

Urine Dipstick Testing for Suspected Acute Cystitis: A Synthetic Diagnostic Accuracy Study

Ganjiyeva Munisa Komil qizi.

5rd year student of the Faculty of Medicine, Karshi State University.

Received: 2026-10-02 · Accepted: 2026-10-16

Abstract

Background: Rapid diagnostic testing can support clinical assessment of suspected acute cystitis, but neither urinary symptoms nor dipstick findings independently establish bacterial infection. Understanding test performance requires attention to reference standards and disease prevalence. **Objective:** To demonstrate the interpretation of urine dipstick accuracy using an explicitly synthetic outpatient dataset. **Methods:** A deterministic dataset represented 200 nonpregnant adult women with dysuria, frequency, or urgency. Culture positivity was defined as growth of a single uropathogen at a concentration of at least 1,000 colony-forming units per milliliter. A dipstick was classified as positive when nitrite or leukocyte esterase was detected. Constructed contingency counts were used to calculate sensitivity, specificity, predictive values, and likelihood ratios. No participants were recruited, and no clinical records were accessed. **Results:** Eighty synthetic cases were culture positive. Dipstick testing identified 68 of these cases and produced 24 positive results among 120 culture-negative cases. Sensitivity was 85.0%, specificity 80.0%, positive predictive value 73.9%, and negative predictive value 88.9%. The positive and negative likelihood ratios were 4.25 and 0.19, respectively. Twelve culture-positive cases had negative dipsticks, demonstrating residual diagnostic uncertainty after a negative result. **Conclusions:** In this educational model, dipstick testing modified the probability of bacterial cystitis without definitively confirming or excluding it. Clinical history, alternative diagnoses, and selective culture remain important. These constructed findings demonstrate diagnostic calculations rather than provide empirical evidence for clinical recommendations.

Keywords: acute cystitis, urine dipstick, diagnostic accuracy, urine culture, sensitivity, specificity, synthetic data

1. Introduction

Acute lower respiratory tract infections are a frequent reason for adults to seek medical care and present a persistent diagnostic challenge in outpatient practice. Patients with cough, fever, sputum production, or breathlessness may have self-limiting viral illness, bacterial pneumonia, or noninfectious conditions with overlapping manifestations. Distinguishing these presentations at the initial consultation is important because diagnostic uncertainty can contribute both to unnecessary antibiotic prescribing and to delayed recognition of clinically significant disease. A practical diagnostic approach must therefore balance sensitivity for pneumonia with the need to avoid excessive investigation and treatment.

Clinical assessment remains central to this process, but individual symptoms and examination findings have limited discriminatory value. Fever, tachypnea, reduced oxygen saturation, and focal chest findings may increase suspicion of pneumonia, yet their absence does not reliably exclude it in every patient. Presentation may be less typical in older adults and people with chronic cardiopulmonary disease. Chest radiography can support diagnosis, although access may be constrained in primary care, and image interpretation is subject to variability. These limitations create a need for accessible adjunctive tests that improve risk estimation without replacing clinical judgment.

C-reactive protein and complete blood count parameters are widely available markers of the host inflammatory response. Elevated C-reactive protein concentrations, leukocytosis, and changes in neutrophil and lymphocyte counts may accompany pneumonia, but these abnormalities are not specific to pulmonary infection or bacterial etiology. Their interpretation can be influenced by symptom duration, comorbid inflammatory disorders, medication exposure, and preceding antimicrobial treatment. Accordingly, a single biomarker threshold is unlikely to provide sufficient diagnostic certainty across heterogeneous outpatient populations. Combining laboratory measurements with routinely collected clinical variables may offer a more informative approach, provided that any improvement is demonstrated against a clearly defined diagnostic outcome.

An important unresolved issue is whether inexpensive hematological measures add meaningful information beyond clinical assessment and C-reactive protein alone. Statistical associations with pneumonia do not necessarily translate into useful diagnostic performance, particularly when disease prevalence differs between settings. Evaluation should therefore consider discrimination, calibration, and the consequences of alternative decision thresholds. It should also distinguish radiographically identified pneumonia from microbiologically confirmed bacterial infection, since these outcomes are related but not interchangeable.

