



## OPEN ACCESS

## EDITED BY

Benjamin Y. Q. Tan,  
National University Hospital, Singapore

## REVIEWED BY

Nicholas Ngiam,  
National University Health System, Singapore  
Thierry Caus,  
University of Picardie Jules Verne, France

## \*CORRESPONDENCE

Guangzhi Liu  
✉ guangzhi2002@hotmail.com

<sup>†</sup>These authors have contributed equally to this work

RECEIVED 14 September 2023

ACCEPTED 20 November 2023

PUBLISHED 05 January 2024

## CITATION

Yuan B, Wang C, Fan Z, Liu C, Fang L, Ma L, Zou W, Yuan G and Liu G (2024) A Bayesian network-based approach for identifying risk factors and predicting ischemic stroke in infective endocarditis patients.  
Front. Cardiovasc. Med. 10:1294229.  
doi: 10.3389/fcvm.2023.1294229

## COPYRIGHT

© 2024 Yuan, Wang, Fan, Liu, Fang, Ma, Zou, Yuan and Liu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# A Bayesian network-based approach for identifying risk factors and predicting ischemic stroke in infective endocarditis patients

Boyi Yuan<sup>1†</sup>, Chaobin Wang<sup>2†</sup>, Zexin Fan<sup>1</sup>, Chun Liu<sup>3</sup>, Libo Fang<sup>4</sup>, Lin Ma<sup>1</sup>, Wenlong Zou<sup>1</sup>, Guobin Yuan<sup>1</sup> and Guangzhi Liu<sup>1\*</sup>

<sup>1</sup>Department of Neurology, Beijing Anzhen Hospital, Capital Medical University, Beijing, China,

<sup>2</sup>Department of Neurology, Beijing Fangshan District Liangxiang Hospital, Beijing, China, <sup>3</sup>Department of Neurology, Meichinka Hospital, Dnipro State Medical University, Dnipro, Ukraine, <sup>4</sup>Department of Neurology, Fuxing Hospital, Capital Medical University, Beijing, China

**Objective:** This study aimed to seek the risk factors and develop a predictive model for ischemic stroke (IS) in patients with infective endocarditis (IE) utilizing a Bayesian network (BN) approach.

**Methods:** Data were obtained from the electronic medical records of all adult patients at three hospitals between 1 January 2018, and 31 December 2022. Two predictive models, logistic regression and BN, were used. Patients were randomly assigned to the training and test sets in a 7:3 ratio. We established a BN model with the training dataset and validated it with the testing dataset. The Bayesian network model was built by using the Tabu search algorithm. The areas under the receiver operating characteristic curve (AUCs), calibration curve, and decision curve were used to evaluate the prediction performance between the BN and logistic models.

**Results:** A total of 542 patients [mean (SD) age, 49.6 (15.3) years; 137 (25.3%) female] were enrolled, including 151 (27.9%) with IS and 391 (72.1%) without IS. Hyperlipidemia, hypertension, age, vegetation size (>10 mm), *S. aureus* infection, and early prosthetic valve IE were closely correlated with IS. The BN models outperformed the logistic regression in training and testing sets, with accuracies of 76.06% and 74.1%, AUC of 0.744 and 0.703, sensitivities of 25.93% and 20.93%, and specificities of 96.27% and 90.24%, respectively.

**Conclusion:** The BN model is more efficient than the logistic regression model. Therefore, BN models may be suitable for the early diagnosis and prevention of IS in IE patients.

## KEYWORDS

Bayesian network, infective endocarditis, stroke, risk factor, prediction model

## 1 Introduction

Infective endocarditis (IE) is an uncommon, life-threatening disease caused by endocardial surface infections. Neurological complications in IE patients include ischemic/hemorrhagic stroke, mycotic aneurysms, brain abscesses and meningitis, spondylodiscitis, spinal cord abscesses, and encephalopathy (1). Among these, ischemic stroke (IS) is one of the most disastrous complications of IE, accounting for 16%–25% of patients (2–4), and ultimately leading to high mortality and morbidity (5, 6). Hence, the accurate and timely identification of IS in IE patients is crucial, as it could improve disease prognosis

and quality of life. The risk factors for ischemic and hemorrhagic stroke include male sex, older age, prior IE or stroke, *S. aureus* infection, fungal infection, atrial fibrillation, multiple vegetation, mitral valve vegetation, valvular abscess, large vegetation (>10 mm), and rheumatic heart disease (4, 7–10). Although predictive models (e.g., ER French Calculator) have been proposed to assess the risk of embolic events in IE patients, they lack external validation (11, 12). Given the unavailability of universally accepted methods for predicting future embolic events, particularly IS, in IE patients, highly predictive models are urgently needed for early IS identification and diagnosis. Herein, we built models separately for IS in IE patients using the Bayesian network (BN) method and logistic regression. We compared the performance of the models using areas under the curve (AUCs), calibration curve, and decision curve analysis (DCA).

## 2 Material and methods

### 2.1 Setting and data sources

Data were obtained from the electronic medical records of all adult patients at three hospitals between 1 January 2018, and 31 December 2022. A randomization method was used for patient's allocation. Data were analyzed from 1 February to 31 September 2023. The inclusion criteria were: (i) age  $\geq 18$  years; (ii) diagnosis of possible and definite IE in accordance with the modified Duke criteria (13). The exclusion criteria were: (i) right-sided IE cases were excluded unless a concomitant left-sided infection was present, (ii) patients with other tumors or gravidity (14), and (iii) patients with missing clinical data such as demographic information, medical history, comorbidities, echocardiographic characteristics, or laboratory examination. All patient underwent transesophageal echocardiography (TEE), and most of them ( $n = 425$ ) also underwent transthoracic echocardiography (TTE) examination. M-mode was used to measure left ventricular ejection fraction [LVEF], left atrial diameter [LAD], left ventricular end-diastolic diameter [LVEDD]. High-mobility refers to highly mobile phenomena of vegetation mass on transesophageal echocardiography (TEE) (15, 16), but so far there are no literatures dealing with the threshold value of the width of mobility. To ensure the standardization of variable across operators, we selected 10 mm as the threshold, since one previous study revealed vegetation size >10 mm were more frequently seen in IE patients complicated by embolic events (17). Early prosthetic valve IE refers to IE that occurs within 1 year after implantation of a prosthetic valve. All participating hospitals' ethics and research committees approved the study.

### 2.2 Outcome variables

To diagnose the target outcome, IS, medical histories, clinical examinations, magnetic resonance angiography scan, and cranial magnetic resonance imaging results were acquired prior to cardiac surgery. Two independent, blinded neurologists confirmed the diagnosis of IS according to medical history,

clinical examination and results of cerebral magnetic resonance imaging (MRI) and magnetic resonance angiography scans.

