

Enhanced Risk Stratification in Infective Endocarditis Surgery: A Comprehensive External Validation of All Available Mortality Prediction Scores

Elisa Mikus^{1,*}, Diego Sangiorgi^{1,*}, Simone Calvi¹, Mariafrancesca Fiorentino¹, Elena Tenti¹, Flavia Dalle Mura², Carlo Savini^{1,2}

¹Cardiovascular Department, Maria Cecilia Hospital, GVM Care & Research, Cotignola, Italy; ²Department of Experimental Diagnostic and Surgical Medicine (DIMEC), University of Bologna, Bologna, Italy

*These authors contributed equally to this work

Correspondence: Diego Sangiorgi, Cardiovascular Department, Maria Cecilia Hospital, GVM Care & Research, Cotignola, Ravenna, 48033, Italy, Tel +39-0545-217579, Email dsangiorgi@gvmnet.it

Background: Several prognostic models have been developed to estimate operative mortality in patients undergoing surgery for infective endocarditis (IE). However, their external validity and performance remain uncertain, limiting clinical applicability. This study aimed to externally validate and compare the performance of contemporary IE-specific and generic cardiac surgery (EuroSCORE II) risk scores in a large single-center cohort.

Methods: Eighteen operative IE-specific risk scores, along with EuroSCORE II, were retrospectively applied to a database of 689 patients undergoing cardiac surgery for IE. Discrimination was evaluated using the area under the receiver operating characteristic curve (AUC), while calibration was assessed using the Hosmer–Lemeshow test, Brier score, and calibration slopes/intercepts. For each score, the type of validation reported in the original study was critically examined, noting that validation was not always performed. Additionally, the inclusion of IE-specific variables, such as pathogen type and valvular complications, was assessed to evaluate the reliability and clinical applicability of each score.

Results: Among the 689 patients, 30% were female, with a median age of 69 years. The most frequent pathogens were *Streptococcus* (26%), *Staphylococcus aureus* (18%), coagulase-negative staphylococci (18%), and *Enterococcus faecalis* (16%). Operative mortality was 10.6% (n = 73). The RISK-E score showed the highest discrimination (AUC: 0.742), followed by APORTEI (0.734) and modified MELD-XI (0.730). All scores demonstrated good calibration, with scaled Brier scores above 0.8. Scores incorporating IE-specific variables generally performed better, while several widely used generic scores, including EuroSCORE II, overestimated operative risk. External validation revealed lower AUCs for many scores compared to original reports, highlighting the importance of rigorous evaluation.

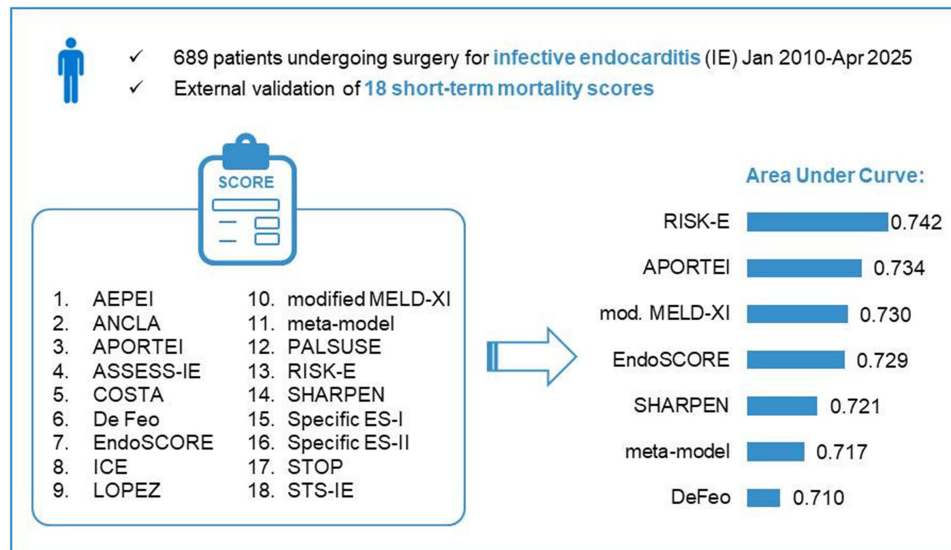
Conclusion: The RISK-E score demonstrated the highest discriminative ability and satisfactory calibration for predicting operative mortality in patients undergoing surgery for infective endocarditis. These results support the role of externally validated, IE-specific prognostic tools in guiding clinical assessment and selecting appropriate perioperative strategies.

