Improving management of infectious diseases at the emergency department: from risk prediction to treatment strategies

Kaal, A.G.

Chapter 4

Bacteremia prediction models to reduce unnecessary blood cultures: external validation in a large US emergency department

Anna Kaal, Laraine Washer, Anastasia Wasylyshyn, Ewout Steyerberg, Prashant Mahajan, Cees van Nieuwkoop

Scientific Reports, 2026 Apr; doi: 10.1038/s41598-026-47024-5

ABSTRACT

Bacteremia prediction models can avoid unnecessary blood cultures while missing few bacteremia patients in European cohorts. We aimed to assess the performance of these models in the United States. We externally validated five prediction models in patients ≥ 18 years old evaluated in an academic emergency department (ED) in the United States, between October 2013 and October 2023. Patients needed to have one or more blood cultures and a procalcitonin (PCT) test within 6 hours of obtaining blood cultures. The models used basic clinical information such as age and vital signs, with or without the addition of C-reactive protein (CRP), PCT or both. Missing data were imputed. The outcome of the models was bacteremia, defined as at least one true positive blood culture obtained within 48 hours of the ED visit. Performance of the models was assessed with the C-statistic for discrimination and calibration plots for calibration. Clinical usefulness was evaluated by the potential reduction in blood cultures, missed bacteremic episodes and decision curve analysis.

We included 12,864 patients (55% male, median age 66 years). 1164 out of 12,864 (9.0%) had bacteremia, while 843 (6.6%) patients had contaminated blood cultures. We observed substantial improvement in performance for models that included PCT, compared to those based on clinical information alone or with CRP. The best performing models included PCT with or without clinical information, with C-statistics of 0.80 (95% CI 0.79-0.82 and 0.79-0.81), adequate overall calibration (predicted 10.4%-13.0%, observed 9.0%). Both models were able to reduce ~30% of blood cultures while missing 5% of bacteremic episodes.

We conclude that previously proposed bacteremia prediction models demonstrated robust applicability in a large United States ED population. Models including procalcitonin have promise to safely avoid many blood cultures in the ED. Further validation and impact studies may support the implementation of these models.

INTRODUCTION

According to the Surviving Sepsis Campaign guidelines it is considered best practice to obtain blood cultures in patients with suspected sepsis before starting antimicrobial therapy.[1] Therefore, it is customary at emergency departments (EDs) to maintain a low threshold to perform blood cultures, as the signs and symptoms of sepsis are nonspecific and may mimic a variety of diseases.[1] Only 4-11% of all blood cultures obtained at the ED yield a true positive pathogen, while up to 50% of all positive cultures may be contaminated.[2-7] Contaminated blood cultures are associated with increased length of stay and more exposure to antibiotics. Additionally, blood cultures are costly (around 90-120 USD per set).[5-7]

Previously, we developed, validated, and updated multiple bacteremia prediction models based on two European cohorts, with the goal to identify ED patients at low risk for bacteremia and hence, safely reduce unnecessary blood cultures.[8] The optimal model consisted of routinely available clinical information combined with the biomarker procalcitonin (PCT). This model calculates the individual pre-test probability of bacteremia of adult ED patients with clinical suspicion of infection. Guiding cultures with this model at a 5% risk threshold might avoid a quarter of all blood cultures while only missing 1% of all true positive cultures.[8]

Although these results were very promising, we observed geographical variation in performance, and risks in the lower risk-ranges were slightly overestimated. In this study, we aimed to assess the performance of these models in a large emergency department in the United States through a retrospective chart review.

METHODS

Data and participants

We extracted data from the electronic health records of all patients aged 18 years or older who had a blood culture taken during their ED visit of the University of Michigan Hospital, Ann Arbor, United States, from October 2013 until October 2023. In case of multiple ED visits only the first was included. We excluded patients who left the ED before they received treatment, those whose blood culture was taken before the ED visit or if the first culture was taken more than 24 hours after ED arrival. Patients suspected of endocarditis or those with an intravascular catheter were excluded since we would not recommend using the models in these patient categories. The TRIPOD guideline was followed for reporting on the validation of multivariable prediction models.[9]

Outcome

The outcome of the prediction models was bacteremia, defined as at least one true positive blood culture taken within 48 hours of the start of the ED visit. In a multidisciplinary team consisting of multiple Infectious Disease specialists and hospital epidemiologists, we developed a flowchart to categorize blood cultures with growth of any organism(s) in the following categories: positive, inconclusive, negative and contaminated. We performed a chart review for all inconclusive cultures to evaluate whether the treating clinicians had considered the culture to be truly positive. The flowchart and all retrieved micro-organisms are listed in the Supplement. In summary, Coagulase-negative staphylococci (CNS), Bacillus spp., Corynebacterium spp., Cutibacterium spp., and Lactobacillus spp., were considered contaminants unless there were multiple blood cultures with the same bacterium in a patient with a non-vascular prosthetic device (e.g. prosthetic joint). Other reported outcomes are 30-day mortality, admission, ICU admission and start of antibiotic treatment.

Laboratory methods

Blood cultures were collected in an aerobic and anaerobic BD BACTEC bottle and incubated for a minimum of five days using the BD Bactec FX Blood Culture System. Since September 2022, a FilmArray Blood Culture Identification PCR Panel (BCID2) was run on the first positive set of bottles. PCT was first tested using the VIDAS BRAHMS PCT assay (Enzyme Linked Fluorescence Assay), range 0.05 to 200 ng/mL, through May 2018. After that moment, the Roche PCT Assay (Electrochemiluminescence Immunoassay) was used, range 0.02 to 100 ng/mL. CRP was tested with a turbidimetric immunoassay.

