Systematic review finds risk of bias and applicability concerns for models predicting central line-associated bloodstream infection (CLA-BSI)

Journal of Clinical Epidemiology 161 (2023) 127−139

Journal of Clinical Epidemiology

REVIEW ARTICLE

Systematic review finds risk of bias and applicability concerns for models predicting central line-associated bloodstream infection

Shan Gao[a], Elena Albu[a], Krizia Tuand[b], Veerle Cossey[a,c], Frank Rademakers[d], Ben Van Calster[a,e,f,1,*], Laure Wynants[a,f,g,1]

[a]*Department of Development and Regeneration, KU Leuven, Leuven, Belgium*
[b]*2Bergen - Learning Centre Désiré Collen, KU Leuven Libraries, KU Leuven, Leuven, Belgium*
[c]*Department of Infection Control and Prevention, University Hospitals Leuven, Leuven, Belgium*
[d]*Faculty of Medicine, KU Leuven, Leuven, Belgium*
[e]*Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands*
[f]*EPI-Center, KU Leuven, Leuven, Belgium*
[g]*Care & Public Health Research Institute, Maastricht University, Maastricht, The Netherlands*

Accepted 26 July 2023; Published online 2 August 2023

Abstract

Objectives: To systematically review the risk of bias and applicability of published prediction models for risk of central line-associated bloodstream infection (CLA-BSI) in hospitalized patients.

Study Design and Setting: Systematic review of literature in PubMed, Embase, Web of Science Core Collection, and Scopus up to July 10, 2023. Two authors independently appraised risk models using CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) and assessed their risk of bias and applicability using Prediction model Risk Of Bias ASsessment Tool (PROBAST).

Results: Sixteen studies were included, describing 37 models. When studies presented multiple algorithms, we focused on the model that was selected as the best by the study authors. Eventually we appraised 19 models, among which 15 were regression models and four machine learning models. All models were at a high risk of bias, primarily due to inappropriate proxy outcomes, predictors that are unavailable at prediction time in clinical practice, inadequate sample size, negligence of missing data, lack of model validation, and absence of calibration assessment. 18 out of 19 models had a high concern for applicability, one model had unclear concern for applicability due to incomplete reporting.

Conclusion: We did not identify a prediction model of potential clinical use. There is a pressing need to develop an applicable model for CLA-BSI.

Keywords: Risk prediction; Central line-associated bloodstream infection; CLA-BSI; Prediction model; Central venous catheter; Bloodstream infection

1. Introduction

Central line-associated bloodstream infections (CLA-BSIs) are bloodstream infections associated with an onset at least 48 hours after the insertion of a central line in the absence of infection at another site [1]. As the most common source of hospital-acquired infection [2], this type of infection is a priority target for prevention, as they cause not only higher morbidity and mortality, but also longer length of stay (LOS) and increased hospital expenditures [3−5]. In the United States, up to 41,000 patients in hospitals acquire CLA-BSI each year [6], and a CLA-BSI was associated with an estimated mean attributable cost of $55,646 and attributable LOS of 19 days compared with those without CLA-BSIs [7]. A study performed in Germany has reported the cost to be €29,909 per CLA-BSI with median attributable LOS of 7 days [8].

Studies have shown that up to 70% of CLA-BSIs are preventable with evidence-based strategies for central line insertion and maintenance [9,10]. For the improvement of infection prevention and control in hospital, some tools

Data availability: I have shared the data extraction form as well as the corresponding results in Supplementary File 3.

[1] B. Van Calster and L. Wynants contributed equally to this work.

* Corresponding author. Herestraat 49 - Box 805, Leuven 3000, Belgium. Tel: +32-16-377788; fax: +32-16-343842.
E-mail address: ben.vancalster@kuleuven.be (B. Van Calster).

https://doi.org/10.1016/j.jclinepi.2023.07.019
0895-4356/

What is new?

Key findings

- Nineteen models which were developed and/or validated to predict central line-associated bloodstream infection (CLA-BSI) for hospitalized patients were all at high risk of bias and had high or unclear concerns regarding applicability.

What this adds to what was known?

- Though increasing number of models and tools have been developed to help improve the infection control in hospital with the popularity of electronic health records, there is still a pressing need to develop a clinically and practically useable model for the risk prediction of CLA-BSI.

What is the implication and what should change now?

- There is an urgent need to improve the methodological conduct of risk prediction model development studies. Moreover, further research may look for the potential for dynamic risk prediction models which allow timely adaption of patient management if needed.

have been developed to predict the infection risks for individuals, in conjunction with hospital-wide prevention strategies. In this systematic review, we summarized and evaluated the current risk prediction models developed for CLA-BSI. The findings of this review synthesize the advantages and disadvantages of known risk prediction models for CLA-BSI and raise some questions about the practical implementation of these models.

2. Methods

We conducted a systematic review of CLA-BSI risk prediction models to investigate previously published literature predicting the risk of CLA-BSI. The research protocol was registered (April 27, 2022) in the International Prospective Register of Systematic Review (PROSPERO; ID CRD42022328706). This report was prepared using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement and the PRISMA for Searching (PRISMA-S) [11,12].

2.1. Search strategy

A systematic search of the literature was performed (July 10, 2023) in PubMed (including MEDLINE), Embase (Embase.com), Web of Science Core Collection and