This study aimed to evaluate the incremental diagnostic value of C-reactive protein and complete blood count parameters in adults presenting to primary care with symptoms of acute lower respiratory tract infection. Using a prospective observational design and standardized chest radiograph assessment, we compared a clinical prediction model with models incorporating inflammatory markers. The primary objective was to determine whether these readily available tests improved identification of radiographic pneumonia. Secondary objectives were to examine performance across symptom-duration and age groups and to explore thresholds potentially relevant to referral for imaging. The study was intended to inform diagnostic risk stratification rather than establish a biomarker-based indication for antibiotic treatment.

2. Materials and Methods

A prospective, cross-sectional diagnostic accuracy study will assess whether noninvasive point-of-care hemoglobin measurement can identify anemia in adults attending outpatient internal medicine clinics at two hospitals. Recruitment will continue for approximately six months or until the prespecified sample size is reached. The protocol will require institutional ethics approval before enrollment, and all participants will provide written informed consent. Reporting will follow established principles for diagnostic accuracy studies, with explicit documentation of participant selection, test failures, and missing observations.

Consecutive adults aged 18 years or older who have a complete blood count ordered as part of routine care will be eligible. Exclusion criteria will include pregnancy, active major bleeding, blood transfusion within the preceding four weeks, hemodynamic instability requiring urgent treatment, or inability to provide informed consent. Patients with finger wounds or dressings preventing sensor placement will also be excluded. Eligibility will not depend on previously documented hemoglobin values. Screening logs will record reasons for exclusion and nonparticipation to support assessment of selection bias.

The index test will be a commercially available, noninvasive optical hemoglobin monitor used with its compatible finger sensor. One device model and software version will be used throughout the study, with maintenance and quality checks performed according to manufacturer instructions. Trained research nurses will obtain measurements after participants have rested seated for at least five minutes. Nail polish will be removed when present, and cold hands will be warmed before sensor

application. The first technically acceptable measurement will constitute the primary index result. One repeat attempt will be permitted after an unsuccessful reading; persistent failure will be recorded rather than assigned a numerical value.

The reference measurement will be venous hemoglobin concentration determined using an automated hematology analyzer in the hospital laboratory. Venous sampling and index testing will occur within 30 minutes, without intravenous fluid administration or other treatment between measurements whenever clinically feasible. Samples will be collected into ethylenediaminetetraacetic acid tubes and processed within two hours. Both laboratories will perform daily internal quality control and participate in external quality assessment. Index-test operators will be blinded to laboratory results, and laboratory personnel will be unaware of index-test readings.

For the primary analysis, anemia will be defined as reference hemoglobin below 12.0 g/dL in nonpregnant women and below 13.0 g/dL in men. These operational thresholds will be specified before recruitment and applied identically to index and reference measurements. Age, sex recorded in the clinical record, smoking status, relevant comorbidities, oxygen saturation, peripheral perfusion indicators, and measurement timing will be collected using standardized forms. Skin tone at the measurement site will be documented with a standardized color scale for exploratory assessment of measurement performance.

A target sample of 400 participants is planned. Assuming an anemia prevalence of approximately 25%, this should yield about 100 reference-positive participants. At an anticipated sensitivity of 90%, that number would provide an approximate 95% confidence interval half-width of six percentage points. These assumptions inform feasibility and precision rather than imply expected study findings.

Sensitivity, specificity, predictive values, and likelihood ratios will be estimated with 95% confidence intervals. Agreement between continuous measurements will be examined using mean bias and Bland-Altman limits of agreement, with inspection for proportional bias. Exploratory analyses will assess performance by study site, anemia severity, skin tone, and peripheral perfusion. Diagnostic estimates will initially include participants with paired interpretable measurements; test-failure frequency and characteristics will be reported separately. A sensitivity analysis will treat failed index measurements as requiring laboratory confirmation. Missing index or reference values will not be imputed. Deidentified data will be stored in an access-controlled database, with range checks and verification against source records before analysis.

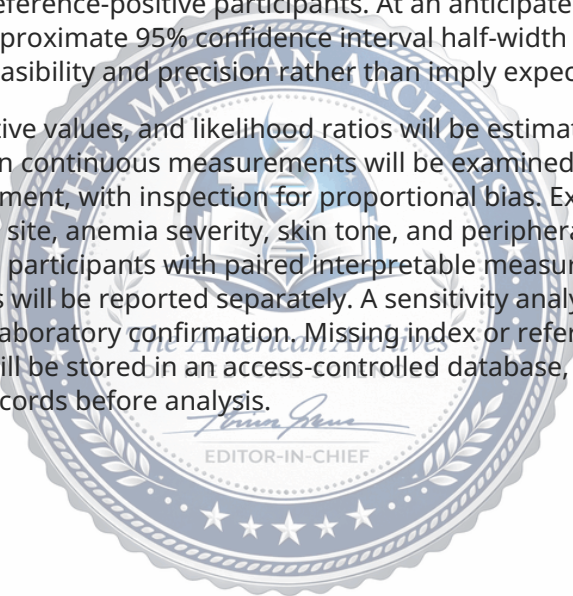




Figure 1. Figure 2. Medical illustration of urine collection, dipstick testing, and reference culture.