### 2.3 Data extraction and quality control

Two researchers extracted the data. Input data, including demographic information (sex, age), medical history [hypertension, hyperlipidemia, diabetes, and atrial fibrillation (AF)], comorbidities, echocardiography (affected valve, vegetation size, vegetation mobility, LVEF, LAD, LVEDD, and regurgitation), electrocardiogram, and laboratory tests (blood culture results, high sensitivity troponin I (Hs-TNI), estimated glomerular filtration rate [eGFR], hemoglobin, and D-dimer, were gained from the hospital's electronic medical record system upon completion of initial admission. The data collection process was standardized, and the researchers were already familiar with this process before initiating data retrieval. A double-entry approach was used. Medical records were reviewed, and discrepancies were corrected during the review period.

### 2.4 Data processing for potential variables

To establish the predictive model, we consulted relevant literature and processed the data accordingly based on previous studies. In two meta-analysis studies, larger vegetations (>10 mm) significantly increase the rates of embolism and mortality in patients (9). An infection with *S. aureus* is linked to a higher risk of short-term embolic events (EEs) (18). EEs are common among patients with IE and are associated with mitral valve endocarditis, as well as the usage of antiplatelet drugs or intravenous drug (19). Moreover, a remarkable association was identified between elevated plasma D-dimer levels and the risk of IS (20). Thoker et al, conducted a study which indicated that patients with IE and elevated levels of cardiac troponin I were associated with an increased odds of adverse clinical outcomes, including cerebral mycotic aneurysm and meningitis (21). Biostatistical literature suggests that if the rate of missing data in a dataset exceeds 30%, the data will lose its measure of confidence. Most of these attributes are missing due to test failures and time conflicts. Hence, in our study, instances with more than 9 missing attributes (out of 27) were excluded from the dataset. Finally, combined with traditional IS risk factors, we identified 27 variables, including age, hypertension, hyperlipidemia, diabetes, and AF as the dataset. We first screened for potential IS-related risk factors to build a predictive model using logistic regression. The data were split into training and testing sets in a 7:3 ratio using a random number table. IS was then predicted using BN and logistic regression models. A Tabu search algorithm was utilized to establish the BN model. Model performance was evaluated by use of test sets. The IS-related factors were measured and coded before building the BN model.

When constructing predictive models for dichotomous outcomes, it is recommended to have a sample size that is at least 10 times larger than the number of independent variables. Due to the inclusion of 11 variables in the multivariate analysis, a minimum sample size of 110 per group was required.

Eventually, we included 151 and 391 IE patients with and without IS, respectively. Therefore, our sample size was sufficient.

## 2.5 BNs

The BN model is represented as a directed acyclic graph (22). The nodes in the graph represent random variables, while the directed edges symbolize the probabilistic dependencies between variables in the model. If a directed arc exists from  $X_1$  to  $X_2$ , it represents  $X_1$  leading to  $X_2$ ;  $X_1$  is the parent node, and  $X_2$  is the child node. The state of each node's parent node is represented by a conditional probability distribution table associated with that node. The BN represents the joint probability distributions of random variables  $X = \{X_1, \dots, X_n\}$ ; thus, a probability expression can be obtained.

$$P(X_1, \dots, X_n) = P(X_1)P(X_2|X_1) \dots P(X_n|X_1, X_2, \dots, X_{n-1}) \\ = \prod_1^n P(X_i|\pi(X_i))$$

where  $P(X_i)$  represents the collection of the parents of  $X_i$ ;  $\pi(X_i) \subseteq \{X_1, \dots, X_{i-1}\} \subseteq \{X_1, \dots, X_{i-1}\}$  (23).

For each instance, 27 random variables were derived from the patient data. Given that too many variables add unnecessary complexity to the BN structure, we used univariate analysis to screen the nodes. The Tabu search algorithm was then used to construct the best model.

## 2.6 Statistical analysis

SPSS Statistics for Windows (version 23.0) was utilized for conducting the statistical analyses. Continuous variables were expressed as mean  $\pm$  standard deviation or median (interquartile

range). Categorical variables were presented as number of subjects ( $n$ ) and percentage (%). Categorical variables were analyzed utilizing the  $\chi^2$  test. Normally distributed data and non-normally distributed data are analyzed using the Student's  $t$ -test and the Mann–Whitney  $U$ -test, respectively. Variables associated with IE-related IS were evaluated using binary logistic regression analysis. In the univariate analysis, variables in association with the outcome at a significance level of  $<0.05$  are deemed potential candidates for further consideration in the subsequent multivariate analysis. Receiver operating characteristic (ROC) curve analysis was utilized to evaluate the predictive models and calculate the AUCs. The Delong test was utilized to evaluate the statistical significance of the disparity among AUC values. To assess the difference between predicted values and true values and plot a calibration curve, the Hosmer–Lemeshow (HL) test was utilized. DCA was performed to assess the net benefits of each predictive model. The Tabu Search algorithm was employed to build a BN model. The maximum likelihood estimation method was utilized for parameter estimation of the BN model. R Studio 4.2.2 was used to establish the BN model. To visualize the BN topology, we employed Netica32 software (Norsys Software Corp., Vancouver, Canada).

## 3 Result

### 3.1 Patients selection

Of the 605 patients with IE, 63 individuals were excluded from the study due to right-sided IE, other comorbidities, or missing data. Finally, 542 patients [mean (SD) age, 49.6 (15.3) years; 137 (25.3%) female] were enrolled, including 151 (27.9%) with IS and 391 (72.1%) without IS (Figure 1 and Table 1). IS was confirmed before cardiac surgery (interventions) among all IE patients. In this study, some IE patients were on anticoagulation due to prosthetic valve replacement or AF, but no significant difference in anticoagulation management was found between the patients