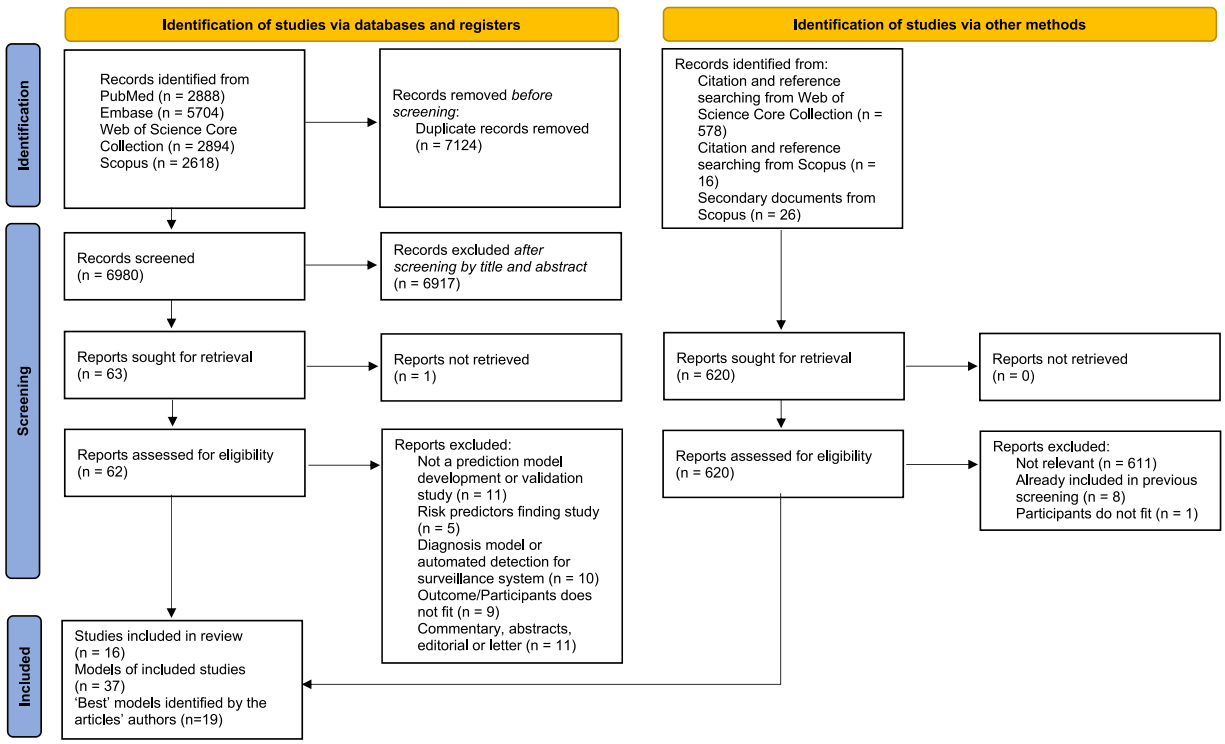

Fig. 1. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1. Characteristics of central line-associated bloodstream infections risk prediction models

Characteristic	Overall (*N* = 19)	Development and internal validation (*N* = 16)	External validation (*N* = 3)
Study type			
Development and internal validation	9 (47%)	9 (56%)	
Development only	7 (37%)	7 (44%)	
External validation only	3 (16%)		3 (100%)
Data type			
Nested case-control	1 (5%)	1 (6%)	
Non-nested case-control	2 (11%)	1 (6%)	1 (33%)
Prospective cohort	1 (5%)		1 (33%)
Retrospective cohort	15 (79%)	14 (88%)	1 (33%)
Country			
Brazil	1 (5%)		1 (33%)
China (mainland)	1 (5%)	1 (6%)	
Denmark	1 (5%)	1 (6%)	
Germany	1 (5%)		1 (33%)
Japan	1 (5%)		1 (33%)
Saudi Arabia	1 (5%)	1 (6%)	
Taiwan	1 (5%)	1 (6%)	
United States	12 (63%)	12 (75%)	
Center			
Monocenter	15 (79%)	12 (75%)	3 (100%)
Multicenter	4 (21%)	4 (25%)	
Setting			
Hospitalwide	8 (42%)	7 (44%)	1 (33%)
Intensive care unit (ICU)	6 (32%)	5 (31%)	1 (33%)
Other specific hospital units/wards	5 (26%)	4 (25%)	1 (33%)
Age			
Adults	5 (26%)	3 (19%)	2 (67%)
All patients	7 (37%)	7 (44%)	
Children	7 (37%)	6 (38%)	1 (33%)
Catheter type			
All central lines	16 (84%)	14 (88%)	2 (67%)
Peripherally inserted central venous catheters (PICC) only	2 (11%)	1 (6%)	1 (33%)
Permanent implantable venous ports (Port-A) catheter only	1 (5%)	1 (6%)	
Outcome event			
Catheter-dependent infection[a]	1 (5%)	1 (6%)	
Catheter-related bloodstream infection (CRBSI)[b]	3 (16%)	1 (6%)	2 (67%)
Central line-associated bloodstream infection (CLA-BSI)	12 (63%)	12 (75%)	
PICC-associated bloodstream infection (PBSI)	2 (11%)	1 (6%)	1 (33%)
Port-A-associated bloodstream infection (PABSI)-free survival[c]	1 (5%)	1 (6%)	
Type of endpoints			
Binary endpoint	15 (79%)	13 (81%)	2 (67%)
Time-to-event endpoint	4 (21%)	3 (19%)	1 (33%)
Dynamic nature			
Static	16 (84%)	13 (81%)	3 (100%)

(*Continued*)

Table 1. Continued

Characteristic	Overall ($N = 19$)	Development and internal validation ($N = 16$)	External validation ($N = 3$)
Dynamic	1 (5%)	1 (6%)	
Unclear	2 (11%)	2 (13%)	
Sample size			
Median sample size (interquartile range)	n.c.	9 862 (210, 22,414)	267 (185, 863)
Median events (interquartile range)	n.c.	123 (46, 241)	66 (59, 78)

Abbreviation: n.c., not calculated.

[a] Catheter-dependent infection is uniquely defined by Lücking et al. as bacteremias due to an external pathogen from the skin or surroundings assumed to have entered the bloodstream via the CVC. However, it is unclear when and how the blood cultures are taken, thus, contaminants might also be included and introduce bias.

[b] Catheter-related bloodstream infection (CRBSI) requires a definitive pathological diagnosis through quantitative culture of the catheter tip or the growth time differences between catheter and peripheral blood culture specimens, which is a narrower scope definition than central line-associated bloodstream infections (CLA-BSI). The definition of CLA-BSI is primarily used for surveillance purposes, assuming that the presence of a bloodstream infection in patients without any other identified source can be attributed to the central line, thus might identify a larger number of cases that may not truly be related to the line.

[c] Port-A-associated bloodstream infection (PABSI)-free survival is defined by Chen et al. as the duration between Port-A implantation and development of the first PABSI in the case group or last follow-up date in the control group.

Scopus. We searched studies developing or externally validating risk prediction models to predict the occurrence of CLA-BSI in hospitalized patients with a catheter. Therefore, an AND-combination was made of the following three concepts: ''CLA-BSI'', ''prediction models'', and ''central venous catheter''. The search strategies were peer reviewed by an experienced information specialist (K.T.) prior to execution. The full search strings were reported in Supplementary File 1 and preserved on searchRxiv https://searchrxiv.org/. Additionally, we did forward and backward citation searching of the included articles via Web of Science Core Collection and Scopus to identify additional relevant studies. (July 16, 2023).