Keywords: infective endocarditis, risk score, mortality, external validation

Introduction

Infective endocarditis (IE) remains a severe, life-threatening disease with considerable heterogeneity in its clinical presentation, course, and outcomes. It continues to pose a significant challenge to modern cardiovascular medicine despite major advances in diagnostic imaging, microbiological techniques, antimicrobial therapy, and surgical interventions.¹

Graphical Abstract



Epidemiological data indicate a progressive increase in the incidence of IE worldwide.² This trend has been primarily attributed to several factors: the rising number of patients with prosthetic heart valves and cardiovascular implantable electronic devices; the growing prevalence of invasive procedures in high-risk populations; and the aging of the population, with an associated increase in multimorbidity and frailty.

Although early diagnosis and the development of targeted antimicrobial regimens have improved patient outcomes, a substantial proportion of cases ultimately requires surgical intervention. According to large multicenter registries (the ENDO registry), surgery is indicated in up to 50% of patients with IE, primarily for management of severe valvular dysfunction, uncontrolled infection, or prevention of embolic events.³

Given these challenges, current international guidelines strongly recommend that therapeutic strategies for IE be determined by a dedicated multidisciplinary “endocarditis team”, typically comprising cardiologists, cardiac surgeons, infectious disease specialists, microbiologists, neurologists, and radiologists.⁴ Such an approach ensures that treatment decisions—including the timing and modality of surgical intervention—are individualized and evidence-based.

A crucial component of this decision-making process is the accurate assessment of operative risk. Reliable risk stratification is essential not only to guide surgical indications and optimize timing but also to facilitate patient counseling, shared decision-making, and benchmarking of institutional outcomes. Integrating clinical risk scores with individualized patient assessment represents a cornerstone of contemporary management, enabling physicians to balance the potential benefits of surgical intervention against the inherent risks in this highly vulnerable population. Inaccurate or poorly calibrated risk estimation may have direct clinical consequences: overestimation of operative risk can unjustifiably delay or even preclude surgery in patients who would otherwise benefit, while underestimation may lead to premature or overly aggressive surgical intervention in clinically unstable individuals. Moreover, discordant or unreliable risk predictions complicate discussions within the multidisciplinary endocarditis team and may undermine patient counseling, particularly when clinicians attempt to communicate individualized operative risk in the context of complex comorbid profiles.

A wide array of prognostic models have been developed to predict operative mortality in cardiac surgery, with some specifically designed for the IE population. Traditional cardiac surgery scores such as EuroSCORE I and II and the Society of Thoracic Surgeons (STS) score were not originally intended for IE and may not account for disease-specific variables such as pathogen type, embolic phenomena, or valve destruction. More recently, several IE-specific models

have emerged: AEPEI,⁵ ANCLA,⁶ APORTEI,⁷ ASSESS-IE,⁸ COSTA,⁹ De Feo,¹⁰ EndoSCORE,¹¹ ICE,¹² LOPEZ,¹³ modified MELD-XI,¹⁴ meta-model,¹⁵ PALSUSE,¹⁶ RISK-E,¹⁷ SHARPEN,¹⁸ Specific ES-I and Specific ES-II,¹⁹ STOP,²⁰ STS-IE,²¹ and one machine learning-based score proposed by Luo L.²²

Despite their promise, these models underwent limited external validation; moreover, most scores were developed without external validation or internal data partitioning (train/test split) and relying only on cross-validation of the overall sample, raising concerns about their generalizability and risk of overfitting.

This lack of robust external testing limits clinicians' confidence in applying these tools to real-world populations, where comorbidities, microbiological profiles, and disease severity may differ substantially from derivation cohorts. As a result, multidisciplinary teams are often forced to rely on subjective clinical judgment rather than data-driven predictions, increasing variability in practice patterns and potentially affecting the timing and appropriateness of surgical referral.

This study aims to assess and compare the external performance of 18 prognostic models for operative mortality in a real-world cohort of patients with surgically treated IE. Our objective is to provide insights into the models' utility in contemporary clinical practice and to identify the most accurate tools for risk stratification.

By clarifying which scores perform reliably in a contemporary external cohort, our findings aim to strengthen operative decision-making, improve consistency within endocarditis teams, and ultimately enhance patient counseling and surgical planning.

Methods

A retrospective analysis of 689 consecutive adult patients who underwent cardiac surgery for definite infective endocarditis between January 2010 and April 2025 was conducted. Diagnosis was established according to modified Duke criteria. Clinical, echocardiographic, microbiological, and surgical data were systematically extracted to compute each of the risk scores.