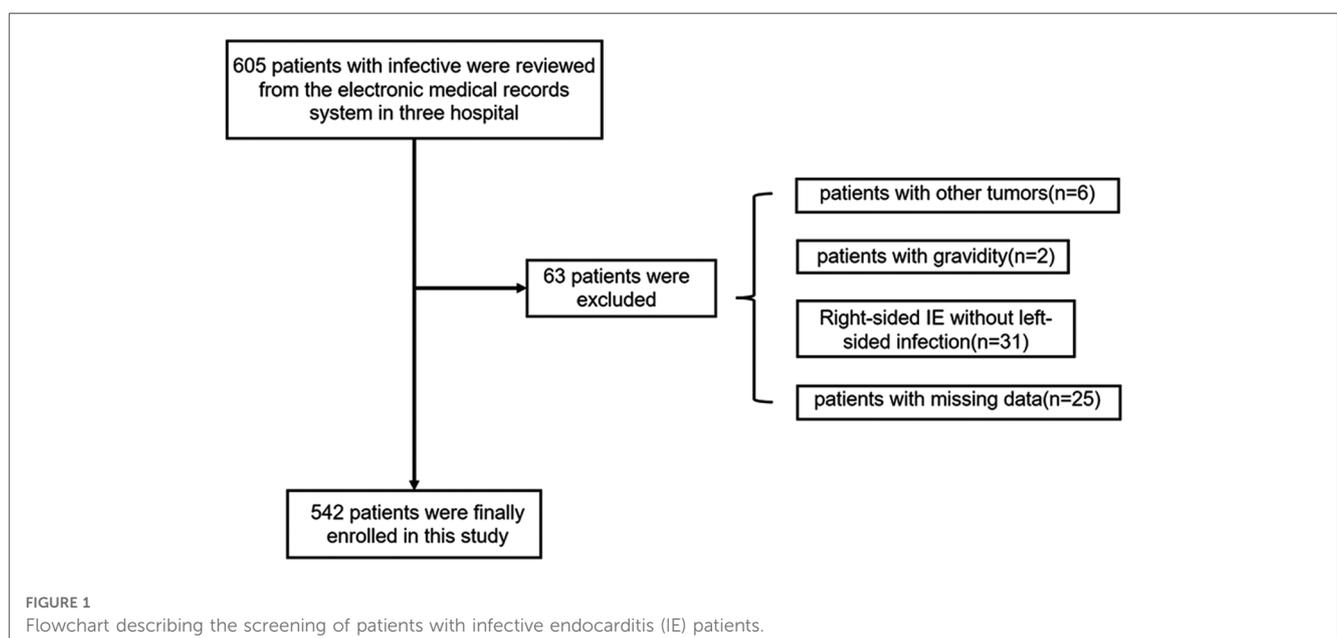


TABLE 1 Baseline data of patients with infective endocarditis (IE).

Variables	IE without IS (N = 391)	IE with IS (N = 151)	P-value
Sex (Male)	299 (76.5)	106 (70.2)	0.1627
Age, mean (SD), years	47 (34,58)	59 (51,65)	<b>&lt;0.001</b>
Current smoking	108 (27.6)	46 (30.5)	0.5813
Current drinking	80 (20.5)	33 (21.9)	0.8102
Hypertension	72 (18.4)	57 (37.7)	<b>&lt;0.001</b>
Hyperlipidemia	10 (2.6)	12 (7.9)	<b>0.009</b>
Diabetes	53 (13.6)	30 (19.9)	0.09
Atrial fibrillation	20 (5.1)	20 (13.2)	<b>0.0022</b>
Cardiac function (class III-IV)	145 (37.1)	52 (34.4)	0.6349
Early prosthetic valve IE	13 (3.3)	15 (9.9)	<b>0.0037</b>
Late prosthetic valve IE	32 (8.2)	9 (6.0)	0.4861
Affected valve			
Mitral	222 (56.8)	98 (64.9)	0.1038
Aortic	237 (60.6)	90 (59.6)	0.9062
Vegetation size (>10 mm)	145 (37.1)	74 (49.0)	<b>0.0148</b>
Vegetation mobility	202 (51.7)	97 (64.2)	<b>0.011</b>
LVEF (%)	61 (57, 66)	61 (57, 66)	0.8057
LAD (mm)	42 (37, 46)	43 (38, 46)	0.201
LVEDD (mm)	57 (52, 62)	56 (51, 61)	0.1394
Mitral regurgitation	235 (60.1)	99 (65.6)	0.2831
Aortic regurgitation	201 (51.4)	79 (52.3)	0.9248
Perivalvular abscess	25 (6.4)	15 (9.9)	0.2187
<i>S. aureus</i>	7 (1.8)	9 (6.0)	<b>0.0193</b>
Hypohepatia			
Normal	340 (87.0)	128 (84.8)	
Low	41 (10.5)	19 (12.6)	
Middle	4 (1.0)	3 (2.0)	
High	6 (1.5)	1 (0.7)	
Hemoglobin (g/L)	109 ± 21.4	104 ± 21.5	<b>0.047</b>
eGFR (ml/min/1.73 m <sup>2</sup> )	103.5 (85.7, 118.5)	88.9 (58.9, 104.6)	<b>&lt;0.001</b>
D-dimer (ng/ml)	339 (207, 674)	481 (323, 891)	<b>&lt;0.001</b>

IE, infective endocarditis; AF, atrial fibrillation; eGFR, estimated glomerular filtration rate; IS, ischemic stroke; LVEDD, left ventricular end-diastolic diameter; LVEF, left ventricular ejection fraction; LAD, left atrium diameter; *S. aureus*, *Staphylococcus aureus*. P-values less than 0.05 have been highlighted in bold.

with IS and those without IS [55 (14.1%, 55/391) vs. 26 (17.2%, 26/151),  $P = 0.43$ ]. Moreover, thirteen patients with cardiac device-associated IE were included in this study. All IE patients underwent blood culture, and the results are presented in Table 2. Although *Viridans group streptococci* was the most common pathogen, no significant difference was found between it and other pathogen isolated ( $P = 0.072$ ).

TABLE 2 Microbiological etiology of infective endocarditis.

Causal agent	IE With IS	IE Without IS
<i>S. aureus</i>	9	7
Coag Neg staph	10	18
<i>Viridans group streptococci</i>	22	55
<i>Enterococci</i>	5	8
HACEK	3	0
other	6	13
Culture negative	99	287

*S. aureus*, *Staphylococcus aureus*; Coag Neg staph, coagulase-negative staphylococci; HACEK, Haemophilus spp., Aggregatibacter (formerly Actinobacillus) actinomycetemcomitans, Cardiobacterium hominis, Eikenella corrodens, and Kingella species.

## 3.2 Risk factors for IS

The risk factors associated with IS were hypertension [odds ratio (OR), 2.687; 95% confidence interval (CI), 1.771–4.076;  $P < 0.001$ ], older age (OR, 1.049; 95% CI, 1.034–1.064;  $P < 0.001$ ), hyperlipidemia (OR, 3.289; 95% CI, 1.390–7.784;  $P = 0.007$ ), atrial fibrillation (AF) (OR, 2.832; 95% CI, 1.477–5.430;  $P = 0.002$ ), early prosthetic valve IE (OR, 3.207; 95% CI, 1.488–6.913;  $P = 0.003$ ), vegetation size (>10 mm) (OR, 1.630; 95% CI, 1.116–2.383;  $P = 0.012$ ), vegetation mobility (OR, 1.681; 95% CI, 1.141–2.476;  $P = 0.009$ ), *S. aureus* infection (OR, 3.477; 95% CI, 1.271–9.511;  $P = 0.015$ ), estimated glomerular filtration rate (eGFR) (OR, 0.983; 95% CI, 0.977–0.989;  $P < 0.001$ ), hemoglobin (OR, 0.990; 95% CI, 0.982–0.999;  $P = 0.035$ ), D-dimer (OR, 1.000; 95% CI, 1.000–1.000;  $P = 0.262$ ).

In the logistic regression models, we incorporated four variables that exhibited statistical significance, including older age (OR, 1.037; 95% CI, 1.019–1.055;  $P < 0.001$ ), vegetation mobility (OR, 1.642; 95% CI, 1.042–2.590;  $P = 0.017$ ), *S. aureus* infection (OR, 4.5; 95% CI, 1.375–14.733;  $P = 0.013$ ), and early prosthetic valve IE (OR, 3.529; 95% CI, 1.544–8.066;  $P = 0.003$ ) (Table 3).

