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

Kaal, A.G.

Chapter 3

Procalcitonin for safe reduction of unnecessary blood cultures in the emergency department: Development and validation of a prediction model

Anna Kaal, Soufian Meziyerh, Nathalie van Burgel, Martijn Dane, Nikki Kolfschoten, Prashant Mahajan, Agustin Julián-Jiménez, Ewout Steyerberg, Cees van Nieuwkoop

Journal of Infection, 2024 Oct;89(4):106251

ABSTRACT

Objectives

Blood cultures (BCs) are commonly ordered in emergency departments (EDs), while a minority yields a relevant pathogen. Diagnostic stewardship is needed to safely reduce unnecessary BCs. We aimed to develop and validate a bacteremia prediction model for ED patients, with specific focus on the benefit of incorporating procalcitonin.

Methods

We included adult patients with suspected bacteremia from a Dutch ED for a one-year period. We defined 23 candidate predictors for a “full model”, of which nine were used for an automatable "basic model”. Variations of both models with C-reactive protein and procalcitonin were constructed using LASSO regression, with bootstrapping for internal validation. External validation was done in an independent cohort of patients with confirmed infection from 71 Spanish EDs. We assessed discriminative performance using the C-statistic and calibration with calibration curves. Clinical usefulness was evaluated by sensitivity, specificity, saved BCs, and Net Benefit.

Results

Among 2111 patients in the derivation cohort (mean age 63 years, 46% male), 273 (13%) had bacteremia, versus 896 (20%) in the external cohort (n = 4436). Adding procalcitonin substantially improved performance for all models. The basic model with procalcitonin showed most promise, with a C-statistic of 0.87 (0.86–0.88) upon external validation. At a 5% risk threshold, it showed a sensitivity of 99% and could have saved 29% of BCs while only missing 10 out of 896 (1.1%) bacteremia patients.

Conclusions

Procalcitonin-based bacteremia prediction models can safely reduce unnecessary BCs at the ED. Further validation is needed across a broader range of healthcare settings.

INTRODUCTION

Prompt identification of bacteremia in patients with suspected infection is crucial due to potential high morbidity and mortality without timely treatment.[1] Currently, there is no satisfying rapid test to diagnose bacteremia in the emergency department (ED). According to sepsis guidelines, the standard is to obtain two sets of blood cultures (BCs).[2] Unfortunately, culture results are only available after 24–48 h. In the meantime, physicians often decide on the start of empirical antibiotics. The yield of ED BCs is disappointing, only 4–11% identify true pathogens, while 0.2–3% actually impact treatment decisions.[3–12] Moreover, 2–8% of BCs are contaminated by skin flora, potentially resulting in longer hospital stay, increased costs and unnecessary exposure to antibiotics.[3–8] Procalcitonin is a biomarker that has been suggested to be useful to identify patients at low risk of bacteremia, with sensitivity for bacteremia ranging from 76–99% depending on the cut-off level.[13–16] Furthermore, clinical decision rules such as the Shapiro score have a sensitivity over 90%. [17–21] It has been suggested that combining the Shapiro score with PCT could save 20% of BCs while identifying all bacteremia cases.[15] However, biomarkers and decision rules have not found their way into routine clinical practice. We hypothesize that this is because simple models lack performance, while extensive decision rules might be too labor-intensive for busy EDs.[22] Recently, multiple studies have developed promising automatable models that could be implemented in electronic health records (EHRs).[23–26] Performance of such models might be improved by including PCT.

We aimed to develop an easy-to-use, automatable bacteremia prediction model to safely improve blood culture efficiency, preventing possible negative effects of false-positive BCs, while evaluating the value of incorporating procalcitonin. Moreover, we validated the potential clinical usefulness of such a model in an unselected and heterogeneous group of patients presenting with suspected infection at the ED.

MATERIALS AND METHODS

Study design and setting

We included consecutive patients aged 18 years and older with a BC taken between April 2019 and May 2020 during their ED visit at the Haga Teaching Hospital (a large hospital in The Hague, the Netherlands, with ± 60.000 ED visits annually). In the Dutch healthcare system, patients visit the ED if referred by their general practitioner or ambulance service. An automated PCT ordering procedure was installed for this prospective study. The decision to obtain a BC was left to the attending physician. PCT was measured using leftover blood, drawn for routine

chemistry concurrently with the BC. PCT was measured with an automated rapid sensitive assay (Elecsys Brahsm PCT on Roche Cobas E601, Basel, Switzerland; measuring range: 0.02–100 ng/mL). The PCT result was unavailable for the attending physician. We only included the first visit in case of multiple ED visits during the study period. The TRIPOD guideline was followed for reporting on the development and validation of multivariable prediction models.[27]

Data collection

Patient records were reviewed by one of five GCP-certified investigators. We collected data on socio-demographics, comorbidities as defined by the Charlson Comorbidity Index (CCI), medication use, and the first vital signs at the ED.[28] All numerical values were checked for impossible values. Additionally, we collected details on laboratory and microbiological results, and the clinical diagnosis in the ED.

Outcomes

The primary outcome was bacteremia. Patients were considered to have bacteremia if at least one BC obtained within 48 h of ED presentation yielded a pathogen that was not deemed a contaminant. For most cultures, this was assessed by the clinical microbiologist. In case of doubt, this was done by the hospital antibiotic stewardship team. Coagulase-negative staphylococci (CNS), Bacillus spp., Corynebacterium spp., Propionibacterium spp., Viridans group streptococci, and Lactobacillus spp. were considered contaminants unless associated with an intravascular catheter or device, suspected endocarditis, or multiple positive blood cultures with the same bacterium in a patient with a non-vascular prosthetic device (e.g. prosthetic joint). BCs were incubated using the BacTEC/ Alert automated BC system. Determination was done with gram stain and the Bruker Maldi-tof-MS Biotyper. Contaminated BCs were considered negative. Secondary outcomes were hospital admission, ICU-admission, length of hospitalization and 30-day mortality (see Supplement).

Definitions

Altered mental status was defined as a Glasgow coma scale < 15 or a note of confusion or changed behavior. Recent use of antibiotics was defined as use in the previous seven days. Diagnoses were divided into different suspected infection sites: Respiratory, Urogenital, Skin and soft tissue, Central nervous system, Abdominal, Intravascular, Bone and joint, Infection of unknown origin and No or other infection.[29] Two investigators (AK, ED) independently categorized the patients and in case of disagreement, an infectious disease specialist was consulted (CvN). Cohen's kappa was 0.87, indicating excellent agreement between assessors.