2.2. Inclusion and exclusion criteria

Studies that developed or validated a multivariable model to predict CLA-BSI risks for inpatients with central lines were included. The National Healthcare Surveillance Network (NHSN) has established the definition of CLA-BSI, in collaboration with the Centers for Disease Control (CDC) [1]. Although this gold standard has been adopted and adapted worldwide by central reporting agencies such as the European CDC [13], there is variability of adopted CLA-BSI definitions across various countries. Furthermore, in practice, CLA-BSI identification and classification involve subjective judgment by infection preventionists [14]. We included studies using (local variations of) the NHSN/CDC criteria. We also included studies using a rigorous clinical definition of catheter-related bloodstream infection (CRBSI), which requires definitive diagnosis of same pathogen in blood and catheter culture, as outcome.

We included studies based on any design (e.g., randomized-controlled trials, retrospective or prospective cohort studies). There were no language or other restrictions on any search of the databases. Diagnostic models that detect the presence of CLA-BSI for surveillance were excluded. Predictor finding studies that focus on the association between potential risk factors and CLA-BSI, were also excluded. Studies whose populations are outpatients (e.g., patients who received home parental nutrition) or inpatients without central lines placement were not included. Other exclusion criteria for research articles include the following: qualitative study reports, lack of access to full text, and articles that do not report original research such as reviews, editorials, and conference abstracts.

After removing duplicates (using EndNote 20 (Clarivate), Rayyan, and manual checking [15,16]), titles and abstracts were initially screened for exclusion by at least two authors. As inter-reviewer agreement was considered sufficient (see Supplementary File 2), the remaining title-abstract screening was done by one author (S.G.) and irrelevant articles were excluded. Then, full texts of the potentially relevant articles were screened independently by two authors (E.A. and S.G.). Discrepancies were resolved through discussion with a third author (B.V.C. or L.W.).

2.3. Data extraction and quality assessment

Data extraction from included studies was carried out independently by two investigators (E.A. and S.G.) using standard data extraction forms (Supplementary File 3) based on the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modeling Studies (CHARMS) [17]. Where one article investigated more than one model (e.g., same model development strategy however, different algorithms), data extraction focused on the model identified as the best model by the article's authors. If multiple papers discussed the

Box 1 Advantages and disadvantages of different algorithms applied by the developed prediction models.

Among all the 16 developed models, nine models applied logistic regression, three were Cox proportional hazards model, two were XGBoost, one was random forest, and one on naïve Bayes.

Regression models still dominate the landscape of risk prediction models for central line-associated bloodstream infections (CLA-BSI). Among traditional regression models, logistic models are commonly used to handle binary or categorical outcomes. They are easy to implement but require human input to address nonlinear problems and interactions between variables, which may occur in real-word scenarios.

In comparison with logistic regression models, survival models such as Cox proportional hazards models focus more on the time to an event which contains more clinical information than simply whether the event occurred or not. They can handle the censoring when subjects did not experience the event before the end of study. Nevertheless, the assumptions of proportional hazards (that predictor effects are constant over time) may not always hold, which may lead to biased estimates. Time-dependent coefficients can be used when the proportional hazard assumption is not reasonable.

In contrast to the aforementioned regression models, machine learning algorithms such as tree-based algorithms automatically capture complex nonlinear and nonadditive relationships. However, they are prone to overfitting if not properly regularized or tuned. A disadvantage of machine learning algorithms is that they are 'black box': there is no simple equation that reflects how each predictor contributes to the prediction. Time-of -to-event extensions for machine learning algorithms exist but are not often used.

Naïve Bayes has the advantages of computationally efficiency and are straightforward to implement. However, the assumption of conditional independence between features may not hold in real-world settings. Moreover, naïve Bayes can encounter the ''zero-frequency problem'' when there are unseen combinations of feature values in the test data that was not present in the training data during prediction.

same model (e.g., development and one or more external validation studies), each study was assessed separately. The following data were extracted: authors, year of publication, study design, data source, definition of CLA-BSI, modeling algorithm, validation method, missing data approach, type of variable selection, number of events and event rate, model performance, and evaluation metrics.

Two independent investigators (E.A. and S.G.) conducted the risk of bias and applicability assessments for included studies, utilizing the Prediction model Risk Of Bias ASsessment Tool (PROBAST) [18,19]. Overall and domain-specific risk of bias and concern regarding applicability were classified as low, high, or unclear. Disagreements were resolved by a third and fourth adjudicator (B.V.C. and L.W.), who informed the final decision.

2.4. Data synthesis

No meta-analysis was performed given the heterogeneity across the included studies [20]. Results were presented narratively, using descriptive statistics and graphical plots to summarize the characteristics of the included studies.

3. Results

Sixteen studies were included after evaluating 6,980 papers (Fig. 1). The PRISMA flow diagram [21] represented the selection process and reasons for article exclusion. These 16 studies developed and/or validated 37 prediction models. Five studies developed more than one model for CLA-BSI, in which case we restricted our focus to the model identified as best by the authors of the paper (Supplementary File 4). Hence, we included sixteen newly developed models for CLA-BSI. Three externally validated models were also included, of which one was initially developed for CLA-BSI in another country and two were initially developed for another outcome (Table 1). Retrospective cohort data were most commonly used (79%). All studies used single country data and 12 (63%) models were developed or validated based on data from the United States. Four (21%) models used multicenter data. Sixteen (84%) models were intended for all types of central line catheters, 2 (11%) for peripherally inserted central venous catheters, and 1 (5%) for permanent implantable venous ports catheters. Twelve (63%) models were intended to predict the risk of CLA-BSI, while 3 (16%) models were developed and validated to predict CRBSI, which requires a definitive diagnosis of same pathogen in cultures of blood and catheter to identify the catheter as the source of the infection. In addition, one model predicted an ad hoc outcome definition of catheter-dependent infection. Model predictors were described in detail in Supplementary File 5 and also summarized by categories such as demographics, medications, vital signs, and laboratory data in Table 2. Age, history of CLA-BSI, total parenteral nutrition (TPN), and neutrophils were the most frequently included predictors in the 19 models.

Table 2. The number and type of predictor variables in the reviewed, developed, and externally validated models

	Development								
Predictors	**Bearman et al. 2010 [22]**	**Wylie et al. 2010 [23]**	**Chen et al. 2012 [24]**	**Lücking et al. 2013 [25]**	**Herc et al. 2017 [26]**	**Beeler et al. 2018 [27]**	**Parreco et al. 2018 [28]**	**Waterhouse et al. 2018 [29][a]**	**Waterhouse et al. 2018 [29]**
Demographics	0	0	0	0	0	2	2	0	0
Age[e]						×			
Admission and patient stay data	0	1	0	0	0	1	0	0	0
Inpatient days before CVL placement						×			
Administration related to catheters	2	0	2	0	2	3	0	1	0
CVL days	×							×	
Medical condition	1	1	0	0	3	1	27	0	0
History of CLA-BSI					×	×			
Medication/ Treatment	1	0	3	1	1	1	1	0	0
TPN			×		×				
Laboratory test	0	0	1	4	0	0	0	0	0
Neutrophils			×	×					
Vital signs	0	0	0	0	0	0	0	0	0
Temperature									
Scores[f]	0	1	0	0	0	0	6	1	1
RACHS								×	×
Others	0	2	0	0	0	0	0	0	2
Total number[g]	4	5	6	5	6	8	36	2	3

Abbreviations: CVL, central venous line; CLA-BSI, central line-associated bloodstream infections; TPN, total parental nutrition; RACHS: Risk Adjustment for Congenital Heart Surgery.