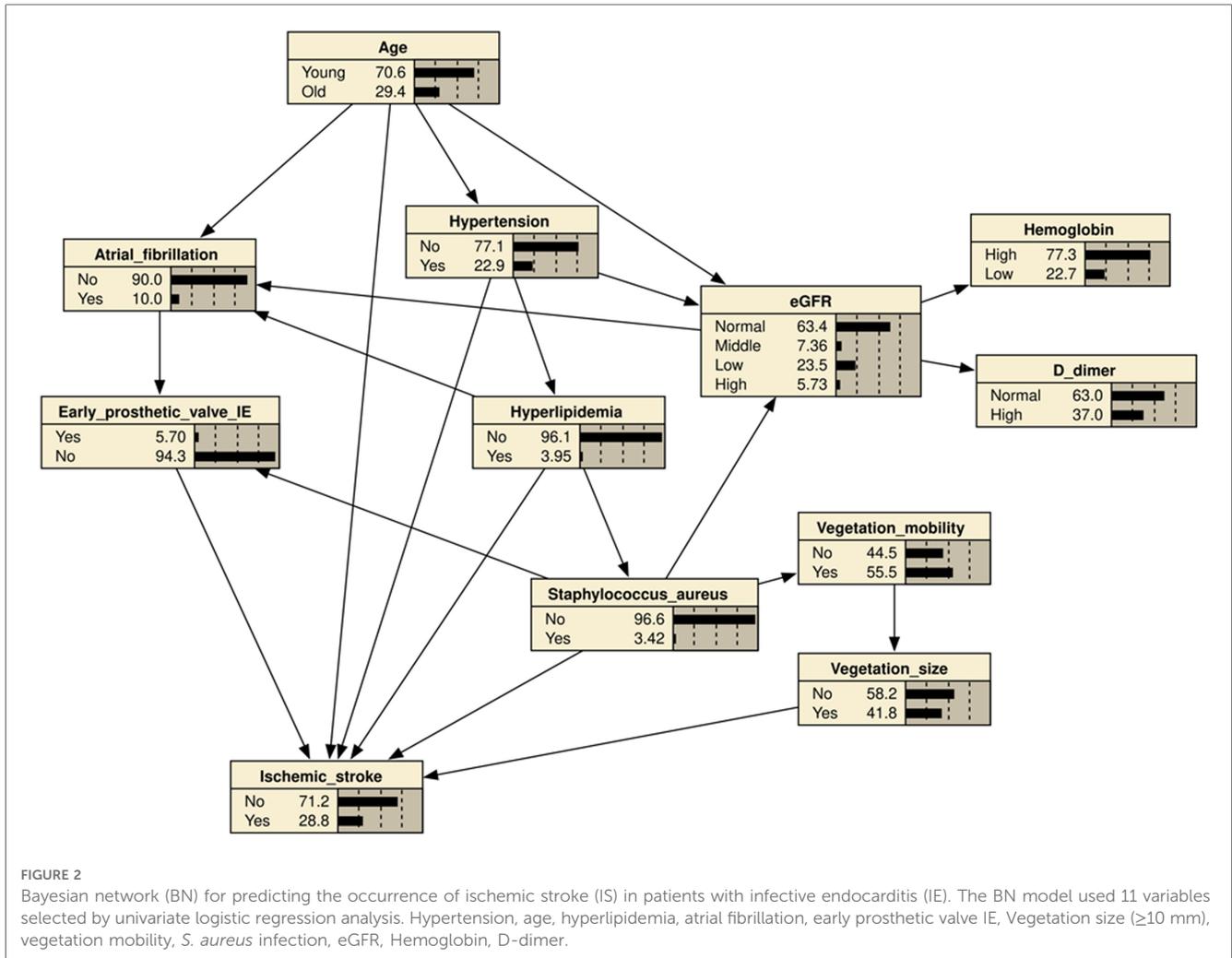
## 3.3 BN structure

It was based on a BN model consisting of 21 directed edges and 12 nodes. The nodes represented IS, eGFR, hyperlipidemia, hypertension, age, AF, early prosthetic valve IE, vegetation mobility, vegetation size (>10 mm), *S. aureus* infection, D-dimer, and hemoglobin. The nodes directly linked to IS included age, hyperlipidemia, hypertension, vegetation size (>10 mm), early prosthetic valve IE, and *S. aureus* infection. AF was indirectly associated with IS through its association with early prosthetic valve IE. Vegetation mobility was indirectly associated with IS through its correlation with vegetation size (>10 mm). The eGFR was indirectly associated with IS through its link with AF and early prosthetic valve IE (Figure 2). Based on maximum likelihood estimation, the conditional probabilities of each node in the network (Table 4) were estimated. The variables predictive

TABLE 3 Risk factors of IS in infective endocarditis (IE) patients: multivariate binary logistic regression analysis.

Category	B	SE	Wald	OR (95% CI)	P-value
Age	0.036	0.009	16.764	1.037 (1.019, 1.055)	<b>&lt;0.001</b>
Hypertension	0.478	0.245	3.804	1.613 (0.998, 2.607)	0.051
Hyperlipidemia	0.461	0.501	0.849	1.586 (0.595, 4.232)	0.357
D-dimer	0.000	0.000	0.123	1.000 (1.000, 1.000)	0.725
Early prosthetic valve IE	1.261	0.422	8.939	3.529 (1.544, 8.066)	<b>0.003</b>
<i>S. aureus</i>	1.504	0.605	6.179	4.5 (1.375, 14.733)	<b>0.013</b>
Vegetation size (>10 mm)	0.383	0.232	2.731	1.466 (0.931, 2.309)	0.098
Vegetation mobility	0.496	0.232	4.558	1.642 (1.042, 2.590)	<b>0.033</b>
AF	0.283	0.376	0.569	1.328 (0.636, 2.772)	0.451
eGFR	-0.005	0.004	1.932	0.995 (0.987, 1.002)	0.165
Hemoglobin	-0.001	0.005	0.049	0.999 (0.989, 1.009)	0.825

IE, infective endocarditis; IS, ischemic stroke; AF, atrial fibrillation; eGFR, estimated glomerular filtration rate; *S. aureus*, *Staphylococcus aureus*. P-values less than 0.05 have been highlighted in bold.



of IS were hyperlipidemia, age, hypertension, early prosthetic valve IE, vegetation size ( $>10$  mm), and *S. aureus* infection.