Ethics

This study was approved by the Medical Ethics Committee and the Institutional Scientific Review Board (protocol T18–040). The need for informed consent was waived given the observational nature without impact upon patient's management. This study was registered at the National Trial Register: https://onderzoekmetmensen.nl/nl/trial/23345.

Validation dataset

External validation was performed in an independent Spanish dataset from the INFURG-SEMES study group, including patients with suspected infection and BC testing from 71 EDs.[21] The inclusion criteria and collection of variables were comparable to the Dutch cohort, apart from slightly different definitions in use of immunosuppressants, use of antibiotics, altered mental status and site of infection. This validation set differed from the development as ultimately, only patients with confirmed infection were included. See the original article for an extensive description of the methodology used in the Spanish dataset.[21] As breathing frequency was dichotomized in the Spanish dataset, we used multiple imputation of the continuous breathing frequency with an indicator variable for missingness.

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 done using $\chi 2$ and Mann-Whitney-U tests, as appropriate. We predefined 23 candidate predictors based on previous literature and expert knowledge (Table 2). We balanced the number of candidate predictors to the number of bacteremia cases aiming for a minimum of 10 events per predictor to prevent over-fitting, including all patients for a one-year period to cover all seasons.[30] All candidate predictors were used for a 'full' model, whereas only readily automatable predictors were used in a 'basic' model. Prediction models were fitted without biomarkers, with either CRP or PCT alone, and with both combined. All numerical predictors were used as a continuous variable. We considered linear, logistic or spline transformations of candidate predictors as indicated upon visual inspection. Variables with extreme values were winsorized towards the 1st and 99th percentile.[30] There was a maximum of 9% missing data per candidate predictor, which were multiple imputed using chained equations.[31] Results were compared to those from complete case analysis (eTable 5). We created a stacked dataset of 10 imputed datasets, with weights of 1/10 per patient. Regression analysis with the Least Absolute Shrinkage and Selection Operator (LASSO) was used to

define the models. Odds ratios and intercepts are given without 95% confidence (CI) interval since penalization with LASSO leads to biased estimates that are optimal for prediction (eTable 6, eTable 7). Predictive performance was assessed by discrimination and calibration. Discriminative ability was assessed by the C-statistic, which is the area under the ROC curve. Calibration was evaluated by the calibration intercept, slope and graphically by plotting observed outcomes versus predicted risk. Internal validation was conducted with a 200 bootstrap resampling procedure, where the LASSO procedure was repeated in each bootstrap sample to estimate the optimism-corrected C-statistic.[30] Predictions were calculated using the predict function from the R package 'stats'. Final models were constructed in the combined development and validation dataset without predictors with coefficients close to zero. For these models, we repeated development and internal validation using a weighted LASSO regression with dataset (Spanish vs Dutch) as an additional predictor and using equal weighting for each dataset. As a safety assessment, we performed a subgroup analysis for immunocompromised patients. P-values of < 0.05 were considered statistically significant. Analyses were performed using RStudio version 2023.03.0.

Potential clinical usefulness

We calculated sensitivity, specificity, positive predictive value, negative predictive value, and the corresponding saved BCs and missed positive BCs. We report outcomes at a risk threshold of 5% for comparability to previous literature and for thresholds from 1% to 10% (eTable 3).[26] For PCT, we considered preset cut-offs of < 0.05 ng/ mL, < 0.10 ng/mL, < 0.25 ng/mL and < 0.50 ng/mL. Clinical usefulness was summarized as Net Benefit (NB), expressed as net interventions avoided per 100 patients, also known as NBOpt-out.[32–34] NB quantifies the potential impact on patient outcomes by decision-making based on a prediction model compared to reference strategies of testing all or none of the patients (see Supplement). We constructed decision curves to show NB between 1% and 25% thresholds. We consider risk thresholds of 1% to 10% most clinically relevant, and above 25% clinically implausible since any clinician would order a BC if the risk of bacteremia would be over 25%.

RESULTS

Of 2550 ED visits in the Dutch cohort, 218 were excluded because PCT was not correctly measured and 221 were patients with multiple ED visits (eFig. 1). Of 2111 unique patients, 273 were diagnosed with bacteremia (13%) and 80 patients had contaminated BCs (3.8%). A total of 2980 BCs were taken from these 2111 patients

of which 404 yielded a true pathogen (14%) and 84 (2.8%) were contaminated. For detailed culture results, see eTable 1.

The mean age in the Dutch dataset was 63 years and 54% (1147/2111) was male (Table 1). Bacteremia patients had more cardiovascular disease (146/273, 54% versus 668/1838, 36%), more diabetes mellitus (78/273, 29% versus 388/1838, 21%) and more often a history of malignancy (78/ 273, 29% versus 394/1838, 21%) (eTable 2). They suffered from more comorbidities (median CCI 4 versus 3). Furthermore, vital parameters reflected more severe disease for those with bacteremia (e.g. lower blood pressure, higher temperature, heart rate and breathing frequency).

Median PCT in bacteremia patients was 1.3 (IQR [0.43–9.3]) versus 0.18 (IQR [0.08–0.60]) for those without bacteremia ($p < 0.001$). Clinical outcomes were worse for bacteremia patients compared to those without bacteremia (See Supplement). The 4436 Spanish patients had similar demographics (mean age 67 years; 2646/4436, 60% male) and less comorbidity (CCI median of 2 [1–4], Table 1). Mean systolic blood pressure was lower and inflammatory markers were higher. Bacteremia was more frequent (896/4436, 20%), and more BCs were contaminated (227/4436, 5.1%).