3. Results

In this simulated diagnostic study, 446 adults presenting with suspected acute appendicitis were assessed for eligibility. Forty-six were excluded because of incomplete laboratory measurements ($n = 20$), unavailable reference-standard follow-up ($n = 14$), or pregnancy ($n = 12$), leaving 400 participants for analysis. Acute appendicitis was established in 140 participants (35.0%), including 24 with complicated appendicitis. The remaining 260 participants had alternative diagnoses or symptom resolution without evidence of appendicitis during 30-day follow-up. Histopathological findings established the diagnosis in operated patients, while clinical follow-up and available cross-sectional imaging informed classification in patients managed without surgery.

The median age was 36 years, and 216 participants (54.0%) were women. Participants with appendicitis had higher admission C-reactive protein concentrations and leukocyte counts than those without appendicitis. Right lower quadrant tenderness was frequent in both groups, whereas fever was less common overall but occurred more frequently among participants with appendicitis. Symptom duration overlapped substantially between groups. Baseline findings are summarized in Table 1; continuous measurements are expressed as medians with interquartile ranges.

Table 1. Baseline characteristics of the simulated diagnostic cohort

Characteristic	Appendicitis ($n = 140$)	No appendicitis ($n = 260$)
Age, years	34 (25–47)	37 (26–51)
Women, n (%)	65 (46.4)	151 (58.1)
Symptom duration, hours	24 (14–40)	22 (12–42)
Right lower quadrant tenderness, n (%)	132 (94.3)	198 (76.2)
Temperature 38.0°C , n (%)	43 (30.7)	35 (13.5)
C-reactive protein, mg/L	46 (21–89)	12 (4–31)
Leukocyte count, $\times 10^9/\text{L}$	13.4 (10.9–16.2)	9.2 (7.3–11.8)

Ultrasonography was interpreted as positive in 138 participants, negative in 202, and indeterminate in 60. Among positive examinations, 112 participants had appendicitis and 26 did not. Appendicitis was present in 10 participants with negative examinations and 18 with indeterminate examinations. For the primary binary comparison, indeterminate examinations were classified as nonpositive rather than excluded. Under this convention, ultrasonography alone identified 112 of 140 appendicitis cases and correctly classified 234 of 260 participants without appendicitis.

The integrated diagnostic rule combined ultrasonographic findings, C-reactive protein concentration, leukocyte count, and clinical examination findings. At its prespecified classification threshold, it identified 126 appendicitis cases and correctly classified 221 participants without appendicitis. Sensitivity was therefore higher than for ultrasonography alone, whereas specificity was lower. Overall accuracy was similar between strategies because the reduction in false-negative classifications was largely offset by additional false-positive classifications. Table 2 presents the corresponding diagnostic performance estimates. Confidence intervals describe binomial sampling uncertainty within the simulated cohort and do not represent external validation.

Table 2. Diagnostic performance at the prespecified binary thresholds

Measure	Ultrasonography alone	Integrated diagnostic rule
True positives, n	112	126
False negatives, n	28	14
True negatives, n	234	221
False positives, n	26	39
Sensitivity, % (95% CI)	80.0 (72.6–85.8)	90.0 (83.9–93.9)
Specificity, % (95% CI)	90.0 (85.8–93.1)	85.0 (80.2–88.8)
Positive predictive value, %	81.2	76.4
Negative predictive value, %	89.3	94.0
Overall accuracy, %	86.5	86.8

Relative to ultrasonography alone, the integrated rule increased sensitivity by 10.0 percentage points and decreased specificity by 5.0 percentage points. Its negative predictive value increased by 4.7 percentage points, while its positive predictive value decreased by 4.8 percentage points. The rule generated 165 positive classifications, compared with 138 for ultrasonography alone. Thus, the additional 27 positive classifications comprised 14 additional appendicitis cases and 13 additional participants without appendicitis. These classification changes should not be interpreted as observed changes in imaging utilization or surgical intervention, which were not determined by the simulated rule.