### 3.4 Model performance evaluation

The performance of the two models was assessed by examining metrics including accuracy, AUC, specificity, sensitivity, calibration curve, and decision curve (Table 5). The logistic regression predictive model exhibited accuracies of 73.4% and 74.1%, AUCs of 0.734 and 0.693, sensitivities of 25.9% and 30.23%, and specificities of 92.5% and 89.43% in the training and testing datasets, respectively. On the other hand, The BN model exhibited accuracies of 76.06% and 74.1%, AUCs of 0.744 and 0.703, sensitivities of 25.93% and 20.93%, and specificities of 96.27% and 90.24%. No statistically significant difference in the AUC values was shown between the logistic regression and BN models (Figure 3). The Delong test was performed in the training and testing cohorts, respectively, with  $P$ -values  $>0.05$  ( $P = 0.8102$  and  $P = 0.849$ ). Moreover, the calibration curves demonstrated that the BN model outperformed the logistic model regarding the degree of fit between the actual and predicted probabilities (Figure 4). Both the logistic regression

model and BN model demonstrated good calibration in either training sets [BN:  $P = 0.999$ ,  $\chi^2 < 0.001$ , degrees of freedom (df) = 8; logistic regression:  $P = 0.6256$ ,  $\chi^2 = 6.193$ , df = 8] or testing sets (BN:  $P = 0.3496$ ,  $\chi^2 = 8.914$ , df = 8; logistic regression:  $P = 0.1469$ ,  $\chi^2 = 12.097$ , df = 8) as indicated by the Hosmer–Lemeshow test. DCA showed a non-significant difference in net benefits between the BN and logistic regression models for predicting IS in the training or test sets (Figure 5).

## 4 Discussion

In the present study, we employed a BN-based approach with the available data to predict the occurrence of IS. Hyperlipidemia, *S. aureus* infection, early prosthetic valve hypertension, age, and vegetation size ( $>10$  mm) were directly correlated with IS, whereas AF, eGFR, and vegetation mobility were indirectly related with IS. On the other hand, the BN model achieved superior or equivalent predictive performance for IS to the logistic regression model.

Consistent with previous studies (4, 16, 18), our study demonstrated that *S. aureus* infection was associated with an

TABLE 4 The conditional probability table of the training set based on ischemic stroke (IS) as the target node.

Age ≥60	Vegetation size (>10 mm)	Hyperlipidemia	<i>S. aureus</i>	Early prosthetic valve IE	Hypertension	IE with IS	IE without IS
Yes	No	No	No	No	No	0.34	0.66
Yes	No	No	No	No	Yes	0.35	0.65
Yes	No	No	Yes	No	No	1	0
Yes	No	No	Yes	No	Yes	1	0
Yes	Yes	No	No	No	No	0.45	0.55
Yes	Yes	No	No	No	Yes	0.6	0.4
Yes	Yes	No	Yes	No	No	1	0
Yes	Yes	No	Yes	No	Yes	1	0
Yes	No	No	No	Yes	No	1	0
Yes	No	No	No	Yes	No	1	0
Yes	Yes	No	No	Yes	No	1	0
No	No	No	No	No	No	0.09	0.91
No	No	No	No	No	Yes	0.3	0.7
No	No	No	Yes	No	No	0.4	0.6
No	Yes	No	No	No	No	0.24	0.76
No	Yes	No	No	No	Yes	0.3	0.7
No	No	No	No	Yes	No	0.29	0.71
No	Yes	No	Yes	No	No	1	0
No	No	No	No	Yes	Yes	0.66	0.33
No	Yes	No	No	Yes	No	0.5	0.5
No	Yes	No	No	Yes	Yes	1	0
Yes	No	Yes	No	No	Yes	0.66	0.33
Yes	Yes	Yes	No	No	Yes	0.5	0.5
No	No	No	No	Yes	Yes	0.67	0.33
Yes	Yes	No	No	Yes	No	1	0
No	Yes	No	No	Yes	No	0.5	0.5
Yes	No	No	No	Yes	Yes	1	0
No	No	Yes	No	No	Yes	0.5	0.5
No	No	Yes	No	No	No	1	0
No	Yes	Yes	No	No	No	0	1
No	Yes	Yes	No	No	Yes	0.5	0.5

IE, infective endocarditis; IS, ischemic stroke; *S. aureus*, *staphylococcus aureus*.

increased risk of IS. The mechanisms underlying IS remain largely unknown. Hence, we speculate that infections may cause inflammation and procoagulation, including vascular leakage, endothelial injury, and hypotension. These events lead to atherosclerotic plaque instability and rupture and arterial occlusion, ultimately result in arterial thrombosis (24–26). Further studies are required to address this issue.

The risk of stroke is remarkably increased in multiple vegetation types, including mitral valve vegetation, valvular abscess, and large vegetation (>10 mm) (4, 7, 9, 27). Our study found that larger vegetation was significantly associated with IS. Additionally, vegetation mobility was indirectly related to IS through its association with vegetation size (>10 mm). Concurrently, two distinct studies demonstrated that vegetation size > 10 mm and high vegetation mobility were predictors of

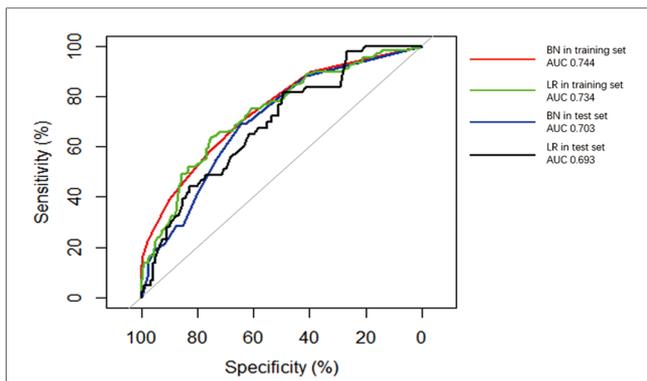
embolic events (28–31). Therefore, to prevent IS, an echocardiographic examination of the cardiac structure and early management should be routinely performed in IE patients.

We observed a significant association between early prosthetic valve IE and IS. This association is noteworthy because prosthetic valve endocarditis (PVE) accounts for approximately 30% of IE cases worldwide, with in-hospital mortality of 20%–40%, partially due to cerebrovascular events or stroke (32–34). *S. aureus* is the most prevalent pathogen implicated in PVE, and patients with *S. aureus* PVE have notably higher rates of mortality in hospital settings (33, 35). Hence, accurate and dynamic monitoring of *S. aureus*-related PVE and timely intervention could improve clinical outcomes.

In our study, hypertension, older age, and hyperlipidemia were directly associated with IS, implying that such risk factors require strict monitoring and management in IE patients, especially in older adults, to prevent IS. In a cohort of 507 patients diagnosed with native left-sided IE, approximately 10.3% experienced new-onset atrial AF, an established risk factor for IS. Notably, patients with pre-existing or newly developed AF demonstrated significantly higher 1-year in-hospital mortality rates compared to patients without AF (36). The impact of IE on heart valves can lead to subsequent heart failure and valvular insufficiency, which in turn can contribute to the onset or progression of AF (37). Our results revealed that AF was indirectly linked to IS

TABLE 5 The performance of different predictive models.