Predictors

The predictors used in the models were age, systolic blood pressure (mmHg), heart rate (beats/min), respiratory rate (breaths/min), temperature (°Celsius), Glasgow Coma Scale, CRP (mg/L) and PCT (ng/mL). Biomarker values closest to blood culture collection time were used, with a maximum six-hour difference. For vital signs, the average value during the ED visit up to the moment of blood culture collection was used.

Definitions

We defined suspected endocarditis as the registration of an ICD-10 code during the ED visit related to endocarditis. We created lists of ICD-10 codes for comorbidities, categorized into renal disease, diabetes mellitus, haematological disease,

neurological disease, cardiovascular disease and malignancy (See Supplement). Finally, we selected ICD-10 codes related to an indwelling vascular catheter and to prosthetic material (intravascular and joint). Recent use of antibiotics was defined as use in the seven days before ED visit, and use of immunosuppression as a minimum of two dispenses of immunosuppressive medication in the year before ED visit, of which the last dispense was within the previous six months (See Supplement for lists of antibiotics and immunosuppressants).

Statistical analysis

Continuous variables are presented as mean and standard deviation (SD) or median and interquartile range (IQR). Categorical data are summarized as counts and percentages. Comparisons were made with the $\chi 2$ test, the unpaired t-test or Mann-Whitney U test, as appropriate. We collected data available from the last 10 years, aiming for as many events as possible, with a minimum of 200 events for reliable validation.[10] We performed single imputation of missing data for all predictors except for PCT using multiple imputations by chained equations. From this dataset, we selected all patients with a known PCT value. We validated the European models on the US data, by calculating predictions using the 'predict' function in the R package 'stats' (eTable 1). Predictive performance was assessed by discrimination and calibration. Discriminative ability was assessed by the C-statistic, which is equivalent to the area under the ROC curve.[11] To assess whether differences in C-statistic were due to a difference in case-mix or due to misfit of the original model we calculated the model-based C-statistic, which provides the C-statistic for the case-mix in the validation data if the model were fully valid.[12] Calibration was evaluated graphically by calibration plots, with intercept and slope as summary statistics (ideal values 0 and 1 respectively).

Additionally, we explored updating strategies by splitting the database 70%/30%. The 70% part was used to develop multiple updated versions: the original model with updated intercept, intercept and slope and a refitted model with equal weights per dataset (Netherlands, Spain and USA). In this refitted model, an interaction term for PCT and database was added if it improved the model.[13] All models were then evaluated in the 30% dataset. Other analyses were subgroup analyses for patients using immunosuppressants, for gender, and a complete case analysis. P-values < 0.05 were considered statistically significant. Analyses were performed using RStudio version 2024.09.0.

Potential clinical usefulness

To express clinical usefulness, we calculated the number of avoided blood cultures and number of missed positive cultures at multiple risk thresholds. We report outcomes at a risk threshold of 2.5% and 5% for comparability with the original study.[8] This threshold reflects the rate of missed bacteremia that is considered acceptable. We consider a plausible range of risk thresholds from 1% to 10%, arguing that any clinician would order blood cultures if the risk would exceed 10%. For PCT, we also reported outcomes at predefined cut-offs of <0.05 ng/mL, <0.10 ng/mL, <0.25 ng/mL and <0.50 ng/mL (Supplement eTable 2 and eTable 3). Clinical usefulness was summarized as Net Benefit (NB), expressed as net interventions avoided per 100 patients, also known as $NB_{Opt\text{-}out}$.[14] This measure quantifies the potential impact on patient outcomes by decision making based on a prediction model. Reference strategies are testing all or none of the patients.

Development versus validation

We assured that inclusion criteria and definitions of predictors and outcome were comparable to the development data. We used histograms to inspect differences in the distribution of continuous predictors with the European data (Supplemental Material). In the Dutch and Spanish setting, blood cultures were ordered based on the judgement of the attending clinician. In Michigan, four screening questions were used to guide whether cultures would be obtained during triage (See Supplement) which could lead to a broader selection of patients. In the Spanish data, patients without a diagnosis of infection within 30 days were excluded. Finally, PCT was tested for all patients in the development cohorts, but only ordered based on the judgement of the attending physician in Michigan.

Ethics

This study was approved by the Institutional Review Board of the University of Michigan (HUM00236866). The requirement for informed consent was waived by the University of Michigan ethics committee due to the retrospective nature of the study. All methods were carried out in accordance with relevant guidelines and regulations.

RESULTS

We identified 46,931 unique ED patients with blood cultures. PCT values were available for 12,864 patients (27%) fulfilling the inclusion criteria. The median age was 65 years, and 7,089/12,864 (55%) were male (Table 1). Bacteremia was diagnosed in 1164 out of 12,864 patients (9.0%), while 843 cultures were

contaminated (6.6%) (Figure 1). The analysis contrasted the 1164 patients with true bacteremia to all others (Negative or contaminated cultures).