The study protocol received approval from the Romagna Ethics Committee on the 30th of June 2023 (protocol number 4497/2023 I.5/95). Since the study involved a retrospective analysis of anonymized data with no direct patient interaction or intervention, individual informed consent was waived. All patient data were handled in accordance with strict confidentiality standards, and the study was conducted in full compliance with the principles outlined in the Declaration of Helsinki. Data collection commenced from clinical charts and was systematically collected in a specific registry. Extensive efforts were made to minimize the occurrence of missing information. In instances where data were missing, it was due to the nature of gaps in clinical documentation and assumed to be missing completely at random.

The included models comprised both general cardiac surgery scores (EuroSCORE II) and IE-specific scores (AEPEI, ANCLA, APORTEI, ASSESS-IE, COSTA, De Feo, EndoSCORE, ICE, LOPEZ, modified MELD-XI, meta-model, PALSUSE, RISK-E, SHARPEN, Specific ES-I and Specific ES-II, STOP, STS-IE).

Statistical Analysis

After checking normal distribution with Shapiro–Wilk test, continuous variables were reported as mean and standard deviation (SD) or median and interquartile range (IQR); categorical variables were reported as absolute number and frequencies; the Absolute standardized mean differences (ASMD) were reported in order to quantify the magnitude of differences between groups. Scores were reconstructed based on their original definitions; missing data were imputed using random forest imputation for categorical variables and predictive mean matching for continuous variables. The imputation was performed with 100 trees and 5 iterations, using predictive mean matching with $k = 3$ to ensure realistic values for continuous variables.

Model performance was assessed by analyzing discrimination using the area under the receiver operating characteristic curve (AUC); the scaled Brier scores were used to assess overall accuracy of the IE scores; firstly, the null Brier score was calculated, which corresponds to the performance of a non-informative model that predicts the same probability for all individuals, equal to the observed death prevalence (p). The null Brier score was computed as $Brier_null = p*(1-p)^2 + (1-p)*p^2$. The scaled Brier scores, which quantify the relative improvement of the model over the null model, were calculated as $Brier_scaled = (1 - (Brier_model/Brier_null))*100\%$; a value of 100% indicates

perfect predictions, while 0% means the model is no better than predicting the prevalence. Calibration was assessed using the Hosmer–Lemeshow test; moreover, calibration intercept and slope were estimated by fitting a logistic regression of the observed outcome on the logit of the predicted probabilities. For some models where the predicted probabilities showed almost no variation, the calibration slope could not be estimated; an intercept equal to 0 reflects perfect average prediction, while above 0 the risk is underestimated and above 0 the risk is overestimated; a slope equal to 1 reflects perfect calibration, while below 1 means that predictions are too extreme (overfitting); conversely, a slope above 1 means that predictions are too conservative, and low and high risks are not well differentiated (underfitting).

All analysis was performed with R 4.5.0 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Among the 689 patients included, 30% were females, with a median age of 69 years (Table 1); more frequent pathogens were Streptococcus (26%), Staphylococcus aureus (18%), Coagulase-negative staphylococci (18%), Enterococcus faecalis (16%) (Table 2). The observed operative mortality was 10.6% (n = 73). Most frequent complications were abscess (28.4%), vegetations (81.7%), leaflet perforation (14.5%), prosthetic valve dehiscence (16.8%); most frequent involved valves were aortic valve (70.1%) and mitral valve (43.8%); previous cardiac surgery was performed on 46.2% of patients; median Log euroscore was 19.9, while patients in Urgency/emergency were 80.7% (Table 3).