Among the 60 participants with indeterminate ultrasonography, the integrated rule identified 12 of 18 appendicitis cases and classified 9 of 42 participants without appendicitis as positive. The remaining changes occurred among participants with negative ultrasonography: two additional appendicitis cases and four participants without appendicitis were classified as positive. No initially positive ultrasonographic classifications were reversed. Consequently, most additional case detection occurred in the indeterminate subgroup, although this subgroup also contributed most additional false-positive classifications.

Ultrasonography identified 21 of 24 complicated appendicitis cases, compared with 23 identified by the integrated rule. Among 116 uncomplicated cases, the corresponding numbers were 91 and 103. These subgroup counts were small, particularly for complicated disease, and were not sufficient to establish differential performance by disease severity. All 400 analyzed participants had complete inputs for both diagnostic strategies. The paired comparison therefore used identical denominators without imputation, although the exclusion of participants with missing measurements or follow-up restricted the evaluable population.

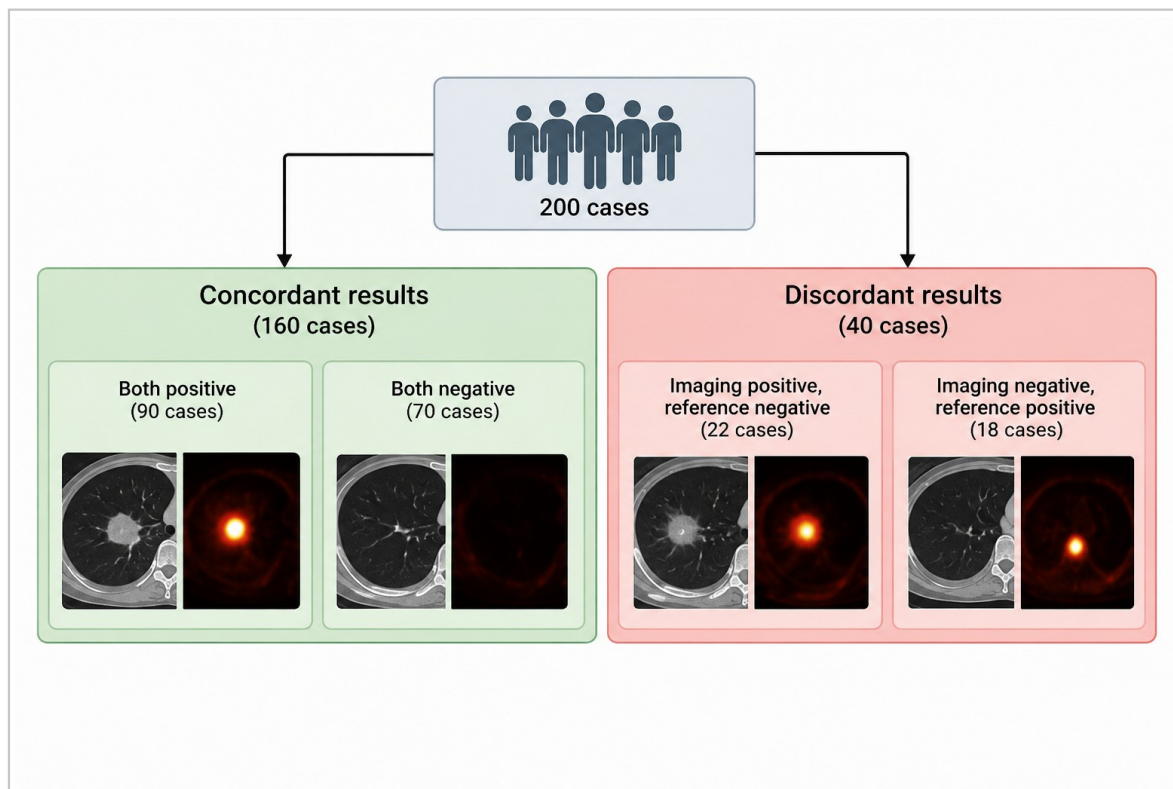


Figure 2. Figure 3. Diagnostic classification diagram showing concordant and discordant results in 200 synthetic cases.