Model	Accuracy	AUC	Sensitivity	Specificity
Bayesian network (training set)	76.06%	0.744	25.93%	96.27%
Logistic regression (training set)	73.4%	0.734	25.9%	92.5%
Bayesian network (test set)	74.1%	0.703	20.93%	90.24%
Logistic regression (test set)	74.1%	0.693	30.23%	89.43%

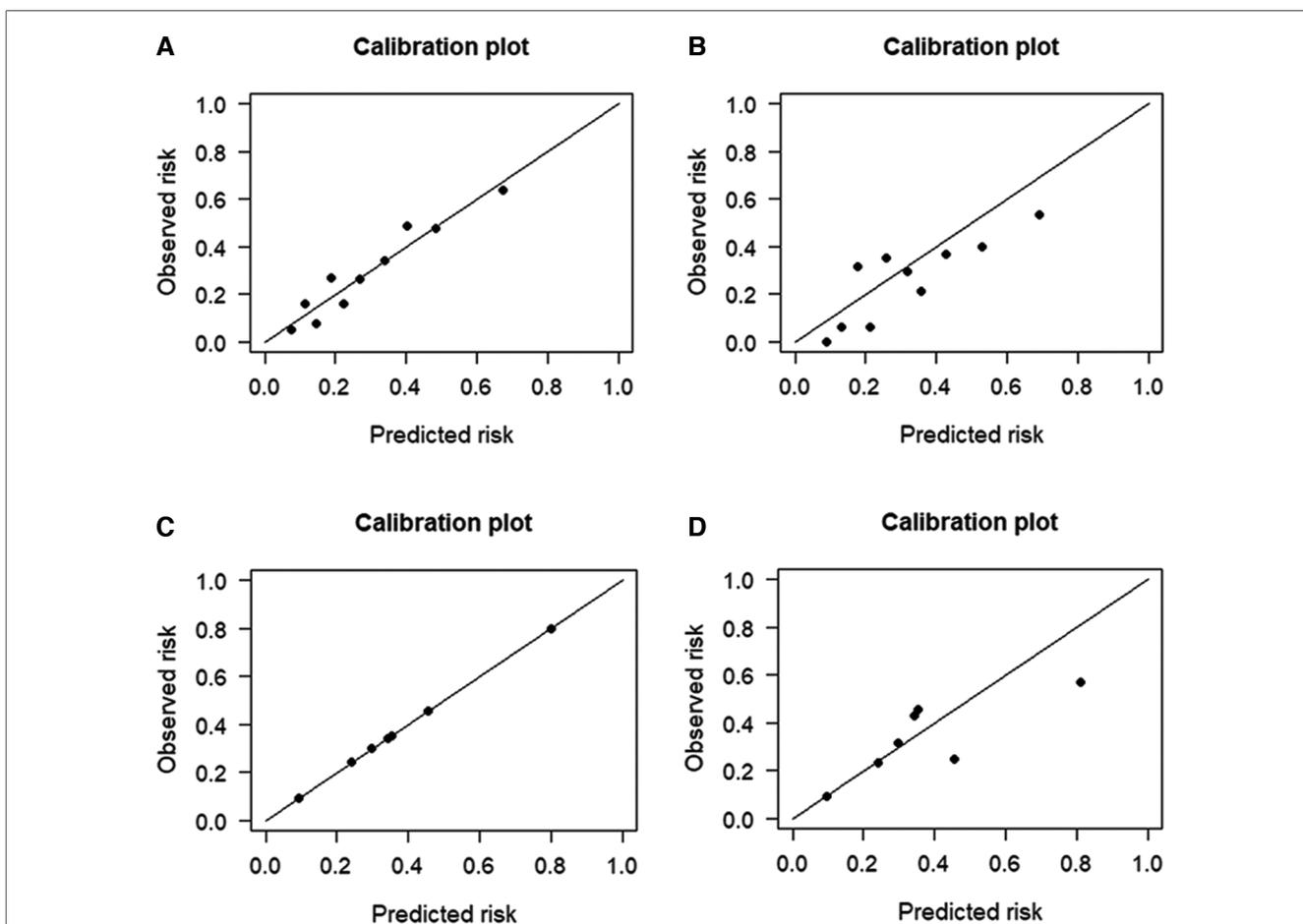


**FIGURE 3**  
Receiver operating characteristic (ROC) curves of Bayesian network (BN) model and logistic regression (LR) model for predicting ischemic stroke (IS) in patients with infective endocarditis (IE). The areas under the curve (AUCs) of the BN model predicting IS was 0.744 and 0.703 in (red line) training and (blue line) test datasets, respectively. The AUC of the logistic regression LR model predicting IS was 0.734 and 0.693 in the (green line) training and (black line) test datasets.

through its association with early prosthetic valve IE utilizing the BN model, further proving our viewpoint. Nonetheless, more evidence is needed to address this issue.

BN models offer several advantages in the medical field. Bayesian networks can handle missing data, a feature distinct from logistic models, in which missing values are not allowed between covariates. Because missing data is very common in clinical practice, BN is well-suited for establishing diagnostic models. Moreover, BN typically models structures in a domain, and the results are intuitively visible. BN enables quantitative risk assessment of selected clinically relevant outcomes. As shown in **Table 4**, an elderly patient without hypertension, *S. aureus* infection, vegetation size (>10 mm), vegetation mobility, hyperlipidemia and early prosthetic heart valve had a probability of 0.34 for concurrent IS; In the presence of vegetation size (>10 mm), the probability increased to 0.45. The probability further rose to 0.6 if these patients also had hypertension. Thus, our results can be easily applied in clinical practice and is very beneficial for early IS detection and diagnosis of patients. Furthermore, these factors may contribute to the prevention of initial and recurrent IS.

This study's calibration results revealed an agreement between the BN and logistic regression models. The DCA curves indicated that BN and the logistic regression model had



**FIGURE 4**  
Calibration plots for the four prediction models in both cohorts. The calibration plots showed that the predicted risk of ischemic stroke (IS) agreed well with the observed risk, in either the logistic regression model of (A) training and (B) testing datasets, or in the Bayesian network model of (C) training and (D) testing datasets.