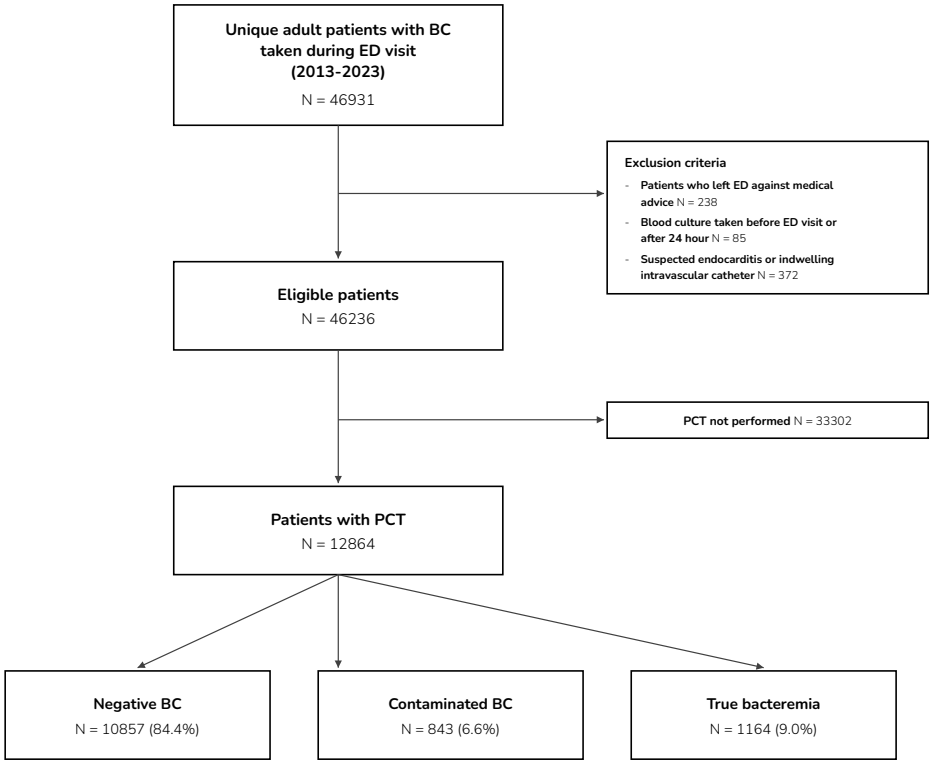

Figure 1. Flowchart of patient inclusions.

The patients in Michigan were comparable to the European patients, and especially similar to the Dutch cohort. (eTable 4) They were less severely ill as reflected in a lower temperature, lower heart rate and lower breathing rate. Patients with a known PCT compared to those without PCT were older, had more comorbidity and suffered from more severe disease as reflected in their vital functions and clinical outcomes. (eTable 5)

Table 1. Baseline characteristics

Characteristics	No bacteremia	Bacteremia	Total	p
	n = 11700	n = 1164	n = 12864	
Demographics and medical history				
Age, years, (median [IQR])	64.6 [51.5, 75.1]	65.5 [54.4, 75.2]	64.8 [51.7, 75.1]	0.030
Sex (Male)	6374 (54.5)	715 (61.4)	7089 (55.1)	<0.001
Sex (Female)	5326 (45.5)	449 (38.6)	5775 (44.9)	<0.001
Nursing home resident	1208 (10.3)	168 (14.4)	1376 (10.7)	<0.001
Cardiovascular disease	5266 (45.0)	588 (50.5)	5854 (45.5)	<0.001
Diabetes mellitus	3610 (30.9)	409 (35.1)	4019 (31.2)	0.003
Hematological disease	1728 (14.8)	174 (14.9)	1902 (14.8)	0.904
Malignancy	4733 (40.5)	512 (44.0)	5245 (40.8)	0.021
Neurological disease	1920 (16.4)	195 (16.8)	2115 (16.4)	0.796
Pulmonary disease	4929 (42.1)	459 (39.4)	5388 (41.9)	0.081
Renal disease	2799 (23.9)	327 (28.1)	3126 (24.3)	0.002
No comorbidities	1500 (12.8)	122 (10.5)	1622 (12.6)	0.025
Use of immunosupressants	2213 (18.9)	243 (20.9)	2456 (19.1)	0.113
Recent use of antibiotics	1422 (12.2)	130 (11.2)	1552 (12.1)	0.349
Vital functions in ED				
Breathing frequency, breaths/min, (median [IQR])	20.0 [18.0, 23.0]	20.0 [18.0, 23.2]	20.0 [18.0, 23.0]	0.573
Glasgow coma scale, (median [IQR])	15.0 [14.5, 15.0]	15.0 [14.0, 15.0]	15.0 [14.5, 15.0]	<0.001
Heart frequency, beats/min, (mean (SD))	97.8 (21.5)	104.7 (22.9)	98.4 (21.7)	<0.001
Saturation, %, (median [IQR])	96.0 [93.5, 97.8]	96.0 [93.9, 97.8]	96.0 [93.5, 97.8]	0.171
Systolic blood pressure, mmHg, (mean (SD))	127.3 (26.6)	118.1 (27.3)	126.5 (26.8)	<0.001
Temperature, °C, (median [IQR])	36.9 [36.6, 37.5]	37.3 [36.8, 38.1]	36.9 [36.7, 37.6]	<0.001
Laboratory values in ED				
Creatinine, µmol/L, (median [IQR])	86.7 [65.4, 128.2]	107.0 [75.2, 170.7]	88.4 [66.3, 131.8]	<0.001
Urea, mmol/L, (median [IQR])	7.1 [5.0, 11.4]	9.6 [6.4, 15.4]	7.5 [5.0, 11.8]	<0.001
Procalcitonin, ng/mL, (median [IQR])	0.2 [0.1, 0.7]	2.5 [0.5, 13.4]	0.2 [0.1, 0.9]	<0.001
C-reactive protein, mg/L, (median [IQR])	68.0 [25.0, 145.0]	145.0 [63.5, 223.0]	74.0 [27.0, 155.0]	<0.001
White blood cell count, x10^9/L, (median [IQR])	10.3 [6.9, 14.8]	11.8 [7.1, 17.4]	10.4 [6.9, 15.0]	<0.001
Neutrophil count, x10^9/L, (median [IQR])	7.8 [4.8, 11.9]	10.3 [6.1, 15.1]	8.0 [4.9, 12.2]	<0.001
White blood cell count < 0.5 x 10^9/L	94 (0.8)	46 (4.0)	140 (1.1)	<0.001