Table 1: Baseline characteristics for the Dutch and Spanish dataset

Characteristics	Netherlands No Bacteremia	Netherlands Bacteremia	Spain No Bacteremia	Spain Bacteremia
	n = 1838	n = 273	n = 3540	n = 896
Demographics and medical history				
Age, years, (mean (SD))	61.9 (19.3)	69.9 (15.2)	66.5 (19.0)	70.3 (16.0)
Sex (Male)	992 (54.0)	155 (56.8)	2099 (59.3)	547 (61.0)
Sex (Female)	847 (46.1)	117 (42.9)	1441 (40.7)	349 (39.0)
Recent use of antibiotics[x]	475 (25.8)	36 (13.2)	566 (16.0)	106 (11.8)
Use of immunosupressants	220 (12.0)	32 (11.7)	256 (7.2)	75 (8.4)
Charlson Comorbidity Index (median [IQR])	3.0 [0.0, 5.0]	4.0 [2.0, 6.0]	2.0 [1.0, 4.0]	3.0 [1.0, 5.0]
Vital functions in ED				
Temperature, °C, (mean (SD))	38.3 (1.1)	38.8 (1.1)	37.8 (1.0)	38.1 (1.1)
Glasgow coma scale (median [IQR])	15.0 [15.0, 15.0]	15.0 [15.0, 15.0]	15.0 [15.0, 15.0]	15.0 [14.0, 15.0]
Heart frequency, beats/min, (mean (SD))	100.0 (21.0)	106.3 (23.9)	99.0 (19.9)	103.6 (22.5)
Systolic blood pressure, mmHg, (mean (SD))	133.8 (27.1)	123.8 (32.3)	124.3 (27.0)	114.4 (27.6)
Breathing frequency> 22 breaths/min	536 (32.3)	113 (44.5)	1115 (34.5)	523 (58.5)
Breathing frequency, breaths/min, (median [IQR])	20.0 [16.0, 24.0]	20.5 [16.0, 28.0]		
Saturation, %, (median [IQR])	96.0 [94.0, 98.0]	95.0 [93.0, 97.0]		
Laboratory values in ED				
C-reactive protein, mg/L, (median [IQR])	68.0 [23.0, 159.5]	122.0 [38.0, 238.0]	110.0 [50.0, 211.0]	159.0 [81.2, 265.5]
Procalcitonin, µg/L, (median [IQR])	0.2 [0.1, 0.6]	1.3 [0.4, 9.2]	0.2 [0.1, 0.5]	3.1 [1.1, 10.7]
White blood cell count, ×10^9/L, (median [IQR])	11.1 [7.8, 15.3]	12.6 [8.9, 17.3]	11.6 [8.0, 16.3]	15.1 [10.3, 19.3]
Creatinine, µmol/L, (median [IQR])	80.5 [65.0, 106.0]	100.0 [78.0, 140.5]		
Suspected site of infection[a]				
Respiratory	751 (40.9)	43 (15.8)	1783 (50.4)	187 (20.9)

Characteristics	Netherlands No Bacteremia	Netherlands Bacteremia	Spain No Bacteremia	Spain Bacteremia
	n = 1838	n = 273	n = 3540	n = 896
Urogenital	263 (14.3)	96 (35.2)	894 (25.3)	382 (42.6)
Abdominal	182 (9.9)	43 (15.8)	355 (10.0)	170 (19.0)
Skin and soft tissue	145 (7.9)	24 (8.8)	163 (4.6)	45 (5.0)
Central nervous system	15 (0.8)	7 (2.6)	16 (0.5)	18 (2.0)
Unknown site of infection	113 (6.1)	20 (7.3)	221 (6.2)	60 (6.7)
Other infection[b]	98 (5.3)	4 (1.5)	108 (3.1)	34 (3.8)
Bone and joint	22 (1.2)	16 (5.9)		
Intravascular	12 (0.7)	11 (4.0)		
No infection	239 (13.0)	9 (3.3)		

Table 1. Baseline characteristics for the Dutch (derivation) and Spanish (validation) dataset. Data are presented as number (percentage) of patients unless otherwise indicated. SD = Standard Deviation, IQR = Interquartile Range, ED = Emergency Department. Missing in Dutch data: (n (%)): Temperature 3 (0.1), Glasgow coma scale 27 (1.3), Heart frequency 21 (1.0), Systolic blood pressure 22 (1.0), Breathing frequency 197 (9.3), Saturation 23 (1.1), Creatinine 7 (0.3). Missing in Spanish data (n (%)): Temperature 14 (0.3), Glasgow coma scale 91 (2.1), Heart frequency 67 (1.5), Systolic blood pressure 30 (0.7), Breathing frequency >22 307 (6.9), C-reactive protein 362 (8.2), Procalcitonin 140 (3.2), White blood cell count 15 (0.3). [x] Recent use of antibiotics in the previous seven days. [a] Two patients were diagnosed with multiple infections. [b] Composite category of multiple infections such as viral infections or ear-nose-throat infections.

Table 2: Bacteremia prediction models with odds ratios per predictor

Candidate predictors	Basic without biomarkers	Basic with CRP	Basic with PCT	Basic with CRP and PCT	Full without biomarkers	Full with CRP	Full with PCT	Full with CRP and PCT	Only PCT
Age	1.4	1.4	1.3	1.3	1.4	1.4	1.3	1.3	
Gender									
SBP	1.9	1.9	1.3	1.2	1.9	1.8	1.3	1.2	
BFR		0.998	0.95	0.95	1.2	1.2	1.1	1.1	
HFR	1.2	1.2	1.2	1.2	1.4	1.4	1.3	1.3	
Temperature	2.2	2.2	2.1	2.1	1.9	2.0	2.0	2.0	
GCS	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	
CRP		1.1		0.95		1.1		0.97	
PCT			1.4	1.4			1.4	1.4	1.4
Antibiotics					0.4	0.4	0.4	0.4	
Immuno						1.03	1.01	1.01	
AMS					0.9	0.9			
Catheter					3.5	3.5	4.4	4.4	
Chills					1.6	1.6	1.4	1.4	
Site of infection[a]					0.9-19	0.7-14	0.7-17	0.7-19	

Table 2. Overview of all models derived from Dutch derivation dataset with corresponding odds ratio per predictor. Predictors in red were not selected for the final model in the LASSO procedure. Grey variables were not considered. CRP = C-reactive protein, PCT = procalcitonin, SBP = systolic blood pressure, BFR = breathing frequency, HFR = heart frequency, GCS = glasgow coma scale, Antibiotics = Use of antibiotics in previous 7 days, Immuno = use of immunosuppressants, AMS = altered mental status, Catheter = indwelling vascular catheter, Chills = reported chills during ED visit. Possible infection sites were: No infection/non-bacterial infection, respiratory, skin and soft tissue, unknown, abdominal, urogenital, central nervous system, bone and joint, intravascular. The odds ratios reflect the following change in values: Age = per 10 years increase, SBP = per 20 mmHg decrease, BFR = increase of breathing frequency of 5/min, HFR = per increase in heartrate of 30 beats/min, GCS for each point decrease, CRP and PCT = for a doubled value. [a] For odds ratios per infection site see Supplemental material.

Table 3: Performance of all models upon external validation in the Spanish dataset.