[a] Total number indicates the number of models that included the corresponding predictors.

[b] The three models developed by Waterhouse et al. used different predictive factors in final models.

[c] Baeissa et al. did not report the number and type of predictors in their final model.

[d] Rahmani et al. developed two eXtreme Gradient Boosting (XGBoost) models using all features and top 13 selected important features, respectively.

[e] Age is the most frequently used predictors among all demographic variables; similar rule applies to the following listed predictors.

[f] Scores include different severity of illness scores such as Oxford Acute Severity of Illness Score (OASIS) and Pediatric Risk of Mortality (PRISM) score.

[g] Total number indicates the overall number of predictors used in each model.

3.1. Development and internal validation studies

There were sixteen developed models and the median sample size was 9, 862 at model development, with a median number of 123 events. Nine models were based on logistic regression, three on Cox regression, two on eXtreme Gradient Boosting, one on random forest, and one on naïve Bayes (Box 1). Thirteen (81%) models predicted a binary outcome and 3 (19%) predicted a time-to-event outcome (Table 1). Ten (63%) models predicted the outcome at any time during admission, five did not specify their prediction time horizon, and one predicted at catheter dwell times of 6–40 days (Table 3). Only 1 (6%) model was developed in a dynamic way, using the daily assessment of positive blood cultures as the proxy of outcome [31]. For 11 (69%) models, neither number/percentage of missing values nor missing data handling method were reported. Three (19%) models were based on complete case analysis, one on multiple imputations, and one on the missing indicator method. Nine (56%) models were internally validated to account for optimism (six random split, one temporal split, one bootstrapping, one cross-validation). The median number of events per variable (EPV) was 3 (ranged from 1 to 19). Fourteen (88%) models had less than 10 EPV, 1 had 19 EPV, and 1 did not report the number of events nor the number of candidate predictors. For 3 (19%) models,

Development							External validation			
Waterhouse et al. 2018 [29]	Baeissa et al. 2019 [30][b]	Bonello et al. 2022 [31]	Hooshmand et al. 2022 [32]	Rahmani et al. 2022 [33][c]	Rahmani et al. 2022 [33]	Wang et al. 2023 [34]	Vilela et al. 2007 [35]	Schalk et al. 2015 [36]	Sakai et al. 2021 [37]	Total number[d]
0		1	1	4	2	0	0	0	0	
		×		×	×					4
0		1	5	1	0	0	0	0	0	
			×	×						3
0		0	2	0	0	2	0	0	2	
						×				3
0		0	1	19	6	0	0	0	3	
			×	×	×				×	6
0		3	1	0	0	0	0	0	1	
		×							×	4
0		3	0	3	3	0	0	0	0	
				×	×					4
0		2	0	1	1	0	0	0	0	
		×		×	×					3
1		0	0	0	0	1	1	1	0	
×										3
1		0	1	0	0	0	0	0	0	
2		10	11	28	12	3	1	1	6	

apparent calibration assessment was performed. Internally validated calibration assessment was not provided for any model. The apparent C-index varied from 0.70 to 0.88, the internally validated C-index from 0.67 to 0.82 (Table 4).

3.2. External validation studies

There were three external validation models and two out of them used scores developed for other outcomes to validate their capability of predicting CRBSI. One study validated the Michigan PICC catheter-associated bloodstream infection (MPC) score, one of the 16 developed models included above, on the data from a Japanese hospital. Only one study assessed calibration, and the reported C-index of these three models varied from 0.53 to 0.77 (Table 5).

3.3. Risk of bias

In our review, a PROBAST risk of bias tool assessed the risk of biased (usually optimistic) predictive performance estimates [18]. All 16 developed models were at high risk of bias (Fig. 2). For the participants domain, 6 (38%) models were at high risk of bias, and 3 (19%) at unclear risk of bias. High risk of bias was due to inappropriate exclusion of patients based on data after the intended moment of prediction or lack of adjustment of sampling fractions following a case-control design. Unclear risk of bias was due to unclear

Table 3. Characteristics of model development studies

Author and yr	Model	Prediction time horizon	Missing handling	Validation method	Predictor selection		Development		Internal validation
					Before modeling	During modeling	Events/*N*	EPV	Events/*N*
Bearman et al. 2010 [22]	CPH	NR	NR	RS (50:50)	Univariate selection	Stepwise selection	123/15,100	6[a]	121/15,097
Wylie et al. 2010 [23]	LR	Any time during admission	NR	RS (67:33)	[b]Other	NR	135/406	4	68/203
Chen et al. 2012 [24]	CPH	NR	NR		All candidate predictors	Stepwise selection	58/232	3	
Lücking et al. 2013 [25]	Logistic Bayes	NR	NR		Univariate selection	Other[c]	34/172	3	
Herc et al. 2017 [26]	CPH	PICC dwell times of 6–40 days	MI	Bootstrap (200 reps)	Univariate selection	Stepwise selection	249/23,088	3	249/23,088
Beeler et al. 2018 [27]	RF	NR	NR	Temporal[d]	All candidate predictors	Other[e]	387/56,174	19	NR/49,669
Parreco et al. 2018 [28]	LR	Any time during admission	NR	10-fold CV	All candidate predictors	All predictors forced in model	333/22,190	3	333/22,190
Waterhous et al. 2018 [29][f]	LR	Any time during admission	NR		Univariate selection	NR	15/66	1	
Waterhouse et al. 2018 [29]	LR	Any time during admission	NR		Univariate selection	NR	15/66	1	
Waterhouse et al. 2018 [29]	LR	Any time during admission	NR		Univariate selection	NR	15/66	1	
Baeissa et al. 2019 [30]	naïve Bayes	NR	NR		NR	NR	NR/28,972	NR	
Bonello et al. 2022 [31]	LR	Any time during admission	MIM	RS (60:40)	Univariate selection	Stepwise selection	240/62,421	3	159/41,614
Hooshmand et al. 2022 [32]	LR	Any time during admission	NR	RS + CV[g]	All candidate predictors	Other[h]	77/4,623	2	19/1,156
Rahmani et al. 2022 [33][i]	XGBoost	Any time during admission	CC	RS (80:20)	All candidate predictors	All predictors forced in model	241/22,095	6	60/5,464
Rahmani et al. 2022 [33]	XGBoost	Any time during admission	CC	RS (80:20)	Other	All predictors forced in model	241/22,095	6	60/5,464
Wang et al. 2023 [34]	LR	Any time during admission	CC		All candidate predictors	Lasso	69/222	2	

Abbreviations: CPH, cox proportional hazards; LR, logistic regression; RF, random forest; XGBoost, extreme gradient boosting; NR, not recorded; MI, multiple imputation; MIM, missing indicator method; CC, complete-case; RS, random split; CV, cross-validation; Events/*N*, number of events/participants; EPV, event per variable.

