

SUPPLEMENTAL DATA

Table S1. Table of urinary tract infection (UTI) flagging vs culture

Index test result	Positive culture (N = 440)	Negative culture (N = 445)	Total
UTI flagging positive (UTI)	295	35	330
UTI flagging negative (non-UTI)	145	410	555
Total	440	445	885

Table S2. Subgroup diagnostic performance of urinary tract infection (UTI) flagging with 2x2 contingency data

Subgroup	N	Positive culture prevalence	2×2 table (Flag+/flag- vs culture+/culture-)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
With urinary catheter	420	59.5% (250/420)	TP=155, FN=95, FP=20, TN=150	0.620 (0.558–0.679)	0.882 (0.821–0.927)	0.886 (0.830–0.927)	0.612 (0.545–0.676)
Without urinary catheter	465	40.9% (190/465)	TP=140, FN=50, FP=15, TN=260	0.737 (0.668–0.798)	0.945 (0.911–0.968)	0.903 (0.845–0.943)	0.839 (0.791–0.879)
Immunosuppressant: Yes	165	66.7% (110/165)	TP=65, FN=45, FP=10, TN=45	0.591 (0.495–0.682)	0.818 (0.689–0.909)	0.867 (0.772–0.930)	0.500 (0.395–0.605)
Immunosuppressant: No	720	45.8% (330/720)	TP=230, FN=100, FP=25, TN=365	0.697 (0.645–0.745)	0.936 (0.906–0.958)	0.902 (0.857–0.935)	0.785 (0.745–0.821)
Female sex	485	55.7% (270/485)	TP=225, FN=45, FP=20, TN=195	0.833 (0.783–0.875)	0.907 (0.861–0.941)	0.918 (0.875–0.949)	0.812 (0.756–0.860)
Male sex	400	42.5% (170/400)	TP=70, FN=100, FP=15, TN=215	0.412 (0.340–0.487)	0.935 (0.895–0.962)	0.824 (0.731–0.893)	0.683 (0.628–0.734)
Age: 18–<65	695	48.9% (340/695)	TP=210, FN=130, FP=30, TN=325	0.618 (0.565–0.668)	0.915 (0.879–0.943)	0.875 (0.826–0.914)	0.714 (0.670–0.755)
Age: ≥65	145	51.7% (75/145)	TP=65, FN=10, FP=0, TN=70	0.867 (0.769–0.931)	1.000 (0.949–1.000)*	1.000 (0.945–1.000)*	0.875 (0.786–0.934)

Note: The perfect specificity and PPV in the ≥65 age group should be interpreted with caution given the small sample size (n = 145) and the absence of false positives in this stratum. The confidence interval lower bound for specificity is 0.949, reflecting uncertainty around the point estimate of 1.000. Age-stratified analyses were restricted to adults; 45 patients aged <18 years (25 culture-positive, 20 culture-negative) were retained in the overall cohort (N = 885) and in all other analyses but were not assigned to an adult stratum, so the two age strata sum to N = 840.

Table S3. Gram agreement of BACT-info Gram (derived from Gram positivity or Gram-negative indicators) vs culture Gram

Culture Gram	Gram-negative	Gram-positive	Total
Gram negative	235	15	250
Gram positive	5	65	70
Total	240	80	320

Note: Denominator (n = 320) represents bacterial culture samples with unambiguous Gram classification (Gram-negative or Gram-positive) after excluding: mixed bacterial growth (n = 60), no growth (n = 445), and samples with indeterminate indicator outputs (both positive or both negative; n=60). Cohen's κ = 0.826 (bootstrap 95% CI 0.746–0.893).

Table S4. Gram agreement of BACT-info category label (flagging BACT-info mapped to Gram positive only) vs culture Gram

Culture Gram	Gram-negative	Gram-positive	Total
Gram-negative	115	90	205
Gram-positive	0	40	40
Total	115	130	245

Note: The denominator (n = 245) comprises 380 single-Gram culture-positive samples (Figure 1: 285 Gram-negative and 95 Gram-positive) after excluding 135 samples who's categorical BACT-info label was non-binary ('Mixed', 'Non-classified', or 'No flag'). These non-binary labels affected both Gram classes — 80 Gram-negative samples (285 reduced to 205) and 55 Gram-positive samples (95 reduced to 40) — rather than Gram-negative samples alone. Culture-negative (no-growth, n = 445) and mixed-growth (n=60) samples were excluded by definition. Cohen's κ = 0.294 (bootstrap 95% CI 0.201–0.397).

Table S5. Bootstrap validation of Youden-optimized cutoffs

Parameter	Training set cutoff	Bootstrap median cutoff	Bootstrap 95% CI	Optimism-corrected sensitivity	Optimism-corrected specificity
WBC (/μL)	76.9	78.2	52.3–101.5	0.728	0.768
BACT (/μL)	487	491	382–612	0.788	0.796

Note: The narrow confidence intervals and minimal optimism (≤ 0.007 difference) suggest that the derived cutoffs are reasonably stable within this dataset. However, external validation in an independent cohort remains necessary before these thresholds can be recommended for clinical use.