	Performance							
	Discrimination		Calibration		Outcomes for a 5% risk threshold			
Model	Internal (95% CI)	External (95% CI)	Intercept (95% CI)	Slope (95% CI)	Saved BC (%)	Missed TP (%)	Sens (%)	NPV (%)
Basic without biomarkers	0.72 (0.69 - 0.75)	0.66 (0.64-0.68)	0.53 (0.45-0.60)	0.74 (0.64-0.85)	10.4	31 (3.5)	96.5	93.3
Basic with CRP	0.73 (0.70 - 0.76)	0.67 (0.65-0.69)	0.45 (0.37-0.53)	0.77 (0.67-0.87)	10.4	26 (2.9)	97.1	94.4
Basic with PCT	0.81 (0.78 - 0.84)	0.87 (0.86-0.88)	0.59 (0.51-0.67)	1.44 (1.34-1.53)	28.5	10 (1.1)	98.9	99.2
Basic with CRP and PCT	0.81 (0.78 - 0.84)	0.88 (0.86-0.89)	0.62 (0.54-0.70)	1.42 (1.33-1.52)	29.8	11 (1.2)	98.9	99.2
Full without biomarkers	0.82 (0.80 - 0.85)	0.76 (0.74-0.77)	0.27 (0.19-0.35)	0.81 (0.74-0.88)	25.7	63 (7.0)	93.0	94.5
Full with CRP	0.83 (0.80 - 0.85)	0.77 (0.75-0.78)	0.20 (0.12-0.28)	0.83 (0.76-0.90)	25.3	52 (5.8)	94.2	95.4
Full with PCT	0.86 (0.84 - 0.88)	0.88 (0.86-0.89)	0.41 (0.33-0.50)	1.20 (1.12-1.28)	37.4	25 (2.8)	97.2	98.5
Full with CRP and PCT	0.86 (0.84 - 0.88)	0.88 (0.87-0.89)	0.43 (0.35-0.52)	1.20 (1.12-1.28)	38.1	25 (2.8)	97.2	98.5
PCT only	0.79 (0.76 - 0.82)	0.90 (0.89-0.91)	0.46 (0.38-0.54)	1.67 (1.57-1.78)	14.5	3 (0.3)	99.7	99.5

Table 3. Performance of all models upon external validation in the Spanish dataset. CRP = C-reactive protein, PCT = procalcitonin, CI = confidence interval, BC = blood cultures, TP = True positive. Outcomes are given at patient-level. For example: applying the basic model without biomarkers would have saved blood cultures in 457 out of 4436 (10.3%) patients, but it would have missed 3.5% of all 896 true positive blood cultures.

Prediction models, performance and clinical usefulness

Various prediction models were constructed (Table 2). Upon development, the basic models including PCT showed the highest discriminative performance with C-statistics of 0.81 (0.78–0.84) (Table 3). Adding clinical information increased the C-statistic beyond those of all basic models. Again, including PCT in the full models led to a substantial increase in C-statistic. Upon external validation, the effect of adding PCT to the basic models led to even higher C-statistics of 0.87 (0.86–0.88). In contrast with the Dutch dataset, adding clinical variables did not show additional benefit over PCT. Upon external validation, the basic model with PCT saved BCs in 29% (1266/4366) of patients while only missing 10 out of 896 (1.1%) bacteremia patients. Adding CRP would have missed one more bacteremia while saving another 1.3% of BCs. A model with procalcitonin only had a C-statistic of 0.79 (0.76–0.82) at internal validation and 0.90 (0.89–0.91) at external validation. This model saved 15% (643/4436) of BCs and missed 3 out of 896 bacteremia patients. Prediction models had limited impact on BC ordering if a risk threshold of 1% was applied, and more impact for a 10% threshold (eTable 3). Overall, calibration showed systemic under-estimation of risk in the Spanish cohort, with intercepts of all models significantly different from 0 (Fig. 1). The calibration plots of models including procalcitonin showed a S-shape with overestimation of risks in the lower risk ranges and underestimation in the higher risk ranges, reflected in slopes significantly different from 1 (Fig. 1). A set of final updated models can be found in eTable 7. Finally, both the original and updated models that included PCT provided a net reduction of BCs across the vast majority of risk thresholds, specifically in the relevant range of thresholds between 1% and 10%. (Fig. 2).

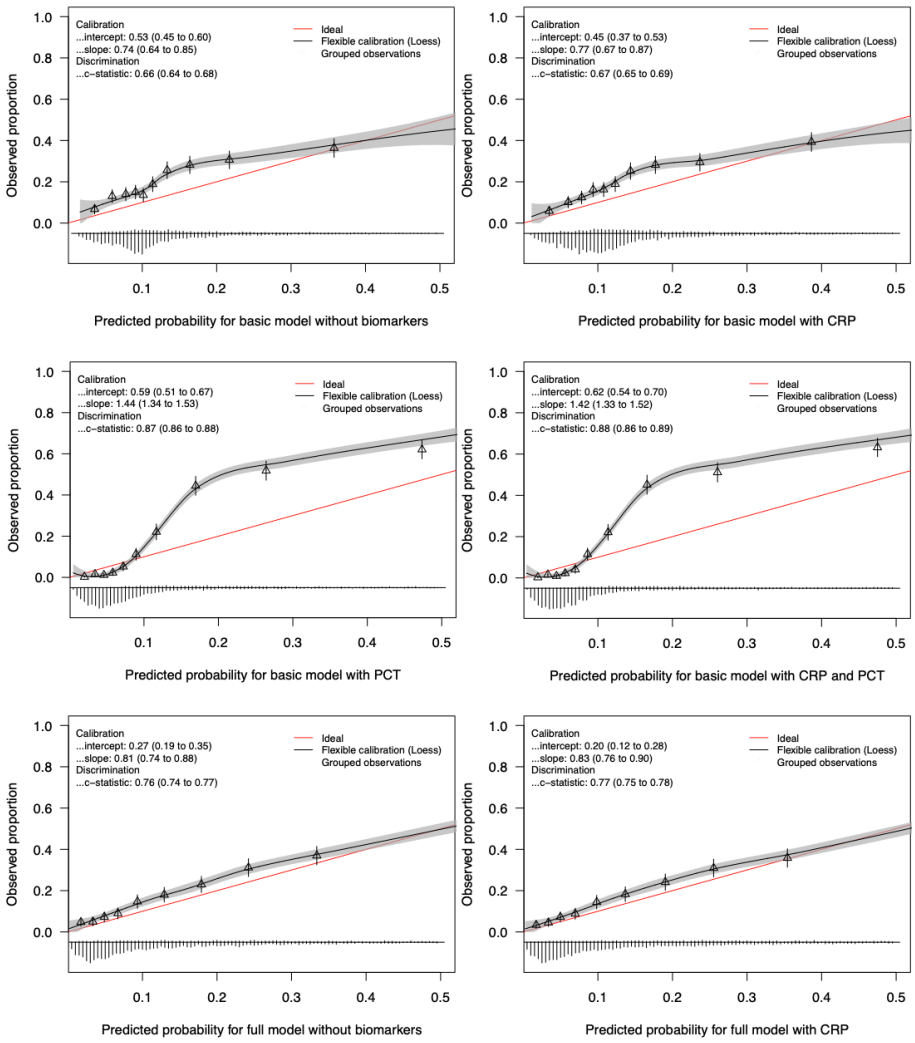
Missed bacteremia

Applying a 5% risk threshold in the basic with PCT model in the Dutch data would have withheld BCs in 733/2111 (35%) of patients. This reduction would imply missing 12 bacteremia patients. A chart review suggested a probable negative impact for 2 out of 12 patients and one case with possible negative impact. For the other 9 patients, no negative impact was expected (eTable 8).