[a] The event per variable (EPV) number is an approximated value as the author did not specify clearly the number of candidate predictors used for model development.

[b] Wylie et al. chose a subset of risk factors that were present and known at the time of line placement without providing any other details regarding the selection criteria.

[c] Lücking et al. selected factors with odds ratio >2 & lower confidence limit >1 or very close to one for inclusion.

[d] Beeler et al. performed the internal validation using 20% randomly split data besides a temporal validation. We considered the temporal validation as a higher priority for the reporting.

[e] Beeler et al. selected based on the random forest algorithm's variable importance rankings following the Gini Impurity criterion, with 15/20 baseline risk factors accounted for the most significant effect on central line-associated bloodstream infections (CLA-BSI) prediction.

[f] The three models developed by Waterhouse et al. used different predictive factors in final models.

[g] Hooshmand et al. built the logistic regression model by splitting the entire dataset into training (80%) and validation (20%) set, and assessed the logistic regression's classification accuracy for the validation set by performing a five-fold cross-validation and aggregating the results from different outcomes.

[h] Hooshmand et al. fitted the logistic regression model on the training set as well as the Cox proportional hazards model on the entire dataset. Eventually logistic regression model was chosen as the 'best' final model. Variables with nonsignificant *P* values (<0.05) in both methods were omitted.

[i] Rahmani et al. developed two eXtreme Gradient Boosting (XGBoost) models using all features and top 13 selected important features, respectively.

Table 4. Performance measurements of the development and internal validation models

		Apparent performance			Internal validation		
Author and yr	**Model**	**Calibration**	**[a]Discrimination**	**Classification**	**Calibration**	**Discrimination**	**Classification**
Bearman et al. 2010 [22]	CPH	NR	NR	Sensitivity: 75%; specificity: 89%; PPV: 0.05	NR	NR	Sensitivity: 69%; specificity: 88%; PPV: 0.046
Wylie et al. 2010 [23]	LR	NR	0.79 (0.75, 0.84)	sensitivity: 74%; specificity: 71%; NPV: 84%; PPV: 56%	NR	0.72 (0.64, 0.79)	sensitivity: 62%; specificity: 74%; NPV: 79%; PPV: 54%
Chen et al. 2012 [24]	CPH	NR	0.844 (0.78, 0.91)	sensitivity: 89%; specificity: 64%			
Lücking et al. 2013 [25]	Logistic Bayes	E/O table	NR	NR			
Herc et al. 2017 [26]	CPH	Calibration table	[b]ranged from 0.70 (0.64–0.76) to 0.80 (0.76–0.84)	NR	NR	ranged from 0.67 to 0.77	NR
Beeler et al. 2018 [27]	RF	NR	NR	NR	NR	0.82	alert rate: 33%
Parreco et al. 2018 [28]	LR	NR	NR	NR	NR	0.72 (0.67, 0.77)	sensitivity: 0%; specificity: 100%; PPV: 0.0%; NPV: 98.6%; Accuracy: 98.6%; Precision: 0.0%
[c]Waterhouse et al. 2018 [29]	LR	NR	0.884	NR			
Waterhouse et al. 2018 [29]	LR	NR	0.790	NR			
Waterhouse et al. 2018 [29]	LR	NR	0.798	NR			
Baeissa et al. 2019 [30]	Naïve Bayes	NR	NR	Sensitivity: 98%; Specificity: 99%; Accuracy: 98%; Classification error rate: 0.95; Kappa: 0.63			
Bonello et al. 2022 [31]	LR	NR	NR	NR	NR	0.82	Sensitivity: 25%; specificity: 99.9%; PPV: 48.2%; NPV: 99.7%
Hooshmand et al. 2022 [32]	LR	NR	NR	NR	NR	0.71	Accuracy: 75%; error rate: 25%; sensitivity: 67%; specificity: 75%; F1 score: 0.08
[d]Rahmani et al. 2022 [33]	XGBoost	NR	NR	NR	NR	0.76 (0.70, 0.83)	Sensitivity 0.8; specificity 0.5; LR+ 1.73; LR- 0.37; DOR: 4.67
Rahmani et al. 2022 [33]	XGBoost	NR	NR	NR	NR	0.75 (0.67, 0.82)	Sensitivity 0.8; specificity 0.5; LR+ 1.54; LR- 0.42; DOR: 3.72
Wang et al. 2023 [34]	LR	Calibration plot	0.84 (0.78, 0.90)				

Abbreviations: CPH, cox proportional hazards; LR, logistic regression; RF, random forest; XGBoost, extreme gradient boosting; NR, not recorded; NPV, negative predictive value; PPV, positive predictive value; E/O, expected/observed number of events; LR, likelihood ratio; DOR, diagnostic odds ratio.

[a] Performance measurements used for discrimination were C-index with 95% confidence interval, if reported.

[b] Herc et al. assessed the predictive performance using time-dependent area under the curve (AUC) values at clinically relevant time points over a range of 6–40 device days.

[c] The three models developed by Waterhouse et al. used different predictive factors in final models.

[d] Rahmani et al. developed two eXtreme Gradient Boosting (XGBoost) models using all features and top 13 selected important features, respectively.

reporting of the eligibility criteria. For the predictor domain, 8 (50%) models were at high risk of bias, and 3 (19%) at unclear risk of bias. High risk of bias resulted from the use of predictors that are unavailable at the time the model is intended to be used in practice. Unclear risk of bias resulted from unknown predictor measurement timing and unspecified definition of predictors. For the outcome domain, 11 (69%) models were at high risk of bias due to the use of a proxy outcome (e.g., International Classification of Disease coding which requires high compliance to coding [38]) or an inappropriate self-defined outcome (i.e., central venous catheter-dependent bloodstream infection which may wrongly include contaminants). One (6%) model had an unclear risk of bias due to an unclear outcome definition. For the analysis domain, all models were at high risk of bias. Typical reasons were insufficient sample size, ignoring missing data, inappropriate model evaluation (e.g., no internal validation at all, or ignoring calibration performance), and problematic variable selection methods (e.g., preselect predictors based on univariate analysis results). Details can be found in Supplementary File 6.