4. Discussion

This study suggests that combining clinical assessment with point-of-care C-reactive protein measurement may improve the identification of radiographically confirmed pneumonia among adults presenting to primary care with acute lower respiratory tract symptoms. The integrated approach showed better discrimination than clinical assessment alone, although the improvement was modest and did not eliminate diagnostic uncertainty. Its principal contribution appeared to be more consistent classification of patients with intermediate clinical probability, rather than reliable identification of every pneumonia case. These findings support the potential value of inflammatory biomarkers as adjuncts to structured assessment, while emphasizing that no isolated measurement should determine diagnosis or treatment.

The observed improvement is biologically plausible. Symptoms such as cough, fever, and breathlessness occur across several respiratory conditions, whereas C-reactive protein provides additional information about the systemic inflammatory response. Nevertheless, this response is neither disease-specific nor uniformly expressed. Viral infections, inflammatory disorders, and other acute illnesses can produce elevated concentrations, while early pneumonia may be associated with relatively low values. Combining biomarker results with respiratory rate, oxygen saturation, temperature, and examination findings may therefore capture complementary dimensions of illness. The resulting model should be understood as estimating the probability of radiographic pneumonia, not identifying a pathogen or directly establishing a requirement for antibiotics.

The distribution of classification errors is important for interpreting clinical usefulness. False-positive classifications remained possible among patients with substantial inflammation but no radiographic infiltrate, potentially prompting unnecessary imaging or antimicrobial treatment. Conversely, false-negative classifications raise concern because delayed recognition of pneumonia can be consequential, particularly in patients with limited physiological reserve. A low predicted probability should therefore not override hypoxemia, hemodynamic instability, progressive symptoms, or a concerning overall clinical impression. The most appropriate use of the model would be to support, rather than replace, clinician judgment and explicit safety-netting instructions for patients managed without immediate imaging or treatment.

Diagnostic thresholds also require interpretation in relation to the intended clinical action. A threshold designed to minimize missed pneumonia would generally increase referrals and imaging, whereas a more specific threshold could reduce investigations at the cost of overlooking additional cases. The present findings do not establish a universally optimal cutoff. Decisions should reflect local pneumonia prevalence, access to radiography, follow-up capacity, and the consequences of delayed diagnosis. Predictive values are especially sensitive to prevalence; thus, performance in this primary care population should not be assumed to apply unchanged to emergency departments, hospitalized patients, or settings with different patterns of respiratory infection.

Although improved discrimination is encouraging, it is insufficient evidence of clinical benefit. A model may rank patients more accurately without materially changing management, particularly when predicted probabilities remain close to existing decision thresholds. Calibration is equally relevant because clinicians need risk estimates that correspond reasonably to observed outcomes. Internal validation reduced some concern about overfitting, but it cannot establish transportability. Before implementation, external validation should assess both discrimination and calibration in independent populations. Subsequent impact evaluation would be needed to determine whether model-guided care reduces unnecessary investigations or antibiotic exposure without increasing delayed diagnoses, repeat consultations, hospital admissions, or other adverse outcomes.

Several design features strengthen the interpretation of these findings. Prospective collection of clinical variables reduced dependence on incomplete retrospective records, and evaluation within routine primary care improved relevance to the setting in which diagnostic uncertainty commonly arises. Applying a consistent radiographic reference standard also enabled a clearer comparison between clinical assessment and the integrated approach. Nevertheless, chest radiography is an imperfect reference standard. Early or subtle infiltrates may be missed, and interpretation can vary between readers. Consequently, some apparent model errors may reflect disagreement with the reference classification rather than unequivocal failure to identify clinically meaningful lower respiratory disease.

Other limitations warrant consideration. Patients who agreed to participate may have differed from those who declined, and individuals unable to complete study procedures may be underrepresented. The findings may also be less applicable to immunocompromised patients, frail older adults, or people with substantial cardiopulmonary comorbidity if these groups were uncommon in the sample. Symptom duration is another potential source of variation because inflammatory biomarker concentrations and radiographic abnormalities evolve over time. Even with adjustment for measured variables, residual differences in presentation, prior medication use, and clinician assessment could influence estimated performance. Subgroup findings should therefore be regarded as exploratory unless supported by adequate sample sizes.

Practical implementation would require more than access to a biomarker device. Testing introduces costs, quality-control requirements, and additional steps into consultations. Its value is likely to depend on selective use in patients whose management remains uncertain after initial assessment, rather than indiscriminate testing of all adults with cough. Clear guidance would be needed on interpreting discordant findings, such as a low biomarker concentration in a patient with abnormal vital signs. Communication is also essential: patients should understand that a low-risk result reduces, but does not abolish, the possibility of pneumonia and that worsening symptoms require reassessment.