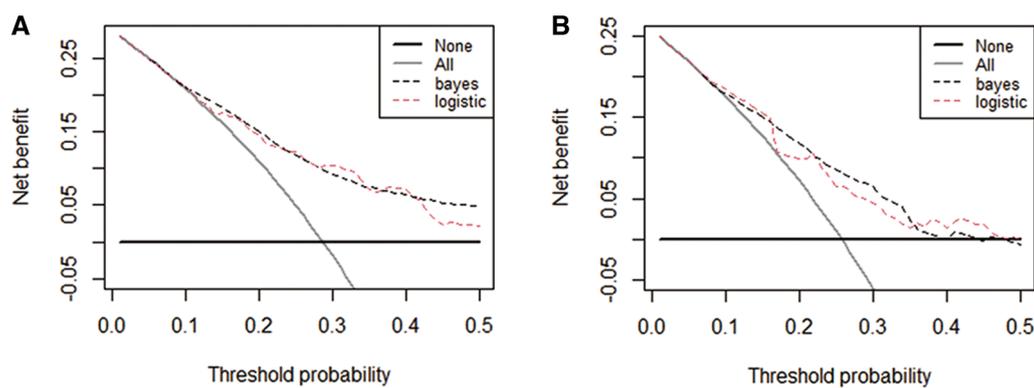


FIGURE 5

Decision curve analysis (DCA) curves plots for the four prediction models in both cohorts. DCA curves of each model for predicting the probability of IS in the training set (A) and test set (B). The horizontal and vertical axes represent the threshold probability and net benefit, respectively. The lines between the horizontal axis and vertical axis display the benefits of the different predictive variables. Compared to the logistic regression model, the decision curve showed that the Bayesian model for IS performed equally well in the training and test sets.

good net benefits in predicting IS. Additionally, the performance of the BN model was better or no worse than the traditional logistic regression model, in the context of accuracy, AUC, specificity and sensitivity. More importantly, the logistic model assumes that each variable is independent and ignores the relationship between risk factors, while the BN model builds a network model through the deep mining of data, which further reveals the interaction between variables and more realistically reflects the effect of each risk factor on IS. However, in clinical trials, missing data for the predictive model can potentially render the prediction unable to proceed. On the contrary, the BN is built upon disease-related knowledge, effectively harnessing the available data to uncover valuable information and unveil the intricate interactions among multiple factors.

To our knowledge, this study, for the first time, construct a BN model for predicting IS in IE patients, and it was more efficient than the logistic regression model. Age, early prosthetic valve IE, vegetation size (>10 mm), hypertension, hyperlipidemia, and *S. aureus* infection were significant predictors of IS in our Chinese cohort. Further research is needed to investigate the risk factors for IS in patients with IE and to determine their causal relationships. This research aims to enhance strategies for disease prevention and eventually reduce the embolic rate and mortality among IE patients.

## 5 Limitations of the study

This study had some limitations. The first limitation in our study is the small sample size, primarily due to the rarity of IE. Second, the BN-directed edges describe the probabilistic dependencies between variables, instead of a causal relationship. Third, the study is exposed to a detection bias for the determination of the baseline characteristics, since the samples were assembled retrospectively. Fourth, considering the predictors

at admission in this study, the model may not study the predictive value of their modifications after the initial evaluation. Last but not least, we didn't handle mortality events and included a composite outcome of stroke and mortality, as this is a cross-sectional study to screen the risk factors and develop a predictive model for IS. Therefore, future large-scale prospective studies are needed.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by Ethics Committee of Beijing Anzhen Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from a by product of routine care or industry. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

## Author contributions

BY: Writing – original draft. CW: Formal analysis, Writing – original draft. ZF: Methodology, Writing – original draft. CL: Writing – original draft. LF: Data curation, Writing – original draft. LM: Methodology, Writing – original draft. WZ:

Methodology, Writing – original draft. GY: Methodology, Writing – original draft. GL: Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article.

The research was supported by a grant from China National Natural Science Foundation (grant no. 82071342).

## Acknowledgments

We thank all the participants, research assistants in this study. We also thank Editage's English language editing.