Characteristics	No bacteremia	Bacteremia	Total	p
	n = 11700	n = 1164	n = 12864	
Neutrophil count <0.5 x 10^9/L	184 (1.6)	32 (2.9)	216 (1.7)	0.003
Platelet count, x10^9/L, (median [IQR])	220.0 [158.0, 298.0]	183.0 [118.0, 257.0]	217.0 [155.0, 295.0]	<0.001
Patient outcomes				
Admission	11078 (94.7)	1136 (97.6)	12214 (94.9)	<0.001
ICU admission	2602 (22.2)	433 (37.2)	3035 (23.6)	<0.001
Start of new antibiotics	9703 (82.9)	1097 (94.2)	10800 (84.0)	<0.001
Duration of antibiotic treatment, days, (median [IQR])	7.3 [2.5, 13.6]	13.8 [10.9, 13.9]	7.9 [2.8, 13.8]	<0.001
Contaminated BCs	843 (7.2)	0 (0.0)	843 (6.6)	<0.001
30 day mortality	1410 (12.1)	215 (18.5)	1625 (12.6)	<0.001

Table 1. Baseline characteristics for all included patients. Data are presented as number (percentage) of patients unless otherwise indicated. Recent use of antibiotics: in prior 7 days. IQR = Interquartile Range, ED = emergency department, SD = standard deviation, ICU = Intensive Care Unit, BCs = Blood cultures. Recent use of antibiotics: in the previous seven days. Missing values: Systolic blood pressure: 215 (1.7%), Heart frequency: 144 (1.1%), Breathing frequency: 303 (2.4%), Temperature 843 (6.6%), Glasgow coma scale: 3195 (24.8%), Saturation: 275 (2.1%), Creatinine: 52 (0.4%), Urea: 52 (0.4%), C-reactive protein: 10547 (82%), White blood cell count: 62 (0.5%), Neutrophil count: 229 (1.8%), Platelet count: 148 (1.2%), Duration of antibiotic treatment: 1207 (9.4%).

Model performance

While all models showed slightly lower performance compared to the European setting upon external validation, the discriminative ability of models with PCT was confirmed with AUC values above 0.8 for models including PCT (Table 2). The difference was partly attributable to a difference in case-mix, as indicated by the model-based concordance around 0.84 rather than around 0.87 in the European setting. Calibration plots showed a minor overestimation of risk (Overall predicted risk (10.1-13.0%) vs observed (9.0%)). (Figure 2) All PCT-containing models performed better than those without PCT. At a risk threshold of 2.5%, the basic model with PCT avoided blood cultures in 3,608 out of 12,864 patients (28%) while missing only 52 out of 1164 (4.5%) bacteremic episodes. Adding CRP did not lead to better performance. The 'PCT only' model led to comparable clinical outcomes, potentially avoiding blood cultures in 30% of patients while missing 5.2% of bacteremic episodes at a 4% risk threshold. This model showed slightly poorer calibration. (Table 2) Net benefit analysis showed most benefit for the PCT-containing models (Figure 3). Updating the original models did not lead to improved performance (eTable 6, eFigures 2 and 3). We observed stable results in the complete case analysis and in a second imputation set (eTable 7 and 8). Subanalyses for patients using immunosuppressants or according to gender showed comparable results (eTable 9).

Clinical outcomes for low-risk patients

We compared clinical outcomes for patients with an estimated risk of bacteremia lower than 2.5% by the basic model with PCT, to those with a higher risk. All outcomes were better for low-risk patients, only contamination rates were comparable (6.5% vs. 6.6%, $p = 0.93$). Low-risk patients had a hospital admission rate of 90% vs. 97% ($p < 0.001$). ICU admission rates were 14% vs. 27% ($p < 0.001$). New antibiotics were started in 75% vs. 88% ($p < 0.001$), bacteremia was present in 1.4% vs. 12.0% ($p<0.001$) and 30-day mortality was 5% vs. 16% ($p<0.001$).

Table 2. Performance of the models in the Michigan data

	Discrimination			Outcomes at a 2.5% risk threshold		Outcomes at a 5% risk threshold		Calibration	
Model	**Development data (95% CI)**	**Michigan (95% CI)**	**Model based C-statistic**	**Avoided BCs (Among 12864)**	**Missed bacteremic episodes (Among 1164)**	**Avoided BCs (Among 12864)**	**Missed bacteremic episodes (Among 1164)**	**Intercept (95% CI)**	**Slope (95% CI)**
Basic with PCT	0.87 (0.86-0.88)	0.80 (0.79-0.82)	0.84	3608 (28.1%)	52 (4.5%)	6893 (53.6%)	166 (14.3%)	-0.21 (-0.27 to -0.14)	0.79 (0.75 to 0.84)
Basic with CRP and PCT	0.87 (0.86-0.88)	0.80 (0.78-0.81)	0.84	3557 (27.7%)	55 (4.7%)	6850 (53.2%)	177 (15.2%)	-0.23 (-0.30 to -0.16)	0.78 (0.73 to 0.82)
PCT only	0.87 (0.85-0.88)	0.80 (0.79-0.81)	0.82	1347 (10.5%)	16 (1.4%)	4906 (38.1%)	93 (8.0%)	-0.52 (-0.59 to -0.45)	0.84 (0.79 to 0.89)
Basic with CRP	0.71 (0.69-0.73)	0.69 (0.67-0.70)	0.67	360 (2.8%)	8 (0.7%)	2343 (18.2%)	64 (5.5%)	-0.13 (-0.19 to -0.07)	1.07 (0.97 to 1.18)
Basic without biomarkers	0.70 (0.68-0.71)	0.64 (0.62-0.65)	0.64	47 (0.4%)	0 (0%)	1431 (11.1%)	58 (5%)	-0.13 (-0.19 to -0.07)	0.94 (0.83 to 1.06)

Table 2. Performance of the models in the Michigan data. The discrimination is reported as the C-statistic and compared to the model-based C-statistic. The difference between the C-statistic in the development data and the model-based C-statistic reflects the difference attributable to a difference in case-mix. Clinical outcomes are given at a 2.5% and 5% risk threshold with the corresponding number of patients in whom blood cultures could have been avoided, and the number of missed bacteremic episodes. Predictions were calculated with the database variable at 0 based on visual inspection of the predictor distributions (eTable 1). CI = confidence interval, BCs = blood cultures, CRP = C-reactive protein, PCT = Procalcitonin.

