and specificity (87.3%, 95%CI 86.6-88.0%) show implausibly narrow, perfectly symmetric confidence intervals despite substantial heterogeneity and only four studies. Under genuine between-study variability,

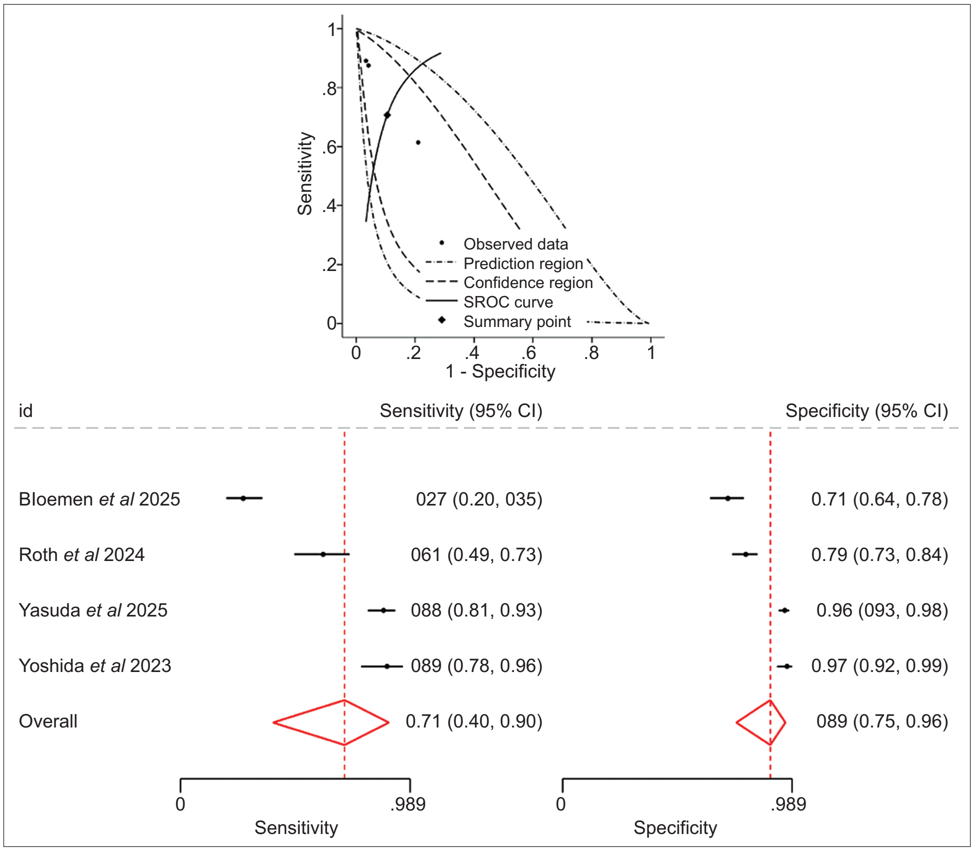

Figure 1 Hierarchical diagnostic meta-analysis of anti-integrin αvβ6 for primary sclerosing cholangitis (PSC). Above: Hierarchical summary receiver operating characteristic (HSROC) plot derived from a hierarchical random-effects model (STATA 19, *metadta*), displaying individual study estimates, the summary operating point, and the corresponding 95% confidence (CI) and 95% prediction regions, thus jointly accounting for the intrinsic correlation between sensitivity and specificity and their between-study heterogeneity. Bottom: Forest plots of sensitivity and specificity from the same hierarchical model. Pooled estimates were sensitivity 0.71 (95%CI 0.40-0.90) and specificity 0.89 (95%CI 0.75-0.96), with substantial between-study variance (τ^2 Se=1.65; τ^2 Sp=1.07). Compared with the previously reported pooled sensitivity (≈0.62), the hierarchical approach yields a materially higher central estimate, while simultaneously demonstrating wide uncertainty and considerable between-study variability. This reinforces the importance of joint modeling and prediction regions when interpreting diagnostic accuracy across heterogeneous settings

precision would be lower; such intervals suggest univariate Wald-type estimation rather than a hierarchical model [2,3]. Hierarchical models jointly account for sensitivity–specificity correlation and heterogeneity, yielding more appropriate uncertainty estimates. In my reanalysis using the reported 2×2 data and a hierarchical random-effects model (Stata 19, *metadta*), pooled sensitivity was higher (0.71 vs. 0.62) but with wide confidence intervals (0.40-0.90), reflecting substantial between-study variance (Fig. 1), a finding that underscores the inherent uncertainty of the estimates, which should be interpreted cautiously given the small number of studies.

Second, the primary sclerosing cholangitis plus inflammatory bowel disease (PSC+IBD) subgroup analysis relies on constructed diagnostic counts. For studies lacking IBD-only controls, false positives and true negatives were derived using an external specificity estimate (ulcerative colitis meta-analysis), assuming a 1:1 ratio. This is not imputation within observed data but creation of hypothetical patients under assumed performance, altering the evidentiary basis and risking artificially precise, model-driven estimates.

Third, thresholds were defined as mean + X standard deviations (SD) within each cohort, making them data-derived rather than prespecified. Under QUADAS-2, this implies high risk of bias in the Index Test domain and potential overfitting [3,4]. Moreover, threshold uniformity is inaccurately reported: while the review states mean +3 SD across studies, Roth *et al* used mean +2 SD, indicating threshold heterogeneity and further supporting hierarchical modeling.

Adherence to the Cochrane and PRISMA-DTA standards [5,6] is essential for valid, transparent, and clinically applicable meta-analyses.

References

1. Papadakos SP, Vogli S, Argyrou A, et al. Serum anti-integrin αvβ6 autoantibodies for diagnosis of primary sclerosing cholangitis: a systematic review and meta-analysis. *Ann Gastroenterol* 2026;**39**:40-47.
2. Arredondo Montero J. Diagnostic test accuracy meta-analysis: a

practical guide to hierarchical models. *J Surg Res* 2025;**315**:768-781.
3. Harbord RM, Whiting P, Sterne JA, Egger M, Deeks JJ, Shang A, Bachmann LM. An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. *J Clin Epidemiol* 2008;**61**:1095-1103.
4. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**:529-536.
5. Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y (editors). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 2.0 (updated July 2023). Cochrane, 2023. Available from: https://training.cochrane.org/handbook-diagnostic-test-accuracy/current [Accessed 2 April 2026].
6. McInnes MDF, Moher D, Thombs BD, et al; and the PRISMA-DTA Group. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA* 2018;**319**:388-396.

Pediatric Surgery Department, Complejo Asistencial Universitario de León, León, Spain

Conflict of Interest: None

Correspondence to: Javier Arredondo Montero, MD, PhD, Department of Pediatric Surgery, Complejo Asistencial Universitario de León, c/ Altos de Nava s/n, 24008 León, Castilla y León, Spain,
e-mail: jarredondo@saludcastillayleon.es,
javier.montero.arredondo@gmail.com

Received 24 February 2026; accepted 19 March 2026;
published online 23 April 2026

DOI: https://doi.org/10.20524/aog.2026.1057