Subgroup analysis

Among 331 immunocompromised Spanish patients, 75 (23%) had bacteremia. Using the basic model + PCT at the 5% risk threshold, none of those would have been missed while avoiding 27% (89/331) of BCs. The C-statistic was 0.90 (0.86–0.93), and the underestimation of risk was similar to the full validation set (calibration intercept 0.71, 0.42–1.00).

Figure 1: Calibration plots of all models applied in the Spanish dataset



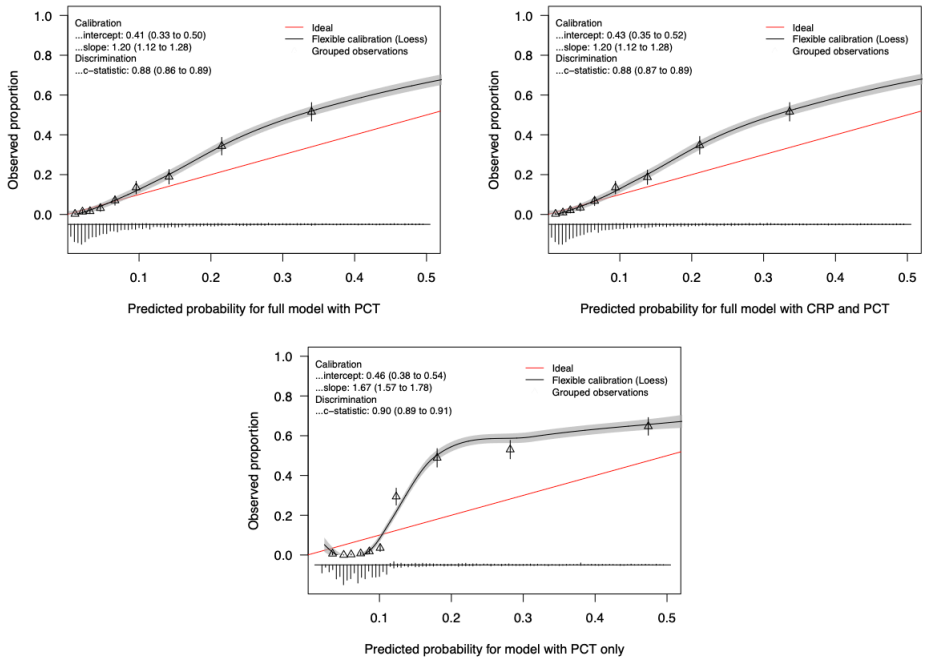


Figure 1. Calibration plots of all models applied in the Spanish dataset. X-axis represents predicted risks, Y-axis is the observed risks in the Spanish dataset. CRP = C-reactive protein, PCT = procalcitonin.

Figure 2: Net reduction of blood cultures

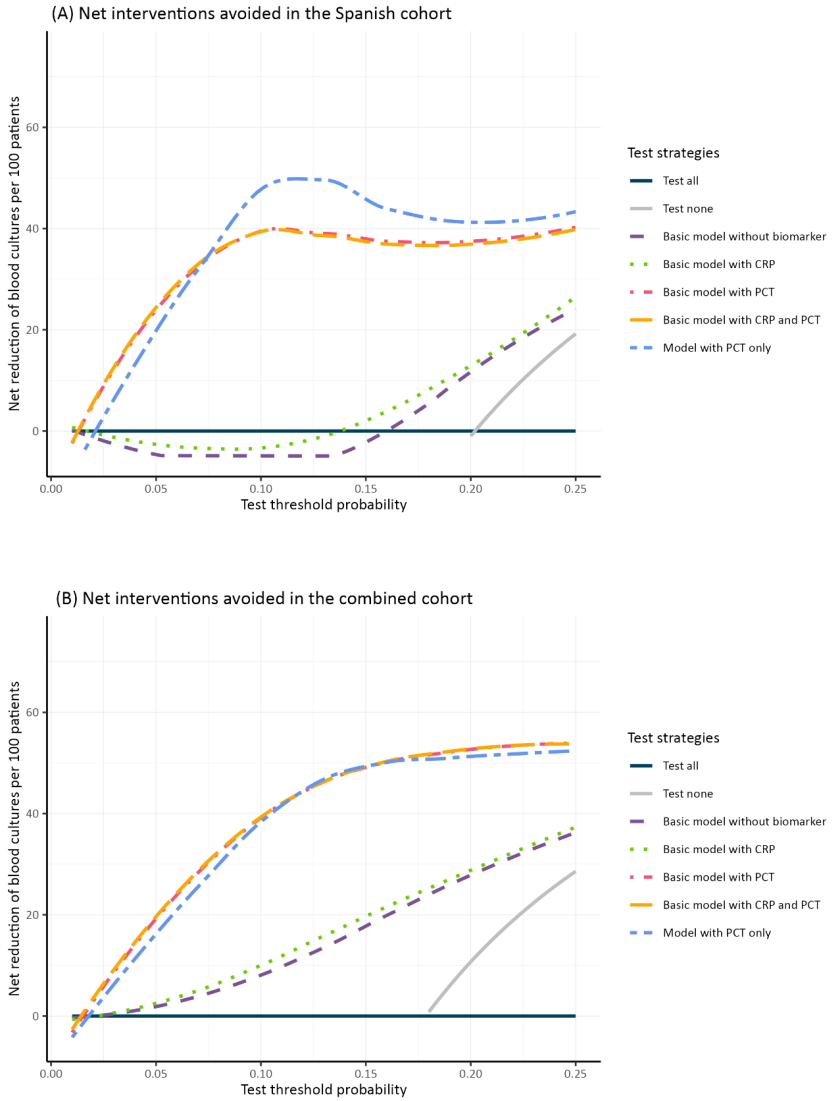


Figure 2. Decision curve analysis showing the expected net reduction of blood cultures per patient for models in the Spanish cohort (A) and in the combined cohorts (B). The net reduction of blood cultures relative to testing everyone ("Test all") is shown. The x-axis displays the possible risk thresholds 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 BenefitOpt-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. At a 5% threshold, the net benefit of the basic model with PCT is 0.24, calculated as ((1256/3540)*0.798)-((10/896)*0.202*(0.95/0.05)). See the supplemental material for more information about Net Benefit.