Table 1 Scores Characteristics

Score	Number of Patients	IE Specific Variables	Validation	Reported AUC	Scoring System	Timepoint of Mortality Assessment
AEPEI	361	No	Internal (bootstrap) and external (161 pts from AEPEI registry)	0.780; 0.715 external	Yes	30-days
ANCLA	138	Yes (large intracardiac destruction)	Internal (bootstrap)	0.828	Yes	In-hospital
APORTEI	7,484 (Meta analysis, 16 studies)	Yes (abscess, pathogen)	External (1,338, GAMES registry)	0.750 (external validation)	Yes	30-days
ASSESS-IE	1,549	Yes (prosthetic IE, symptoms <1 month before admission)	Internal (bootstrap) and temporal Train (Jan09-Dec18, 1141)/Test (Jan19-Feb20, 408) split	0.781/0.779	Yes	6-months
COSTA	186	Yes (abscess, vegetation)	No	0.872	Yes	In-hospital
De Feo	440	Yes (positivity of latest preop. blood cultures, abscess)	No	0.91	Yes	30-days
EndoSCORE	2,715	Yes (abscess, pathogen)	Internal (bootstrap)	0.836; 0.851 (bootstrap)	No (only equation)	30-days
ICE	4,049	Yes (prosthetic IE, symptoms >1 month before admission, pathogen, abscess, persistent bacteremia)	Internal (bootstrap, ICE-PCS cohort 2000–2006) and temporal/external (1,197, ICE-PLUS cohort 2008–2012, a quote of centers overlap with ICE-PCS)	0.715; 0.682 for external validation	Yes	6-months

(Continued)

Table 1 (Continued).

Score	Number of Patients	IE Specific Variables	Validation	Reported AUC	Scoring System	Timepoint of Mortality Assessment
LOPEZ	317	Yes (pathogen, periannular complications)	Temporal (train 1996–2003 317; test 2004–2008, 263) and external (264)	0.67; 0.68 for temporal; 0.74 for external validation	Yes	In-hospital
Modified MELD-XI	858	No	No	0.823	No (only equation)	In-hospital
Meta-model	4,273 (3 selected scores)	Yes (abscess, fistulae, pathogen)	Aggregate of studies (EndoSCORE, Specific ES-I and ES-II)	0.790	No (only equation)	In-hospital
PALSUSE	437	Yes (prosthetic IE, large intracardiac destruction, pathogen)	No	0.84	Yes	In-hospital
RISK-E	671	Yes (prosthetic IE, pathogen, abscess)	Train (2/3 = 424)/ Test (1/3 = 247) split	0.77 train set; 0.82 test set	Yes	In-hospital
SHARPEN	233	No	No	0.86	Yes	In-hospital
Specific ES-I	775	Yes (fistulae, pathogen)	Internal (bootstrap)	0.786; 0.773 (bootstrap)	Yes	In-hospital
Specific ES-II	775	Yes (fistulae, pathogen)	Internal (bootstrap)	0.782; 0.770 (bootstrap)	Yes	In-hospital
STOP	1,181	Yes (active endocarditis, prosthetic IE)	Internal (5-fold cross-validation with bootstrap)	0.941	Yes	In-hospital
STS-IE	19,543	Yes (active endocarditis)	Train (70% = 13,617)/ Test (30% = 5,926) split	0.758 train set; not reported for test set	Yes	In-hospital
SYSUPMIE	476	Yes (vegetations)	Internal (bootstrap) and external (200)	0.810; 0.813 for external validation	None (black box, weights were not explicitly reported)	In-hospital

Overall, 56% of the scores had no validation or only internal validation using bootstrap; 17% (3 scores) divided train/test set for the development, while 22% (4 scores) used an external cohort validation; 3 scores (17%) did not involve Infective Endocarditis correlated variables; mortality was assessed in-hospital or at 30 days in 89% of cases; the remaining 11% studied mortality at 6 months (Figure 1).

Age and NYHA class were the most frequently included variables across the models (61% each). Of note, IE-specific predictors were represented in only a minority of scores: pathogen type or structural valve complications