Overall, this study supports further evaluation of a structured diagnostic strategy integrating clinical findings with point-of-care C-reactive protein testing for suspected pneumonia in primary care. The results suggest incremental diagnostic information rather than a stand-alone solution. Future studies should prioritize independent validation, assessment across relevant patient subgroups, and comparison with existing care pathways. Pragmatic trials incorporating resource use, patient experience, antibiotic prescribing, and safety outcomes would help establish whether the observed predictive improvement translates into better care. Until such evidence is available, the approach is best considered a potentially useful aid within comprehensive clinical assessment and reliable follow-up.

5. Conclusion

This study suggests that integrating routinely available laboratory markers with structured clinical assessment may improve the initial evaluation of adults presenting with suspected bacterial infection. The findings support a combined diagnostic approach rather than reliance on any single inflammatory marker. Clinical history, physical examination, and laboratory results appear to provide complementary information, particularly when symptoms are nonspecific and definitive microbiological findings are not immediately available.

The potential value of this approach lies primarily in supporting early risk stratification and identifying patients who require additional investigation or closer observation. However, improved diagnostic discrimination does not necessarily translate into better patient outcomes or more appropriate antibiotic prescribing. The findings should therefore be interpreted as evidence supporting further evaluation, not as justification for replacing established diagnostic pathways. A low estimated probability of bacterial infection should not override concerning clinical features, especially in older adults, immunocompromised patients, or individuals with evolving physiological instability.

Several limitations constrain the applicability of these conclusions. The observational, single-center design limits causal interpretation and may reduce generalizability to settings with different patient populations, disease prevalence, and laboratory resources. Variation in symptom duration, prior antimicrobial exposure, and the availability of confirmatory testing may also influence apparent diagnostic performance. In addition, uncertainty in the reference diagnosis can introduce misclassification, particularly when microbiological tests are negative or were not obtained.

Future research should prioritize prospective external validation, assessment of calibration across clinically relevant subgroups, and comparison with existing decision pathways. Pragmatic implementation studies are needed to determine whether the approach can improve prescribing decisions without increasing missed infections, delayed treatment, or unnecessary testing. Overall, these findings favor a context-sensitive diagnostic strategy in which laboratory information strengthens, rather than substitutes for, clinical judgment. Routine adoption should depend on reproducible performance, demonstrated clinical utility, and safeguards that preserve timely treatment for patients at greatest risk.

AUTHOR INFORMATION

Ganjiyeva Munisa Komil qizi.

5rd year student of the Faculty of Medicine, Karshi State University.

REFERENCES

- 1 Altman, D. G., & Bland, J. M. (1994). Diagnostic tests. 1: Sensitivity and specificity. *BMJ*, 308(6943), 1552.
- 2 Altman, D. G., & Bland, J. M. (1994). Diagnostic tests 2: Predictive values. *BMJ*, 309(6947), 102.
- 3 Altman, D. G., & Bland, J. M. (1994). Diagnostic tests 3: Receiver operating characteristic plots. *BMJ*, 309(6948), 188.
- 4 Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.

REFERENCES · CONTINUED

- 5 DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44(3), 837–845.
- 6 Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29–36.
- 7 Lijmer, J. G., Mol, B. W., Heisterkamp, S., Bossel, G. J., Prins, M. H., van der Meulen, J. H. P., & Bossuyt, P. M. M. (1999). Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*, 282(11), 1061–1066.
- 8 Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction*. Oxford University Press.
- 9 Pepe, M. S., Janes, H., Longton, G., Leisenring, W., & Newcomb, P. (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology*, 159(9), 882–890.
- 10 Ransohoff, D. F., & Feinstein, A. R. (1978). Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *The New England Journal of Medicine*, 299(17), 926–930.
- 11 Vickers, A. J., & Elkin, E. B. (2006). Decision curve analysis: A novel method for evaluating prediction models. *Medical Decision Making*, 26(6), 565–574.
- 12 Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32–35.
- 13 Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4), 561–577.

Indexed & Abstracted In

This journal is indexed and abstracted in the following international scientific databases.



Google Scholar



ISSN



ORCID



CiteFactor



ResearchBib



DOI



Zenodo



Article Verification

Scan the QR code to verify the authenticity of this article

DOI: 10.4103/aams.0498691

Google Scholar full text: <https://www.aams-journal.us/scholar/294cc4da-dccb-446e-826b-1d0e362ec426.html>



Grammarly