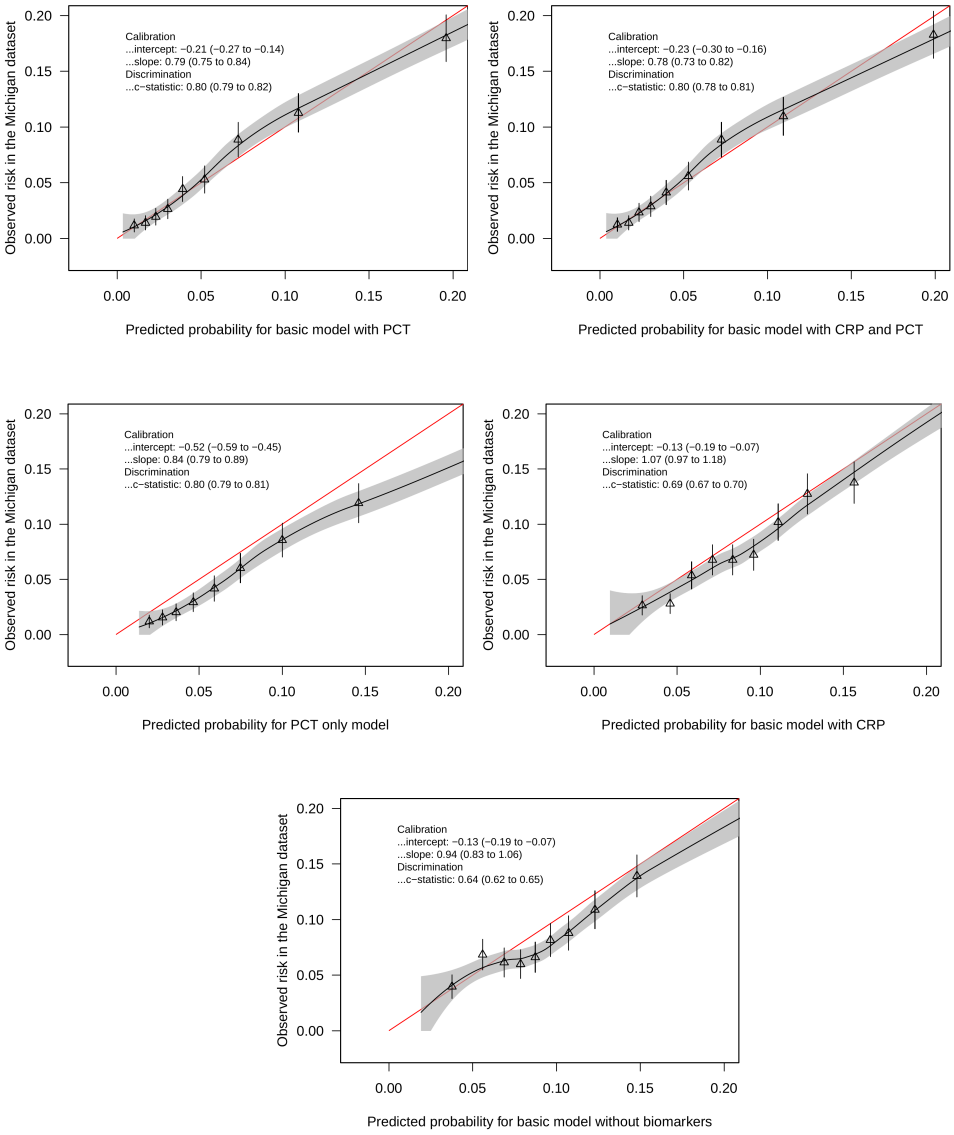

Figure 2: Calibration plots of all models in the Michigan dataset. The X-axis represents the predicted risks, the Y-axis represents the observed risks. CRP = C-reactive protein, PCT = procalcitonin.

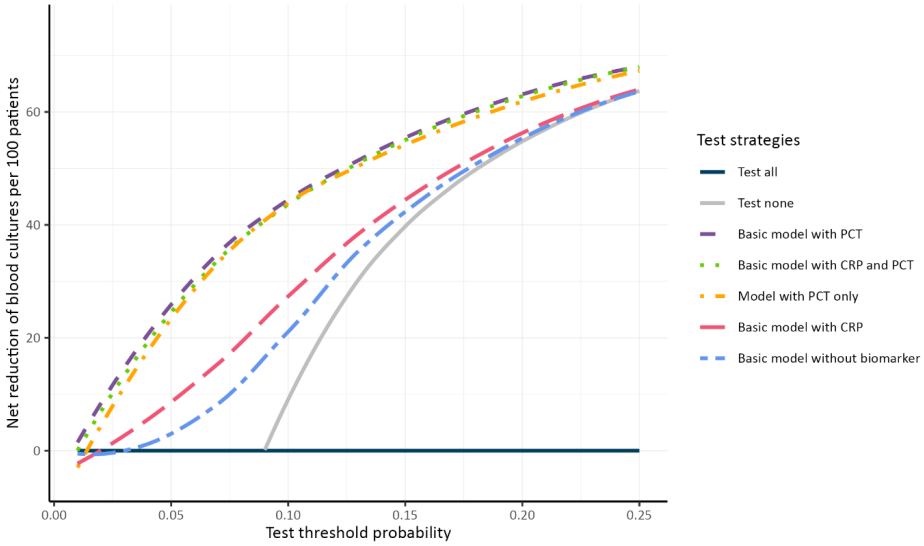

Figure 3: Net benefit curves for all original models. This shows the expected net reduction of blood cultures per 100 patients for all models applied to the Michigan data. cohort relative to testing everyone ("Test all") and no one ("Test none"). The x-axis displays the possible risk thresholds for bacteremia on which to decide to withhold blood cultures. For instance: one physician might accept a maximum bacteremia risk of 5%, taking cultures in most patients. The net reduction of blood cultures, or Net $Benefit_{Opt\text{-}out}$, can be calculated by TNRR*(1-p)-p*(1-R/R)*FNRR, in which TNR is the true negative rate, R is the risk threshold, p is the disease prevalence and FNR is the false negative rate.[8,12]