Table 2 Baseline and Demographic Characteristics

	Overall	Alive	Dead	ASMD
n	689	616	73	
Age, median (IQR)	69 (58, 76)	68 (56, 75)	73 (68, 78)	0.664
Female, n (%)	204 (29.6)	175 (28.4)	29 (39.7)	0.241
BMI, median (IQR)	25.4 (23.1, 28.5)	25.3 (22.9, 28.4)	26.1 (24.2, 28.9)	0.140
NYHA (%)				0.524
1	121 (17.6)	111 (18.0)	10 (13.7)	
2	292 (42.4)	274 (44.5)	18 (24.7)	
3	191 (27.7)	163 (26.5)	28 (38.4)	
4	85 (12.3)	68 (11.0)	17 (23.3)	
Hypertension, n (%)	439 (63.7)	378 (61.4)	61 (83.6)	0.513
Diabetes, n (%)	143 (20.8)	117 (19.0)	26 (35.6)	0.380
Dialysis, n (%)	27 (3.9)	16 (2.6)	11 (15.1)	0.450
COPD (%)	75 (10.9)	64 (10.4)	11 (15.1)	0.141
Neurological disorder, n (%)	136 (19.7)	120 (19.5)	16 (21.9)	0.060
Pulmonary HT, n (%)	43 (6.2)	35 (5.7)	8 (11.0)	0.192
Chronic kidney disease, n (%)	89 (12.9)	66 (10.7)	23 (31.5)	0.527
Heart failure, n (%)	183 (26.6)	146 (23.7)	37 (50.7)	0.581
AF, n (%)	171 (24.8)	157 (25.5)	14 (19.2)	0.152
AV block II/III, n (%)	17 (2.5)	14 (2.3)	3 (4.1)	0.105
LVEF %, median (IQR)	58 (52, 63)	59 (54, 63)	55 (47, 60)	0.365
Drug addict, n (%)	43 (6.2)	40 (6.5)	3 (4.1)	0.107
CRP (mg/dL), median (IQR)	4.0 (1.1, 8.3)	3.7 (1.0, 7.8)	5.4 (2.1, 11.9)	0.310
Creatinine (mg/dL), median (IQR)	1.00 (0.84, 1.33)	1.00 (0.81, 1.25)	1.37 (1.00, 2.03)	0.467
Bilirubine (mg/dL), median (IQR)	0.47 (0.33, 0.71)	0.45 (0.32, 0.71)	0.52 (0.38, 0.84)	0.285
Emoglobine (mg/dL), median (IQR)	10.6 (9.6, 11.9)	10.6 (9.7, 12.0)	10.1 (9.1, 10.8)	0.489
Platelets (10 ⁹ /L), median (IQR)	216 (156, 274)	219 (158, 278)	185 (117, 247)	0.457
Systolic BP (mmHg), median (IQR)	120 (110, 130)	120 (110, 130)	120 (110, 130)	0.076

Abbreviations: AF, Atrial fibrillation; ASMD, absolute standardized mean difference; AV, atrioventricular; BMI, Body Mass Index; BP, Systolic blood pressure (mmHg); COPD, Chronic obstructive pulmonary disease; CRP, C-reactive protein; HT, hypertension; IQR, interquartile range; LVEF, Left ventricular ejection fraction; NYHA, New York Heart Association.

Table 3 Endocarditis and Operative Characteristics

	Overall	Alive	Dead	ASMD
n	689	616	73	
Negative blood culture, n (%)	81 (11.8)	78 (12.7)	3 (4.1)	0.312
Staphylococcus aureus, n (%)	126 (18.3)	108 (17.5)	18 (24.7)	0.175
Coagulase-negative staphylococci, n (%)	125 (18.1)	108 (17.5)	17 (23.3)	0.143
Streptococcus, n (%)	177 (25.7)	170 (27.6)	7 (9.6)	0.476
Enterococcus faecalis, n (%)	109 (15.8)	91 (14.8)	18 (24.7)	0.250
Pseudomonas, n (%)	4 (0.6)	4 (0.6)	0 (0.0)	0.114
Fungus, n (%)	10 (1.5)	10 (1.6)	0 (0.0)	0.182
Other pathogen, n (%)	79 (11.5)	70 (11.4)	9 (12.3)	0.030
Prosthetic endocarditis, n (%)	288 (41.8)	253 (41.1)	35 (47.9)	0.139
Abscess, n (%)	196 (28.4)	165 (26.8)	31 (42.5)	0.334
Prosthetic valve dehiscence, n (%)	116 (16.8)	97 (15.7)	19 (26.0)	0.255
Leaflet perforation, n (%)	100 (14.5)	92 (14.9)	8 (11.0)	0.119
Vegetations, n (%)	563 (81.7)	507 (82.3)	56 (76.7)	0.139
Active endocarditis, n (%)	598 (86.8)	532 (86.4)	66 (90.4)	0.127
Days from diagnosis to surgery, median (IQR)	20 (12, 35)	20 (12, 35)	21 (10, 30)	0.161
Urgency, n (%)	528 (76.6)	476 (77.3)	52 (71.2)	0.207

(Continued)

Table 3 (Continued).