DISCUSSION

We found a substantial incremental value of PCT in multiple bacteremia prediction models, enabling a safe reduction in BC ordering. Five of our nine bacteremia prediction models were designed to be fully automatable, allowing for easy validation and implementation in EHRs. Considering both predictive performance and applicability, the basic model with PCT seems most promising. It could potentially save 29% of BCs, while only missing 1% of bacteremia patients.

While we investigated multiple models, we focused on the basic with PCT model since including PCT led to substantially improved predictive performance. This is in line with previous research which showed an increase in the C-statistic of the Shapiro score from 0.73 to 0.80 by adding PCT.[15] In contrast, Lien et al. showed no added benefit of PCT to a machine learning model based on a complete blood count.[35] The added benefit of detailed clinical information was limited in both cohorts, while the manual data-entry required for such information would be a barrier for implementation at the ED. Although the basic model with both PCT and CRP as biomarkers was able to save slightly more BCs, the incremental value might be considered too limited to justify testing CRP for all patients. PCT had substantial value on its own already, particularly in the Spanish data set.

Interpretation

A previously developed machine learning model demonstrated a C-statistic of 0.81 (0.78–0.83) in a test cohort and, in a prospective validation, saved 30% of BCs while missing 6.6% of bacteremia patients at a 5% risk threshold.[26] This reflects a sensitivity of 93%, while our basic with PCT model had a sensitivity of 99%, which is what clinicians appreciate for a bacteremia prediction model.[22] Our models have the advantage of being transparent with respect to the predictors and their weights, unlike machine learning models in which variables are unavailable to the clinician, a 'black-box'. Second, choosing candidate predictors based on known physiological processes is expected to lead to more robust models with validity in different settings.[36] Finally, being able to provide the regression equation also enables more straightforward validation and implementation in new settings.[37]

The discriminative performance of our models was higher in the Spanish cohort than at model development, but calibration showed systematic underestimation. This might relate to the different patient selection process for the Spanish dataset, as only patients with confirmed infection were included. Our models were unable to fully capture this difference in average bacteremia risk. Furthermore, the predictive value of PCT was very strong in the Spanish dataset, which explains the improved performance for all models containing PCT upon external validation.

We do not have a clear explanation for this difference in diagnostic strength of PCT. The miscalibration motivated updating of the prediction models based on both datasets (eTable 6). These updated models require further external validation before considering clinical implementation. By following an iterative process of model updating and validation, we aim to create a broadly applicable bacteremia prediction model.[38,39]

In times where diagnostic stewardship becomes increasingly important, a bacteremia prediction model like ours meets a pressing need. However, no model will be perfect, and the use of such a model will inevitably miss some bacteremia patients. Although clinicians are hesitant to miss bacteremia, there is extensive literature on the lack of treatment consequences in the majority of positive BCs.[3,9,10] Indeed, upon clinical review, missing bacteremia in patients with a predicted low risk was generally deemed clinically irrelevant, except in those with suspected endocarditis (eTable 8). We therefore do not recommend using our models in patients with a clear indication for BCs, such as suspected endocarditis or central-line associated bacteremia. Notably, the model performed well in the Spanish immunocompromised patients, missing none of 75 bacteremia cases at a threshold of 5%.

Societal value

Apart from the clinical usefulness of our models, they also have potential societal value through cost reduction. Applying the basic with PCT model in the Spanish dataset would save 571 negative BCs per 1000 patients, assuming two BCs per patient. Additionally, it would prevent 18 false positive BCs per 1000 patients. Assuming 96 to 128 USD per negative BC, this would potentially save k$55 to k$73 per 1000 patients.[7] At estimated costs of $2923 to $5812, avoiding the false positive BCs would save an additional $53.000 to $105.000, leading to a total saving of $108.000 and $178.000 per 1000 patients. Testing PCT for 1000 patients is estimated to cost $51.000 to $96.000.[40,41] This would result in a potential net cost-saving of $12.000 to $127.000 per 1000 patients. We recognize that the costs of BCs and PCT testing differ across the world, and PCT testing can be costly, which will influence the potential cost-saving. Additionally, it is difficult to quantify the costs of missing a bacteremia. In the future, a more extensive cost-benefit analysis should be performed to quantify societal value.

Implementation

There are two main things to be considered before implementing these models in clinical practice. First, we believe that another validation study of the updated models is essential to increase our understanding of the observed geographical

differences. Second, an implementation study should evaluate whether the models indeed lead to a safe reduction of BCs, while exploring how clinicians use such models and what risk of bacteremia they are willing to accept.

We believe that our model should only be used in patients where the clinician sees an indication for BCs. This would prevent using the model in a broader range of patients, which could potentially lead to an unwanted increase of BC testing. It is important to realize that we only included patients with a clinical indication for BCs, and do not know the model performance in patients where the clinician would normally not have taken BCs. Finally, it is important to realize that the potential cost-saving will be different per setting, mainly depending on the costs of BC and PCT testing. In places where PCT testing is not already implemented, high costs of PCT might be a barrier for implementation. However, we believe that especially at EDs already using PCT, the model has the potential to lead to significant cost reduction and can prevent patient harm related to false- positive BCs.

Strengths and limitations

The main strength of this study is the development and external validation of our models in two large international datasets with high-quality data and minimal missing values. While some definitions differed slightly between the development and validation cohorts, we consider the data a good reflection of clinical practice across different geographic locations as clinicians will interpret variables slightly different across settings.[39] Furthermore, we applied state-of-the-art statistical methods to develop robust prediction models. This robustness is underlined by the fact that the models performed well even though the validation cohort consisted only of confirmed infections, in contrast to suspected infections in the development data. A limitation is that we observed some unexplained geographic variation in statistical performance (discrimination and calibration) and clinical usefulness (Net Benefit). This motivates validating the updated models in other geographical settings. Finally, we emphasized clinical interpretability and applicability, which will make implementation in clinical practice more feasible.

CONCLUSIONS

PCT should be considered in any bacteremia prediction model because of considerably improved performance beyond clinical information alone. PCT-based bacteremia prediction models can potentially safely reduce a large number of unnecessary BCs at the ED, motivating further validation across a range of geographical locations and healthcare settings.