DISCUSSION

We found adequate applicability for diagnostic prediction models for bacteremia in the US. Different model variants were developed and previously validated in Dutch and Spanish patients, and lessons were drawn to create broadly applicable models.[8] The current validation in US patients confirms the promising performance for models that include PCT to rule out bacteremia. The PCT models were able to avoid over a quarter of blood cultures while missing around 5% of bacteremic episodes, a risk which many consider sufficiently low to withhold blood cultures.[15,16] It is important to note that the models can also be used with a more cautious approach, according to what the clinician would find acceptable. This would lead to missing fewer bacteremic episodes, and avoiding fewer blood cultures.

To guide blood culture stewardship, indication-based algorithms are considered to achieve the highest impact, with an estimated reduction of 20-50% by avoiding blood cultures in patients with a low risk of bacteremia (< 5-10%).[15-18] Especially within the ED setting, implementation of such algorithms is challenging as diagnoses or source of infection are often still uncertain or even faulty.[17,19] Some

prediction models aim to predict the risk of bacteremia, something clinicians tend to overestimate.[20,21] However, many of these tools are too labor-intensive for a busy ED or complex to incorporate into the electronic health record. Moreover, most models lack performance or robust validation, so clinicians are hesitant to use them.[22] As overuse of blood cultures is associated with patient harm, robust models that are readily available for application at EDs are urgently needed.

The five models that we developed previously performed slightly worse in the US than in the European context. A part of this difference is explained by case-mix variation. The other part may be due to true differences not captured by these models. For instance, there are differences in clinical settings. In the Netherlands, general practitioners play a major role in treating patients with infectious diseases, and people who are sent to the ED are generally sicker and more often pretreated with antibiotics compared to Spain and the United States. The University of Michigan Hospital is a large center providing academic care, with 1.5 to 2.5 times more patients using immunosuppressants, and few patients with no medical history. However, multiple strategies for local updating did not improve the performance of the models.

The importance of procalcitonin for bacteremia prediction has been shown before in meta-analyses, with a C-statistic of the summary ROC curve of 0.79 for ED patients.[23] Another recent meta-analysis found C-statistics ranging from 0.68 to 0.98.[24] For high-quality studies using a PCT cut-off point >0.5 ng/mL, the C-statistic of the summary ROC curve was 0.89. Another large study in the US demonstrated limited sensitivity for 'blood stream infections'. However, using a lower cut-off for PCT, a large number of blood cultures could have been avoided.[25,26] In our development study, we observed a very strong relation between PCT and bacteremia in the Spanish dataset (C-statistic of 0.9) which motivated an update of the models based on Dutch data only.[8] In the current validation study, this strong relation was also observed although more in line with the original Dutch development data (C-statistic 0.79).[8] Surprisingly, basic clinical characteristics did not provide relevant increases in performance, as opposed to the results of the Dutch dataset. Why the predictive performance of PCT differs in different settings is unclear. Although we tried to align all datasets, this could be due to a different patient selection, slight differences in definitions of bacteremia, or timing of the PCT test.

A strength of this study is the large size of the validation dataset with high quality data. We included a broad range of patients, giving a realistic perspective on the application of the models in a US emergency department. A limitation is that we excluded a large part (73%) of eligible patients based on lack of PCT results, without a clear guideline for PCT testing. This has led to a selection of patients

that were more severely ill, as reflected in eTable 5. This selection could have been a confounding factor that could have influenced the results. However, the final selection of patients was more comparable to the development dataset, and the predictive performance of procalcitonin was comparable with the European cohorts in which PCT was tested in all patients. Another limitation is that we used patient chart data, which inherently limits the available information. However, since all predictors used in the models were objective values, we believe this will not have impacted the conclusions of our study. Finally, since the number of positive blood cultures was too high for manual chart review we created a flowchart to categorize blood cultures. In this flowchart, we erred on the side of caution which could have led to an overestimation of the rate of positive cultures. However, if so, this would result in even better clinical outcomes in clinical practice.

Utility in current care

Based on adequate performance and the limited effort required to use a model, we recommend moving forward with either the basic model with PCT or the PCT only model for a prospective validation study. If the setting and case-mix are different from the European and US cohorts, we would advise local validation with model updating if needed. The performance of the models without PCT will not be acceptable for most clinicians. In the Dutch dataset, we observed some improvement in discrimination when adding basic clinical predictors. Ideally, one would evaluate both models using a shadow study, meaning that both models would run in the background without changing diagnostic decisions, and evaluating the performance before moving forward with implementation.[13] This could also be used to decide whether there is enough added performance to justify the extra effort of implementing the basic model with PCT rather than the PCT only model.

There will be multiple challenges for implementing these models.[13] We suggest that when the treating clinician considers whether to obtain blood cultures, one would adjust the process as follows. First, if deemed safe to withhold antibiotics for one to three hours one could draw lab testing including PCT and use the model to estimate the risk of bacteremia, using all information available up until the moment that a blood culture would be ordered.[27] We propose withholding blood cultures when the estimated risk is <5%, and obtaining them when the risk exceeds 5%. An alternative is to withhold blood cultures if the estimated risk is below the local contamination rate. The model can be adjusted for the clinician's level of risk tolerance, with the consequence that more (or fewer) blood cultures will be withheld. If antibiotics are already needed before model results are available, one could obtain the cultures and decide whether to send cultures for testing based on the model results. In this way,

it is possible to avoid costs from microbiological analysis and negative consequences of contaminated cultures. However, it must be noted that this will only be possible in hospitals where PCT results are known within the first hours. We would argue that if not all predictors are available, one should use the PCT only model, since the added value of clinical predictors was relatively limited compared to the PCT only model. Of note, the models should not be applied to patients with a vascular catheter or suspected endocarditis.