All three external validation studies suffered from high risk of bias. One study used a case-control design but did not appropriately adjust for the distorted incidence of the event [35]. In addition, its controls were matched on future data. One study did not discuss missing values and how they were handled [36]. In one study, all events (PICC-associated bloodstream infection) were assessed by the author using the linked microbiology data, unblinded to the MPC score [37]. This may influence determination of the outcome and lead to a potential risk of bias. All three studies had low sample size, resulting in a high risk of bias for the analysis domain. Two studies did not assess calibration.

3.4. Applicability

Applicability assesses the transferability of a study finding to a specific population or setting [18]. In our case, we assessed the models' applicability to predict CLA-BSI risk for inpatients at any time during their hospitalization. Among 16 developed models, 15 (94%) had a high concern and one an unclear concern for applicability. Seven (44%) models had a high concern in the participants' domain due to its limited eligible population, for example, solid cancer patients who were seriously ill. One (6%) model had unclear applicability due to the poor reporting of eligibility criteria. 8 (50%) models had a high concern for applicability in the predictors domain as the predictor values are only available when a CLA-BSI has occurred, which made these models unusable. 3 (19%) models had unclear applicability due to unclear predictor definitions and timing. For the outcome domain, 8 (50%) models had a high concern and 2 (13%) models an unclear concern for applicability. Problematic definition of outcome which did not follow the CDC/NHSN criteria is the most common reason of a high concern rating.

All three external validation studies had a high concern for applicability. One study had a high concern for applicability for the participants' domain as it used a case-control design with controls matched on future data (LOS). One study had a high concern for applicability as it predicted CRBSI at catheter removal. As an effective way of preventing CLA-BSI, a catheter may be removed at the suspicion of infection. Thus, predicting its occurrence when CLA-BSI is suspected does not seem to provide enough time to react and prevent it. The validation of the MPC score had a high concern for applicability for the predictors' domain due to the potential temporal leaks. Components of the score like TPN through the PICC might leak future data as TPN receipts may change after catheter placement. The authors did not specify when they were measuring it. It can be after patient discharge or before catheter insertion.

4. Discussion

In this systematic review of risk prediction models related to CLA-BSI, we identified 13 model development studies (16 models) and three external validation studies (three models). All models were labeled as having a high risk of bias. Common reasons were inappropriate proxy outcome,

Table 5. Externally validated models

Study	Model validated	Outcome	[a]Sample size	Calibration	Discrimination	Classification
Vilela et al. 2007 [35]	PRISMA score	CRBSI	51/102	NR	C-index 0.53 (95% CI: 0.43–0.63)	
Schalk et al. 2015 [36]	mIPS	CRBSI	66/267	NR	C-index 0.77 (95% CI: 0.71–0.83)	Sensitivity: 84.9%; Specificity: 60.7%; NPV: 92.4%; PPV: 41.5%
Sakai et al. 2021 [37]	MPC score	PBSI	89/1,459	CS 1.16 (95% CI: 1.02–1.32)	C-index 0.61 (95% CI: 0.54–0.67)	

Abbreviations: MPC score, michigan PICC catheter-associated bloodstream infection score; mIPS, modified infection probability score; PRISMA score, pediatric risk of mortality score; PBSI, PICC-associated bloodstream infection; CRBSI, catheter-related bloodstream infection; CS, calibration slope; NR, not recorded; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

[a] Sample size is presented in a way of the number of outcome/participants.

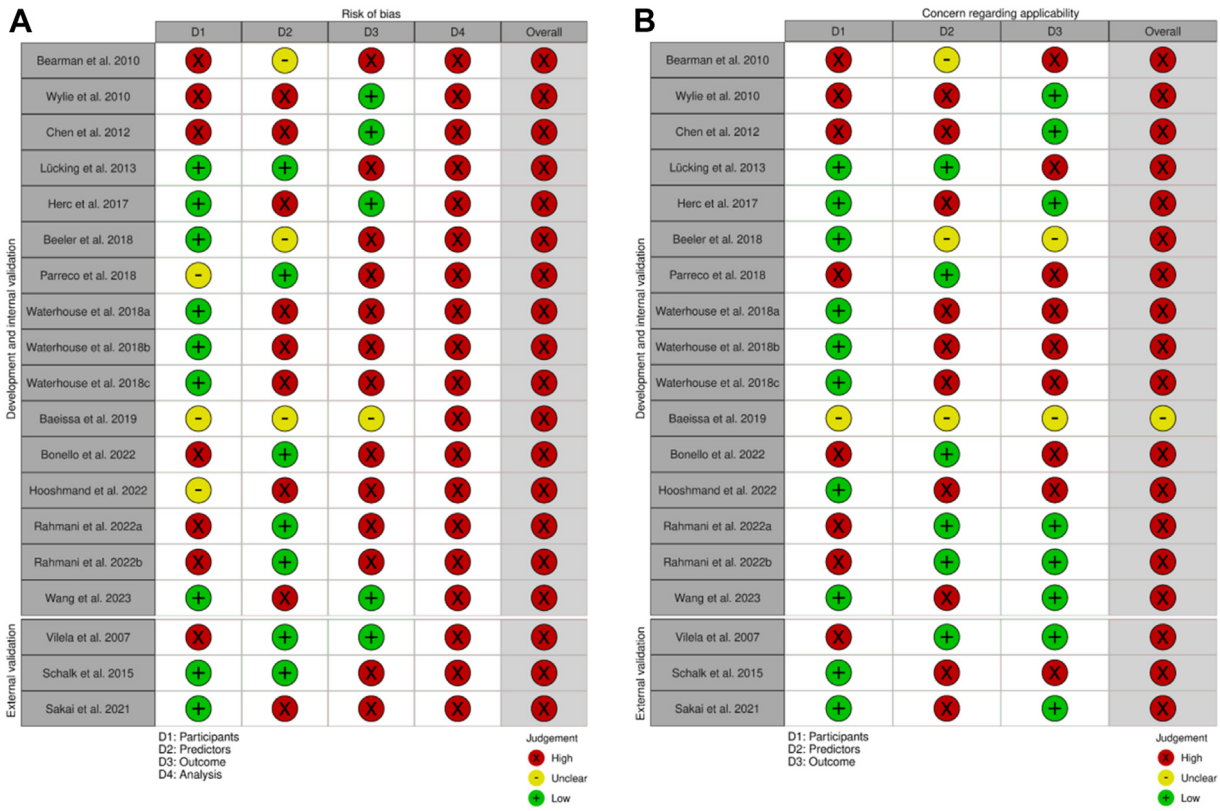

Fig. 2. Traffic light plots of (A) risk of bias and (B) concern regarding applicability assessments using the Prediction model Risk Of Bias ASsessment Tool (PROBAST) in each domain and overall for all models. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