## References

- Sotero FD, Rosário M, Fonseca AC, Ferro JM. Neurological complications of infective endocarditis. *Curr Neurol Neurosci Rep.* (2019) 19(5):23. doi: 10.1007/s11910-019-0935-x
- García-Cabrera E, Fernández-Hidalgo N, Almirante B, Ivanova-Georgieva R, Noureddine M, Plata A, et al. Neurological complications of infective endocarditis: risk factors, outcome, and impact of cardiac surgery: a multicenter observational study. *Circulation.* (2013) 127(23):2272–84. doi: 10.1161/CIRCULATIONAHA.112.000813
- Murdoch DR, Corey GR, Hoen B, Miró JM, Fowler VG, Bayer AS, et al. Clinical presentation, etiology, and outcome of infective endocarditis in the 21st century: the international collaboration on endocarditis-prospective cohort study. *Arch Intern Med.* (2009) 169(5):463–73. doi: 10.1001/archinternmed.2008.603
- Valenzuela I, Hunter MD, Sundheim K, Klein B, Dunn L, Sorabella R, et al. Clinical risk factors for acute ischaemic and haemorrhagic stroke in patients with infective endocarditis. *Intern Med J.* (2018) 48(9):1072–80. doi: 10.1111/imj.13958
- Ruttman E, Willeit J, Ulmer H, Chevtchik O, Höfer D, Poewe W, et al. Neurological outcome of septic cardioembolic stroke after infective endocarditis. *Stroke.* (2006) 37(8):2094–9. doi: 10.1161/01.STR.0000229894.28591.3f
- Sonneville R, Mirabel M, Hajage D, Tubach F, Vignon P, Perez P, et al. Neurologic complications and outcomes of infective endocarditis in critically ill patients: the ENDOcardite en REAnimation prospective multicenter study. *Crit Care Med.* (2011) 39(6):1474–81. doi: 10.1097/CCM.0b013e3182120b41
- Cao GF, Liu W, Bi Q. Stroke in patients with infective endocarditis: a 15-year single-center cohort study. *Eur Neurol.* (2018) 80(3–4):171–8. doi: 10.1159/000495149
- Mishra A, Sahu KK, Abraham BM, Sargent J, Kranis MJ, George SV, et al. Predictors, patterns and outcomes following infective endocarditis and stroke. *Acta Biomed.* (2022) 93(2):e2022203. doi: 10.23750/abm.v93i2.10185
- Mohananey D, Mohadjer A, Pettersson G, Navia J, Gordon S, Shrestha N, et al. Association of vegetation size with embolic risk in patients with infective endocarditis: a systematic review and meta-analysis. *JAMA Intern Med.* (2018) 178(4):502–10. doi: 10.1001/jamainternmed.2017.8653
- Østergaard L, Andersson NW, Kristensen SL, Dahl A, Bundgaard H, Iversen K, et al. Risk of stroke subsequent to infective endocarditis: a nationwide study. *Am Heart J.* (2019) 212:144–51. doi: 10.1016/j.ahj.2019.03.010
- Hubert S, Thuny F, Resseguier N, Giorgi R, Tribouilloy C, Le Dolley Y, et al. Prediction of symptomatic embolism in infective endocarditis: construction and validation of a risk calculator in a multicenter cohort. *J Am Coll Cardiol.* (2013) 62(15):1384–92. doi: 10.1016/j.jacc.2013.07.029
- Saito F, Toyoda S, Arikawa T, Inami S, Watanabe R, Obi S, et al. Prediction of acute-phase complications in patients with infectious endocarditis. *Intern Med.* (2019) 58(16):2323–31. doi: 10.2169/internalmedicine.1813-18
- Li JS, Sexton DJ, Mick N, Nettles R, Fowler VG, Ryan T, et al. Proposed modifications to the duke criteria for the diagnosis of infective endocarditis. *Clin Infect Dis.* (2000) 30(4):633–8. doi: 10.1086/313753
- Katsafanas C, Bushnell C. Pregnancy and stroke risk in women. *Neurobiol Dis.* (2022) 169:105735. doi: 10.1016/j.nbd.2022.105735
- Shapiro S, Kupferwasser LI. Echocardiography predicts embolic events in infective endocarditis. *J Am Coll Cardiol.* (2001) 37(4):1077–9. doi: 10.1016/S0735-1097(01)01149-4
- Yang A, Tan C, Daneman N, Hansen MS, Habib G, Salaun E, et al. Clinical and echocardiographic predictors of embolism in infective endocarditis: systematic review and meta-analysis. *Clin Microbiol Infect.* (2019) 25(2):178–87. doi: 10.1016/j.cmi.2018.08.010
- Sambola A, Lozano-Torres J, Boersma E, Olmos C, Ternacle J, Calvo F, et al. Predictors of embolism and death in left-sided infective endocarditis: the European society of cardiology EURObservational research programme European infective endocarditis registry. *Eur Heart J.* (2023) 44(43):4566–75. doi: 10.1093/eurheartj/ehad507
- Mejer N, Gotland N, Uhre ML, Westh H, Schönheyder HC, Petersen A, et al. Increased risk of arterial thromboembolic events after Staphylococcus aureus bacteremia: a matched cohort study. *J Infect.* (2015) 71(2):167–78. doi: 10.1016/j.jinf.2015.03.010
- Das AS, McKeown M, Jordan SA, Li K, Regenhardt RW, Feske SK. Neurological complications and clinical outcomes of infective endocarditis. *J Stroke Cerebrovasc Dis.* (2022) 31(8):106626. doi: 10.1016/j.jstrokecerebrovasdis.2022.106626
- Xu N, Fu Y, Wang S, Li S, Cai D. High level of D-dimer predicts ischemic stroke in patients with infective endocarditis. *J Clin Lab Anal.* (2020) 34(5):e23206. doi: 10.1002/jcla.23206
- Thoker ZA, Khan KA, Rashid I, Zafar. Correlation of cardiac troponin I levels with infective endocarditis & its adverse clinical outcomes. *Int J Cardiol.* (2016) 222:661–4. doi: 10.1016/j.ijcard.2016.07.130
- Liao Y, Xu B, Wang J, Liu X. A new method for assessing the risk of infectious disease outbreak. *Sci Rep.* (2017) 7:40084. doi: 10.1038/srep40084
- Zhang Z, Zhang J, Wei Z, Ren H, Song W, Pan J, et al. Application of tabu search-based Bayesian networks in exploring related factors of liver cirrhosis complicated with hepatic encephalopathy and disease identification. *Sci Rep.* (2019) 9(1):6251. doi: 10.1038/s41598-019-42791-w
- Dellinger RP. Inflammation and coagulation: implications for the septic patient. *Clin Infect Dis.* (2003) 36(10):1259–65. doi: 10.1086/374835
- Libby P, Simon DI. Inflammation and thrombosis: the clot thickens. *Circulation.* (2001) 103(13):1718–20. doi: 10.1161/01.CIR.103.13.1718
- Corrales-Medina VF, Madjid M, Musher DM. Role of acute infection in triggering acute coronary syndromes. *Lancet Infect Dis.* (2010) 10(2):83–92. doi: 10.1016/S1473-3099(09)70331-7
- Cabell CH, Pond KK, Peterson GE, Durack DT, Corey GR, Anderson DJ, et al. The risk of stroke and death in patients with aortic and mitral valve endocarditis. *Am Heart J.* (2001) 142(1):75–80. doi: 10.1067/mhj.2001.115790
- Thuny F, Di Salvo G, Belliard O, Avierinos JF, Pergola V, Rosenberg V, et al. Risk of embolism and death in infective endocarditis: prognostic value of echocardiography: a prospective multicenter study. *Circulation.* (2005) 112(1):69–75. doi: 10.1161/CIRCULATIONAHA.104.493155

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

29. Di Salvo G, Habib G, Pergola V, Avierinos JF, Philip E, Casalta JP, et al. Echocardiography predicts embolic events in infective endocarditis. *J Am Coll Cardiol.* (2001) 37(4):1069–76. doi: 10.1016/S0735-1097(00)01206-7
30. Vilacosta I, Graupner C, San Román JA, Sarriá C, Ronderos R, Fernández C, et al. Risk of embolization after institution of antibiotic therapy for infective endocarditis. *J Am Coll Cardiol.* (2002) 39(9):1489–95. doi: 10.1016/S0735-1097(02)01790-4
31. Okonta KE, Adamu YB. What size of vegetation is an indication for surgery in endocarditis? *Interact Cardiovasc Thorac Surg.* (2012) 15(6):1052–6. doi: 10.1093/icvts/ivs365
32. Wallace SM, Walton BI, Kharbanda RK, Hardy R, Wilson AP, Swanton RH. Mortality from infective endocarditis: clinical predictors of outcome. *Heart.* (2002) 88(1):53–60. doi: 10.1136/heart.88.1.53
33. Lalani T, Chu VH, Park LP, Cecchi E, Corey GR, Durante-Mangoni E, et al. In-hospital and 1-year mortality in patients undergoing early surgery for prosthetic valve endocarditis. *JAMA Intern Med.* (2013) 173(16):1495–504. doi: 10.1001/jamainternmed.2013.8203
34. Ali N, Baig W, Wu J, Blackman D, Gillott R, Sandoe J. 129 Prosthetic valve endocarditis following transcatheter aortic valve implantation – experience from a UK centre. *BMJ.* (2019) 105(Suppl 6):A105–6. doi: 10.1136/heartjnl-2019-bcs.126
35. Wang A, Pappas P, Anstrom KJ, Abrutyn E, Fowler VG, Hoen B, et al. International collaboration on endocarditis investigators. The use and effect of surgical therapy for prosthetic valve infective endocarditis: a propensity analysis of a multicenter, international cohort. *Am Heart J.* (2005) 150(5):1086–91. doi: 10.1016/j.ahj.2005.01.023
36. Ferrera C, Vilacosta I, Fernández C, López J, Sarriá C, Olmos C, et al. Usefulness of new-onset atrial fibrillation, as a strong predictor of heart failure and death in patients with native left-sided infective endocarditis. *Am J Cardiol.* (2016) 117(3):427–33. doi: 10.1016/j.amjcard.2015.11.012
37. Baddour LM, Wilson WR, Bayer AS, Fowler VG Jr, Tleyjeh IM, Rybak MJ, et al. Infective endocarditis in adults: diagnosis, antimicrobial therapy, and management of complications: a scientific statement for healthcare professionals from the American heart association. *Circulation.* (2015) 132(15):1435–86. doi: 10.1161/CIR.0000000000000296