Finally, the implementation of these models in the US could be hindered by issues related to the severe sepsis and septic shock management bundle (SEP-1) compliance.[28] SEP-1 is a quality metric that evaluates the timely collection of blood cultures before the administration of antibiotics in patients with suspected severe sepsis and septic shock. However, it only assesses whether cultures were performed, without considering the clinical usefulness of those cultures. Therefore, while our proposed models could lead to more appropriate culture use, the reduction of cultures could negatively impact SEP-1 compliance. In turn, this could lead to decreased financial reimbursement for hospitals, as SEP-1 compliance is tied to value-based purchasing and Medicare reimbursement rates. Addressing these challenges at the policy level is crucial to align quality improvement efforts with financial incentives, ensuring that hospitals can implement diagnostic stewardship models without jeopardizing their reimbursement.

Whether the implementation of these models is worthwhile will differ per setting. For EDs that have already implemented PCT testing in their routine clinical practice, the costs of implementing the models will be low. In settings where PCT testing is not implemented, the costs of the extra PCT tests will have to be weighed against the potential savings of avoiding blood cultures but potentially more missed bacteremia. Costs of PCT, blood cultures and the associated costs with false-negative cultures differ around the world, necessitating local evaluation.

Further research

Important next steps would be to evaluate the potential impact of model implementation, for instance with a shadow study and by evaluating model-user interaction. Questions to be addressed include whether clinicians accept the advice from the model or will the blood cultures still be ordered? What is the outcome of cultures that are taken against the recommendation of the model? And, importantly, does the model impact clinical outcomes?

CONCLUSIONS

Five previously proposed bacteremia prediction models perform well for the US. Especially models including procalcitonin show promising performance to avoid blood cultures in the emergency department while missing few bacteremia cases. Further impact studies may support the implementation of these models in the emergency department.

Data Availability Statement

De-identified, individual patient data supporting the findings of this study are available from the corresponding author upon reasonable request, available from publication without end date. R MarkDown files showing the analysis for one of the models are available in the supplementary material.

Use of large language models

ChatGPT was used to improve readability and language.

Potential conflicts of interest

All authors declare no conflicts of interest.

Financial support

This project was made possible because of personal grants appointed to AK, from the Fullbright Commission the Netherlands, the Leids Universitair Fonds, the Catharine van Tussenbroek Fonds, the Cultuurfonds, the Hendrik Mullerfonds and the department of Internal Medicine and Rheumatology, Haga Teaching Hospital, the Netherlands.

Author contributions

Concept and design: AK, ES, CvN. Acquisition of data: AK, PM. Analysis and interpretation of data: AK, LW, AW, ES, PM, CvN. Drafting of the manuscript: AK, critical revision of the manuscript: AK, LW, AW, ES, PM, CvN. Statistical expertise: AK, ES. Acquisition of funding, AK, CvN

SUPPLEMENTARY MATERIAL

REFERENCES

1. Evans L. et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med*. Nov 2021;47(**11**):1181-1247. doi:10.1007/s00134-021-06506-y
2. Long B, Koyfman A. Best Clinical Practice: Blood Culture Utility in the Emergency Department. *J Emerg Med*. 2016 Nov;51(**5**):529-539. doi: 10.1016/j.jemermed.2016.07.003. Epub 2016 Sep 14. PMID: 27639424.
3. Coburn B, Morris AM, Tomlinson G, Detsky AS. Does this adult patient with suspected bacteremia require blood cultures? *JAMA*. 2012 Aug 1;308(**5**):502¬-11. doi: 10.1001/jama.2012.8262. Erratum in: *JAMA*. 2013 Jan 23;309(**4**):343. PMID: 22851117.
4. Nannan Panday RS. et al. Evaluation of blood culture epidemiology and efficiency in a large European teaching hospital. *PLoS One*. 2019 Mar 21;14(**3**):e0214052. doi: 10.1371/journal.pone.0214052. PMID: 30897186; PMCID: PMC6428292.
5. Dargère S, Cormier H, Verdon R. Contaminants in blood cultures: importance, implications, interpretation and prevention. *Clin Microbiol Infect*. 2018 Sep;24(**9**):964-969. doi: 10.1016/j.cmi.2018.03.030. Epub 2018 Apr 3. PMID: 29621616.
6. Dempsey C, Skoglund E, Muldrew KL, Garey KW. Economic health care costs of blood culture contamination: A systematic review. *Am J Infect Control*. 2019 Aug;47(**8**):963-967. doi: 10.1016/j.ajic.2018.12.020. Epub 2019 Feb 20. PMID: 30795840.
7. Bates DW, Goldman L, Lee TH. Contaminant blood cultures and resource utilization. The true consequences of false-positive results. *JAMA*. 1991 Jan 16;265(**3**):365-9. PMID: 1984535.
8. Kaal AG. et al. Procalcitonin for safe reduction of unnecessary blood cultures in the emergency department: Development and validation of a prediction model. *J Infect*. Oct 2024;89(**4**):106251. doi:10.1016/j.jinf.2024.106251
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. Ann Intern Med. 2015 Jan 6;162(1):55-63. doi: 10.7326/M14-0697. Erratum in: Ann Intern Med. 2015 Apr 21;162(8):600. doi: 10.7326/L15-0078-4. PMID: 25560714.
10. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol*. 2005 May;58(**5**):475-83. doi: 10.1016/j.jclinepi.2004.06.017. PMID: 15845334.
11. Steyerberg EW. et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010 Jan;21(**1**):128-38. doi: 10.1097/EDE.0b013e3181c30fb2. PMID: 20010215; PMCID: PMC3575184.
12. Steyerberg EW. (2009) Clinical prediction models: a practical approach to development, validation, and updating. Springer, New York
13. Binuya MAE, Engelhardt EG, Schats W, Schmidt MK, Steyerberg EW. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. *BMC Med Res Methodol*. 2022 Dec 12;22(**1**):316. doi: 10.1186/s12874-022-01801-8. PMID: 36510134; PMCID: PMC9742671.
14. Kerr KF, Brown MD, Marsh TL, Janes H. Assessing the Clinical Impact of Risk Models for Opting Out of Treatment. *Med Decis Making*. 2019 Feb;39(**2**):86-90. doi: 10.1177/0272989X18819479. Epub 2019 Jan 16. PMID: 30649998; PMCID: PMC6374190.