predictors that are unavailable at prediction time, inadequate sample size, negligence of missing data, lack of model validation, absence of calibration assessment, and incomplete reporting. In addition, all models suffered from high or unclear concerns for applicability. Common reasons included inappropriate eligibility criteria, unavailable predictor measures at the moment the prediction is needed in practice, and an inappropriate outcome identification method. For one model, applicability was unclear due to incomplete reporting regarding participants, predictors, and outcome.

Similar problems as mentioned above have also been reported in reviews of prediction models for other clinical outcomes [39–41]. Despite that many reviews recommend to follow the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines, the adherence to the reporting guideline remains poor [42]. Also, with the increasing popularity of machine learning models, the TRIPOD-AI extension is underway [43]. This will include much needed attention for machine learning specific topics such as hyperparameter optimization.

4.1. Implications

Risk prediction models aim to support clinical decision-making for patients; therefore, it is essential that the intended clinical setting and target population for C LA-BSI prediction are clearly defined. This helps to avoid the inclusion of predictors that are not yet available when a prediction is needed. The inclusion of such predictors makes prediction models useless.

We repeat the recommendation to follow the TRIPOD [44] reporting guideline. In addition to the reporting quality, the methodological conduct of prediction model studies raised serious concerns. Inadequate sample size often results in models with suboptimal performance in new individuals [45]. It is recommended to calculate the sample size required for model development, even though further research is needed for machine learning models. In addition, sample size methods for external validation studies have recently been proposed, which are important to guarantee sufficiently precise performance estimates [46,47]. Although it is a difficult problem, missing data need to be handled carefully. Simply excluding patients with missing data to perform a complete case analysis results in a smaller sample size and often leads to bias. In particular, models that give real-time predictions for hospitalized patients using electronic health record's information should also be able to deal with missing data in real time. Finally, 7 of 16 developed models were not even internally validated and hence, did not report any optimism-corrected performance estimate. For the nine models with internal validation, none assessed calibration using for example calibration plots [48]. It is strongly suggested to apply appropriate internal validation strategies such as bootstrapping or cross-validation, and assess calibration and discrimination [49].

Aside from issues with bias and applicability, this systematic review identified only one dynamic model. Dynamic models have the advantage that they can update predictions over time. This allows to dynamically adapt patient management if needed, allowing clinical staff to take timely action to prevent CLA-BSI occurrence. Our review identifies a clear potential for developing dynamic CLA-BSI prediction models.

4.2. Limitations

This review has several limitations. Our data extraction form was mainly designed based on the items in CHARMS checklist and signaling questions from PROBAST. Though most items were applicable for both regression-based and machine learning models, some differences such as hyperparameter tuning complicated harmonization of data extraction. There was also great variability in target populations, predictors, and outcome definitions, which made summaries through meta-analysis impossible.

5. Conclusion

This systematic review critically appraised 19 models which were developed and/or validated to predict CLA-BSI for hospitalized patients. However, all models were at high risk of bias and had high or unclear concerns regarding applicability. We did not identify any model as potentially clinically useable. Moreover, there was considerable variability in target populations, predictors, catheter types, and outcome definitions. Thus, there is a need for a well-developed and applicable model. For example, a dynamic model that enables real-time CLA-BSI risk prediction based on electronic health record data. Further, there is an urgent need to improve the methodological conduct of risk prediction studies for any clinical outcome of interest.

CRediT authorship contribution statement

S. Gao: Conceptualization, Formal analysis, Investigation, Data curation, Writing − original draft. **E. Albu:** Investigation, Data curation, Writing − review & editing. **K. Tuand:** Investigation, Writing − review & editing. **V. Cossey:** Writing − review & editing. **F.E. Rademakers:** Writing − review & editing. **B. Van Calster:** Conceptualization, Investigation, Writing − review & editing. **L. Wynants:** Conceptualization, Investigation, Writing − review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Internal Funds KU Leuven [grant C24M/20/064]. The funding sources had no role in the conception, design, data collection, analysis, or reporting of this study.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2023.07.019.

References

[1] Centers for Disease Control and Prevention. Bloodstream infection event (central line-associated bloodstream infection and non-central line associated bloodstream infection) [internet] 2022. Available at https://www.cdc.gov/nhsn/pdfs/pscmanual/4psc_clabscurrent.pdf. Accessed January 25, 2023.

[2] Surveillance of bloodstream infections in belgian hospitals: report 2021 data up to and including 2020 [internet]. Available at https://www.sciensano.be/sites/default/files/bsi_report_2021.pdf. Accessed February 15, 2023.

[3] Klevens RM, Edwards JR, Richards CL Jr, Horan TC, Gaynes RP, Pollock DA, et al. Estimating health care-associated infections and deaths in U.S. hospitals, 2002. Public Health Rep 2007;122:160−6.

[4] Stewart S, Robertson C, Pan J, Kennedy S, Haahr L, Manoukian S, et al. Impact of healthcare-associated infection on length of stay. J Hosp Infect 2021;114:23−31.

[5] Zimlichman E, Henderson D, Tamir O, Franz C, Song P, Yamin CK, et al. Health care-associated infections: a meta-analysis of costs and financial impact on the US health care system. JAMA Intern Med 2013;173:2039−46.

[6] O'Grady NP, Alexander M, Burns LA, Dellinger EP, Garland J, Heard SO, et al. Guidelines for the prevention of intravascular catheter-related infections. Am J Infect Control 2011;39:S1−34.

[7] Goudie A, Dynan L, Brady PW, Rettiganti M. Attributable cost and length of stay for central line-associated bloodstream infections. Pediatrics 2014;133:e1525−32.

[8] Leistner R, Hirsemann E, Bloch A, Gastmeier P, Geffers C. Costs and prolonged length of stay of central venous catheter-associated bloodstream infections (CVC BSI): a matched prospective cohort study. Infection 2014;42:31−6.

[9] Umscheid CA, Mitchell MD, Doshi JA, Agarwal R, Williams K, Brennan PJ. Estimating the proportion of healthcare-associated infections that are reasonably preventable and the related mortality and costs. Infect Control Hosp Epidemiol 2011;32:101−14.