Funding

This project was supported by an unrestricted grant from the Haga Scientific Research Fund. This did not influence the study design, the collection and analysis of data, the writing of the manuscript or the decision to submit for publication.

CRediT authorship contribution statement

AK: Conceptualization, Methodology, Data curation, Formal analysis, Validation, Visualization, Writing – original draft, Writing – review & editing. **SM:** Conceptualization, Methodology, Data curation, Visualization, Writing – review & editing, Funding acquisition. **NvB:** Resources, Writing – review & editing. **MD:** Resources, Writing – review & editing. **NK:** Writing – review & editing. **PM:** Writing – review & editing. **AJ-J:** Writing – review & editing, Data curation. **ES:** Conceptualization, Methodology, Formal analysis, Validation, Visualization, Writing – review & editing, Supervision. **CvN:** Conceptualization, Resources, Methodology, Visualization, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used OpenAI in order to reduce word count. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content for publication.

Data availability

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.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: ES received royalties from Springer Verlag for his book “Clinical Prediction Models. AJJ declares to have received payments for presentations and lectures for Roche Diagnostics, Thermo Fisher Scientific, B.R.A.H.M.S. AG, Biomerieux and Virogates. All other authors reported no competing interests.

Acknowledgments

We want to thank Dagmar Brouwer, Dide van de Fliert, Nicoline Meijs, Lars Snel and Emma Dregmans for their help with retrieving all clinical information from the electronic patient records. Furthermore, we want to thank the INFURG-SEMES study group for sharing their data.

SUPPLEMENTAL MATERIAL

REFERENCES

1. Jiménez-Aguilar P, López-Cortés LE, Rodríguez-Baño J. *Impact of infectious diseases consultation on the outcome of patients with bacteraemia. Ther Adv Infect Dis* 2019 Dec 6;**6**:2049936119893576. https://doi.org/10.1177/2049936119893576. PMID: 31839942; PMCID: PMC6900613.
2. Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, et al. *Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock. Intensive Care Med* 2016;**43**(3):304–77. https://doi.org/10.1007/s00134- 017-4683-6. Epub 2017 Jan 18. PMID: 28101605.
3. Long B, Koyfman A. *Best clinical practice: blood culture utility in the emergency department. J Emerg Med* 2016 Nov;**51**(5):529–39. https://doi.org/10.1016/j. jemermed.2016.07.003. Epub 2016 Sep 14. PMID: 27639424.
4. 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. https:// doi.org/10.1001/jama.2012.8262
5. Nannan Panday RS, Wang S, van de Ven PM, Hekker TAM, Alam N, Nanayakkara PWB. *Evaluation of blood culture epidemiology and efficiency in a large European teaching hospital. PLoS One* 2019 Mar 21;**14**(3):e0214052. https://doi.org/10.1371/ journal.pone.0214052. PMID: 30897186; PMCID: PMC6428292.
6. Dargère S, Cormier H, Verdon R. *Contaminants in blood cultures: importance, implications, interpretation and prevention. Clin Microbiol Infect* 2018 Sep;**24**(9):964–9. https://doi.org/10.1016/j.cmi.2018.03.030. Epub 2018 Apr 3. PMID: 29621616.
7. 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):963–7. https://doi.org/10.1016/j.ajic.2018.12.020. Epub 2019 Feb 20. PMID: 30795840.
8. 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.
9. Howie N, Gerstenmaier JF, Munro PT. *Do peripheral blood cultures taken in the emergency department influence clinical management? Emerg Med J* 2007 Mar;**24**(3):213–4. https://doi.org/10.1136/emj.2006.039875. PMID: 17351231; PMCID: PMC2660034.
10. Mountain D, Bailey PM, O'Brien D, Jelinek GA. *Blood cultures ordered in the adult emergency department are rarely useful. Eur J Emerg Med* 2006 Apr;**13**(2):76–9. https://doi.org/10.1097/01.mej.0000188231.45109.ec. PMID: 16525233.
11. Fabre V, Milstone AM, Keller SC, Carroll KC, Cosgrove SE. *Prescribers' knowledge, attitudes and perceptions about blood culturing practices for adult hospitalized patients: a call for action. Infect Control Hosp Epidemiol* 2018 Nov;**39**(11):1394–6. https://doi.org/10.1017/ice.2018.224. Epub 2018 Sep 18. PMID: 30226121.
12. Dräger S, Giehl C, Søgaard KK, Egli A, de Roche M, Huber LC. *Do we need blood culture stewardship programs? A quality control study and survey to assess the appropriateness of blood culture collection and the knowledge and attitudes among physicians in Swiss hospitals. Eur J Intern Med* 2022;**103**:50–6. https://doi.org/10. 1016/j.ejim.2022.04.028. Epub 2022 Jun 14. PMID: 35715280.

13. 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. https://doi.org/10.1016/j.cmi.2014.12. 026. Epub 2015 Jan 14. PMID: 25726038.

14. van Nieuwkoop C, Bonten TN, van't Wout JW, Kuijper EJ, Groeneveld GH, Becker MJ, et al. *Procalcitonin reflects bacteremia and bacterial load in urosepsis syndrome: a prospective observational study. Crit Care* 2010;**14**(6):R206. https://doi.org/10. 1186/cc9328. , Epub 2010 Nov 17. PMID: 21083886; PMCID: PMC3220019.

15. Laukemann S, Kasper N, Kulkarni P, Steiner D, Rast AC, Kutz A, et al. *Can we reduce negative blood cultures with clinical scores and blood markers? Results from an observational cohort study. Medicine* 2015;**94**(49):e2264. https://doi.org/10.1097/MD.0000000000002264

16. Goodlet KJ, Cameron EA, Nailor MD. *Low sensitivity of procalcitonin for bacteremia at an academic medical center: a cautionary tale for antimicrobial stewardship. Open Forum Infect Dis* 2020 Mar 16;**7**(4):ofaa096. https://doi.org/10.1093/ofid/ofaa096. PMID: 32322602; PMCID: PMC7162616.

17. Shapiro NI, Wolfe RE, Wright SB, Moore R, Bates DW. *Who needs a blood culture? A prospectively derived and validated prediction rule. J Emerg Med* 2008 Oct;**35**(3):255–64. https://doi.org/10.1016/j.jemermed.2008.04.001. Epub 2008 May 16. PMID: 18486413.

18. Bates DW, Cook EF, Goldman L, Lee TH. *Predicting bacteremia in hospitalized patients. A prospectively validated model. Ann Intern Med* 1990 Oct 1;**113**(7):495–500. https://doi.org/10.7326/0003-4819-113-7-495. PMID: 2393205.