15. Ryder JH, Van Schooneveld TC, Diekema DJ, Fabre V. Every Crisis Is an Opportunity: Advancing Blood Culture Stewardship During a Blood Culture Bottle Shortage. *Open Forum Infect Dis*. Sep 2024;11(**9**):ofae479. doi:10.1093/ofid/ofae479

16. Suleyman G. et al. Blood culture bottle inventory management and clinical conservation during supply shortages. American Society for Microbiology; 2024. https://asm.org/getmedia/8daa65e4-c141-4a0b-9d60-0ae2337505fc/Blood-Culture-shortage-FAQ_FINAL.pdf?ext=.pdf Accessed May 16th, 2025.

17. Fabre V, Sharara SL, Salinas AB, Carroll KC, Desai S, Cosgrove SE. Does This Patient Need Blood Cultures? A Scoping Review of Indications for Blood Cultures in Adult Nonneutropenic Inpatients. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. Aug 22 2020;71(**5**):1339-1347. doi:10.1093/cid/ciaa039

18. Schoffelen T. et al. European society of clinical microbiology and infectious diseases guidelines for antimicrobial stewardship in emergency departments (endorsed by European association of hospital pharmacists). *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*. Jul 17 2024;doi:10.1016/j.cmi.2024.05.014

19. Dregmans E. et al. Analysis of Variation Between Diagnosis at Admission vs Discharge and Clinical Outcomes Among Adults With Possible Bacteremia. *JAMA Netw Open*. 2022 Jun 1;5(**6**):e2218172. doi: 10.1001/jamanetworkopen.2022.18172. PMID: 35737389; PMCID: PMC9226997.

20. Morgan DJ. et al. Accuracy of Practitioner Estimates of Probability of Diagnosis Before and After Testing. *JAMA Intern Med*. Jun 1 2021;181(**6**):747-755. doi:10.1001/jamainternmed.2021.0269

21. Schinkel M. et al. Diagnostic stewardship for blood cultures in the emergency department: A multicenter validation and prospective evaluation of a machine learning prediction tool. *EBioMedicine*. 2022 Aug;82:104176. doi: 10.1016/j.ebiom.2022.104176. Epub 2022 Jul 16. PMID: 35853298; PMCID: PMC9294655.

22. Eliakim-Raz N, Bates DW, Leibovici L. Predicting bacteraemia in validated models--a systematic review. *Clin Microbiol Infect*. 2015 Apr;21(**4**):295-301. doi: 10.1016/j.cmi.2015.01.023. Epub 2015 Feb 10. PMID: 25677625.

23. Hoeboer SH, van der Geest PJ, Nieboer D, Groeneveld AB. The diagnostic accuracy of procalcitonin for bacteraemia: a systematic review and meta-analysis. *Clin Microbiol Infect*. 2015 May;21(**5**):474-81. doi: 10.1016/j.cmi.2014.12.026. Epub 2015 Jan 14. PMID: 25726038.

24. Julián-Jiménez A, Eduardo García D, Merinos-Sánchez G, García de Guadiana-Romualdo L, González Del Castillo J. Precisión diagnóstica de la procalcitonina para la bacteriemia en el servicio de urgencias: una revisión sistemática [Diagnostic accuracy of procalcitonin for bacteremia in the emergency department: a systematic review]. *Rev Esp Quimioter*. 2024 Feb;37(**1**):29-42. Spanish. doi: 10.37201/req/099.2023. Epub 2023 Dec 7. PMID: 38058128; PMCID: PMC10874657.

25. Lawandi A et al. Reliability of Admission Procalcitonin Testing for Capturing Bacteremia Across the Sepsis Spectrum: Real-World Utilization and Performance Characteristics, 65 U.S. Hospitals, 2008-2017. Crit Care Med. 2023 Nov 1;51(**11**):1527-1537. doi: 10.1097/CCM.0000000000005968. Epub 2023 Jul 3. PMID: 37395622.

26. Schuetz P. The Value of a Biomarker Should be Judged on What it Adds to the Clinical Assessment: Not the Area Under the Curve. Crit Care Med. 2024 Jan 1;52(**1**):e30-e31. doi: 10.1097/CCM.0000000000006035. Epub 2023 Dec 14. PMID: 38095530.

27. Venkatesh B. et al. Impact of 1-hour and 3-hour sepsis time bundles on patient outcomes and antimicrobial use: A before and after cohort study. *Lancet Reg Health West Pac.* 2021 Nov 2;18:100305. doi: 10.1016/j.lanwpc.2021.100305. PMID: 35024649; PMCID: PMC8654968.

28. Quality Measures Fact Sheet. Severe Sepsis and Septic Shock: Management Bundle Measure (NQF #0500). CMS, August 2020. https://www.cms.gov/priorities/innovation/media/document/bpci-advanced-alt-fs-my4-sepsis. Accessed on December 1st, 2025.