[10] Saegeman V, Cossey V, Schuermans A. Reducing central-line-associated bloodstream infections by half: it is possible. J Hosp Infect 2022;128:89−91.

[11] Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6:e1000097.

[12] Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev 2021;10:39.

[13] European Centre for Disease Prevention and Control. Surveillance of healthcare-associated infections and prevention indicators in European intensive care units [internet] 2017. Available at https://www.ecdc.europa.eu/sites/default/files/documents/HAI-Net-ICU-protocol-v2.2_0.pdf. Accessed January 9, 2023.

[14] Lin MY, Hota B, Khan YM, Woeltje KF, Borlawsky TB, Doherty JA, et al. Quality of traditional surveillance for public reporting of nosocomial bloodstream infection rates. JAMA 2010;304:2035−41.
[15] Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. J Med Libr Assoc 2016;104:240−3.
[16] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev 2016;5:210.
[17] Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med 2014;11:e1001744.
[18] Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170:51−8.
[19] McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): an R package and Shiny web app for visualizing risk-of-bias assessments. Res Synth Methods 2021;12:55−61.
[20] McKenzie JE, Brennan SE. Synthesizing and presenting findings using other methods. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane Handbook for Systematic Reviews of Interventions. 2nd ed. Chichester, UK: John Wiley & Sons; 2019:321−47.
[21] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.
[22] Bearman GML, Oppenheim MI, Mendonca EA, Hupert N, Behta M, Christos PJ, et al. A clinical predictive model for catheter related bloodstream infections from the electronic medical record. Open Epidemiol J 2010;3:24−8.
[23] Wylie MC, Graham DA, Potter-Bynoe G, Kleinman ME, Randolph AG, Costello JM, et al. Risk factors for central line-associated bloodstream infection in pediatric intensive care units. Infect Control Hosp Epidemiol 2010;31:1049−56.
[24] Chen IC, Hsu C, Chen YC, Chien SF, Kao HF, Chang SY, et al. Predictors of bloodstream infection associated with permanently implantable venous port in solid cancer patients. Ann Oncol 2013;24:463−8.
[25] Lücking V, Rosthøj S. Prediction of bacteremia in children with febrile episodes during chemotherapy for acute lymphoblastic leukemia. Pediatr Hematol Oncol 2013;30:131−40.
[26] Herc E, Patel P, Washer LL, Conlon A, Flanders SA, Chopra V. A model to predict central-line-associated bloodstream infection among patients with peripherally inserted central catheters: the MPC score. Infect Control Hosp Epidemiol 2017;38:1155−66.
[27] Beeler C, Dbeibo L, Kelley K, Thatcher L, Webb D, Bah A, et al. Assessing patient risk of central line-associated bacteremia via machine learning. Am J Infect Control 2018;46:986−91.
[28] Parreco JP, Hidalgo AE, Badilla AD, Ilyas O, Rattan R. Predicting central line-associated bloodstream infections and mortality using supervised machine learning. J Crit Care 2018;45:156−62.
[29] Waterhouse SG, Vergales JE, Conaway MR, Lee L. Predictive factors for central line-associated bloodstream infections in pediatric cardiac Surgery patients with chylothorax. Pediatr Crit Care Med 2018;19:810−5.
[30] Baeissa O, Noaman AY, Ragab AHM, Hagag A. Reduce prediction time for HAI-central line blood stream infection using big data mining model. Int J Comput Sci Netw Secur 2019;19:19−23.
[31] Bonello K, Emani S, Sorensen A, Shaw L, Godsay M, Delgado M, et al. Prediction of impending central-line-associated bloodstream infections in hospitalized cardiac patients: development and testing of a machine-learning model. J Hosp Infect 2022;127:44−50.
[32] Hooshmand MA, Toledo CS, Moghaddas R, Skordilis E. Data analytics for diagnosis and prediction of central line-associated bloodstream infections in critical care units. Comput Inform Nurs 2022;40:365−72.
[33] Rahmani K, Garikipati A, Barnes G, Hoffman J, Calvert J, Mao Q, et al. Early prediction of central line associated bloodstream infection using machine learning. Am J Infect Control 2022;50:440−5.
[34] Wang Y, Li Q, Shu Q, Liu M, Li N, Sui W, et al. Clinical epidemiology and a novel predicting nomogram of central line associated bloodstream infection in burn patients. Epidemiol Infect 2023;151:e90.
[35] Vilela R, Jácomo AD, Tresoldi AT. Risk factors for central venous catheter-related infections in pediatric intensive care. Clinics 2007;62:537−44.
[36] Schalk E, Hanus L, Färber J, Fischer T, Heidel FH. Prediction of central venous catheter-related bloodstream infections (CRBSIs) in patients with haematologic malignancies using a modified Infection Probability Score (mIPS). Ann Hematol 2015;94:1451−6.
[37] Sakai H, Iwata M, Terasawa T. External validation of the Michigan PICC catheter-associated bloodstream infections score (MPC score) for predicting the risk of peripherally inserted central catheter-associated bloodstream infections: a single-center study in Japan. Infect Control Hosp Epidemiol 2021;44:480−3.
[38] Tukey MH, Borzecki AM, Wiener RS. Validity of ICD-9-CM codes for the identification of complications related to central venous catheterization. Am J Med Qual 2015;30:52−7.
[39] Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020;369:m1328.
[40] Erin EK, Kathryn HB, Subhash A. Systematic review of prediction models for postacute care destination decision-making. J Am Med Inform Assoc 2022;29:176−86.
[41] Brown FS, Glasmacher SA, Kearns PKA, MacDougall N, Hunt D, Connick P, et al. Systematic review of prediction models in relapsing remitting multiple sclerosis. PLoS One 2020;15:e0233575.
[42] Heus P, Damen JAAG, Pajouheshnia R, Scholten RJPM, Reitsma JB, Collins GS, et al. Poor reporting of multivariable prediction model studies: towards a targeted implementation strategy of the TRIPOD statement. BMC Med 2018;16:120.
[43] Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. Lancet 2019;393:1577−9.
[44] 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.
[45] Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ 2020;368:m441.
[46] Riley RD, Debray TPA, Collins GS, Archer L, Ensor J, van Smeden M, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. Stat Med 2021;40:4230−51.
[47] Riley RD, Collins GS, Ensor J, Archer L, Booth S, Mozumder SI, et al. Minimum sample size calculations for external validation of a clinical prediction model with a time-to-event outcome. Stat Med 2022;41:1280−95.
[48] Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol 2016;74:167−76.
[49] Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 2016;69:245−7.