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

20. Julián-Jiménez A, García-Lamberechts EJ, González Del Castillo J, Navarro Bustos C, Llopis-Roca F, Martínez-Ortiz de Zarate M, et al. *Validation of a predictive model for bacteraemia (MPB5-Toledo) in the patients seen in emergency departments due to infections. Enferm Infecc Microbiol Clin* 2022;**40**(3):102–12. https://doi.org/10.1016/ j.eimce.2021.12.006. Epub 2022 Jan 4. PMID: 34992000.

21. Julián-Jiménez A, González Del Castillo J, García-Lamberechts EJ, Huarte Sanz I, Navarro Bustos C, Rubio Díaz R, et al. *A bacteraemia risk prediction model: development and validation in an emergency medicine population. Infection* 2022;**50**(1):203–21. https://doi.org/10.1007/s15010-021-01686-7. Epub 2021 Sep 6. PMID: 34487306.

22. Lautenbach E, Localio R, Nachamkin I. *Clinicians required very high sensitivity of a bacteremia prediction rule. J Clin Epidemiol* 2004 Oct;**57**(10):1104–6. https://doi.org/10.1016/j.jclinepi.2004.03.003. PMID: 15528062.

23. Ratzinger F, Dedeyan M, Rammerstorfer M, Perkmann T, Burgmann H, Makristathis A, et al. *A risk prediction model for screening bacteremic patients: a cross sectional study. PLoS One* 2014;**9**(9):e106765. https://doi.org/10.1371/journal. pone.0106765. PMID: 25184209; PMCID: PMC4153716.

24. Choi DH, Hong KJ, Park JH, Shin SD, Ro YS, Song KJ, et al. *Prediction of bacteremia at the emergency department during triage and disposition stages using machine learning models. Am J Emerg Med* 2022;**53**:86–93. https://doi.org/10.1016/j.ajem. 2021.12.065. Epub 2022 Jan 1. PMID: 34998038.

25. Boerman AW, Schinkel M, Meijerink L, van den Ende ES, Pladet LC, Scholtemeijer MG, et al. *Using machine learning to predict blood culture outcomes in the emergency department: a single-centre, retrospective, observational study. BMJ Open* 2022;**12**(1):e053332. https://doi.org/10.1136/bmjopen-2021-053332. PMID: 34983764; PMCID: PMC8728456.

26. Schinkel M, Boerman AW, Bennis FC, Minderhoud TC, Lie M, Peters-Sengers H, 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;**82**:104176. https://doi.org/10.1016/j.ebiom.2022.104176. Epub 2022 Jul 16. PMID: 35853298; PMCID: PMC9294655.

27. 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.

28. Charlson ME, Pompei P, Ales KL, MacKenzie CR. *A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis* 1987;**40**(5):373–83. https://doi.org/10.1016/0021-9681(87)90171-8. PMID: 3558716.

29. Dregmans E, Kaal AG, Meziyerh S, Kolfschoten NE, van Aken MO, Schippers EF, et al. *Analysis of variation between diagnosis at admission vs discharge and clinical outcomes among adults with possible bacteremia. JAMA Netw Open* 2022;**5**(6):e2218172. https://doi.org/10.1001/jamanetworkopen.2022.18172. PMID: 35737389; PMCID: PMC9226997.

30. Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating.* New York: Springer; 2009.

31. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. *Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res* 2011 Mar;**20**(1):40–9. https://doi.org/10.1002/mpr.329. PMID: 21499542; PMCID: PMC3074241.

32. Vickers AJ, van Calster B, Steyerberg EW. *A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res* 2019 Oct 4;**3**:18. https://doi.org/10. 1186/s41512-019-0064-7. PMID: 31592444; PMCID: PMC6777022.

33. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, et al. *Reporting and interpreting decision curve analysis: a guide for investigators. Eur Urol* 2018;**74**(6):796–804. https://doi.org/10.1016/j.eururo.2018.08.038. PMID: 30241973; PMCID: PMC6261531.

34. Kerr KF, Brown MD, Marsh TL, Janes H. *Assessing the clinical impact of risk models for opting out of treatment. Med Decis Mak* 2019 Feb;**39**(2):86–90. https://doi.org/10.1177/0272989X18819479. Epub 2019 Jan 16. PMID: 30649998; PMCID: PMC6374190.

35. Lien F, Lin HS, Wu YT, Chiueh TS. *Bacteremia detection from complete blood count and differential leukocyte count with machine learning: complementary and competitive with C-reactive protein and procalcitonin tests. BMC Infect Dis* 2022 Mar 26;**22**(1):287. https://doi.org/10.1186/s12879-022-07223-7. PMID: 35351003; PMCID: PMC8962279.

36. Steyerberg EW, Eijkemans MJ, Harrell Jr FE, Habbema JD. *Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med* 2000 Apr 30;**19**(8):1059–79. https://doi.org/10.1002/ (sici)1097-0258(20000430)19:8<1059::aid-sim412>3.0.co;2-0

37. Van Calster B, Steyerberg EW, Collins GS. *Artificial intelligence algorithms for medical prediction should be nonproprietary and readily available. JAMA Intern Med* 2019 May 1;**179**(5):731. https://doi.org/10.1001/jamainternmed.2019.0597. PMID: 31058938.

38. 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 Method* 2022 Dec 12;**22**(1):316. https:// doi.org/10.1186/s12874-022-01801-8. PMID: 36510134; PMCID: PMC9742671.

39. de Hond AAH, Shah VB, Kant IMJ, Van Calster B, Steyerberg EW, Hernandez- Boussard T. *Perspectives on validation of clinical predictive algorithms. NPJ Digit Med* 2023 May 6;**6**(1):86. https://doi.org/10.1038/s41746-023-00832-9. PMID: 37149704; PMCID: PMC10163568.

40. Buendía JA, Guerrero Patiño D. *Cost-effectiveness of procalcitonin for detection of serious bacterial infections in children presenting with fever without source. BMC Pediatr* 2022 Apr 26;**22**(1):226. https://doi.org/10.1186/s12887-022-03293-3. PMID: 35473509; PMCID: PMC9040337.

41. Voermans AM, Mewes JC, Broyles MR, Steuten LMG. *Cost-effectiveness analysis of a procalcitonin-guided decision algorithm for antibiotic stewardship using real-world U.S. hospital data. OMICS* 2019 Oct;**23**(10):508–15. https://doi.org/10.1089/omi. 2019.0113. Epub 2019 Sep 11. PMID: 31509068; PMCID: PMC6806362.