	Overall	Alive	Dead	ASMD
Emergency, n (%)	28 (4.1)	22 (3.6)	6 (8.2)	0.207
Shock, n (%)	37 (5.4)	26 (4.2)	11 (15.1)	0.374
Intubation (pre-op), n (%)	49 (7.1)	35 (5.7)	14 (19.2)	0.418
IABP pre, n (%)	4 (0.6)	4 (0.6)	0 (0.0)	0.114
Log EuroSCORE, median (IQR)	19.9 (8.7, 40.7)	18.7 (7.7, 35.3)	42.4 (19.5, 65.1)	0.772
Previous cardiac surgery, n (%)	318 (46.2)	277 (45.0)	41 (56.2)	0.225
Previous CABG, n (%)	35 (5.1)	29 (4.7)	6 (8.2)	0.143
Previous valves (%)				0.289
0	391 (56.7)	358 (58.1)	33 (45.2)	
1	269 (39.0)	233 (37.8)	36 (49.3)	
2	27 (3.9)	24 (3.9)	3 (4.1)	
3	2 (0.3)	1 (0.2)	1 (1.4)	
N valves (%)				0.407
0	5 (0.7)	5 (0.8)	0 (0.0)	
1	524 (76.1)	477 (77.4)	47 (64.4)	
2	150 (21.8)	124 (20.1)	26 (35.6)	
3	10 (1.5)	10 (1.6)	0 (0.0)	
Aortic valve, n (%)	483 (70.1)	428 (69.5)	55 (75.3)	0.131
Mitral valve, n (%)	302 (43.8)	270 (43.8)	32 (43.8)	<0.001
Aortic surgery, n (%)	65 (9.4)	54 (8.8)	11 (15.1)	0.195

Abbreviations: ASMD, absolute standardized mean difference; CABG, coronary artery bypass grafting; IABP, Intra-aortic balloon pump; IE, infective endocarditis diagnosis to surgery; IQR, interquartile range.

were included in just 50% of the models (abscess formation in 50%, and prosthetic valve involvement in 33%) (Figure 2).

RISK-E demonstrated the highest discriminatory ability (AUC 0.742), followed by APORTEI (0.734) and modified MELD-XI (0.730) (Figure 3).

Brier score was below 0.02 for all scores, showing good accuracy; as a consequence, all scores achieved an 80% reduction in prediction error relative to the null mo scaled Brier scores resulted to be: Costa (87.7), Specific ES-II (87.1), STOP (87.1), ANCLA (86.9), Specific ES-I (86.4), ASSESS IE (85.7), AEPEI (84.7), PALSUSE (84.4), LOPEZ (83.9), ICE (83.5), STS-IE (82.7), De Feo (81.9), meta-model (81.8), EuroSCORE II (81.2), SHARPEN (80.8), APORTEI

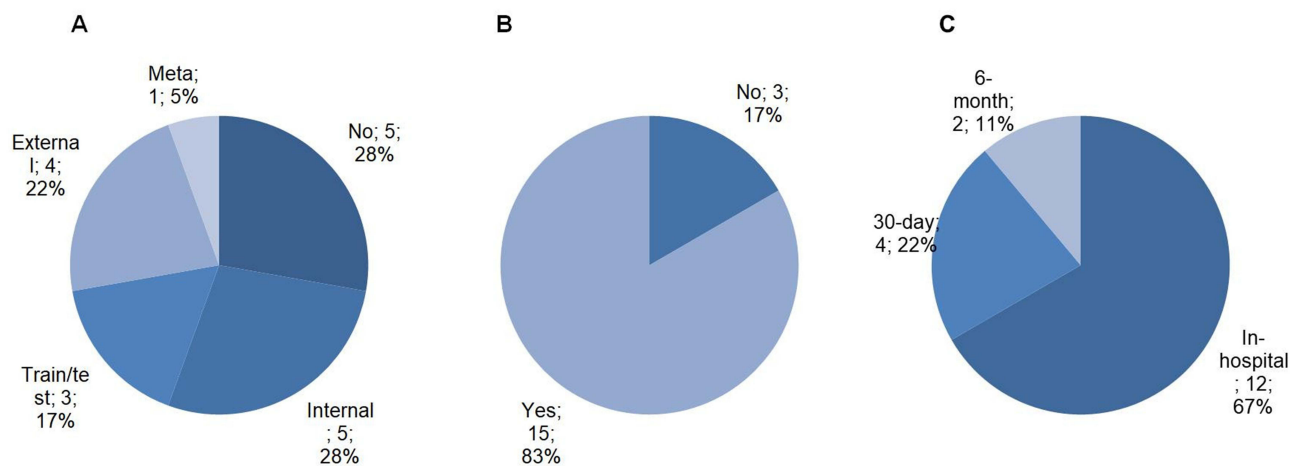


Figure 1 Scores distribution according to: panel (A) type of validation, panel (B) involvement of IE variables, panel (C) timepoint of mortality assessment.

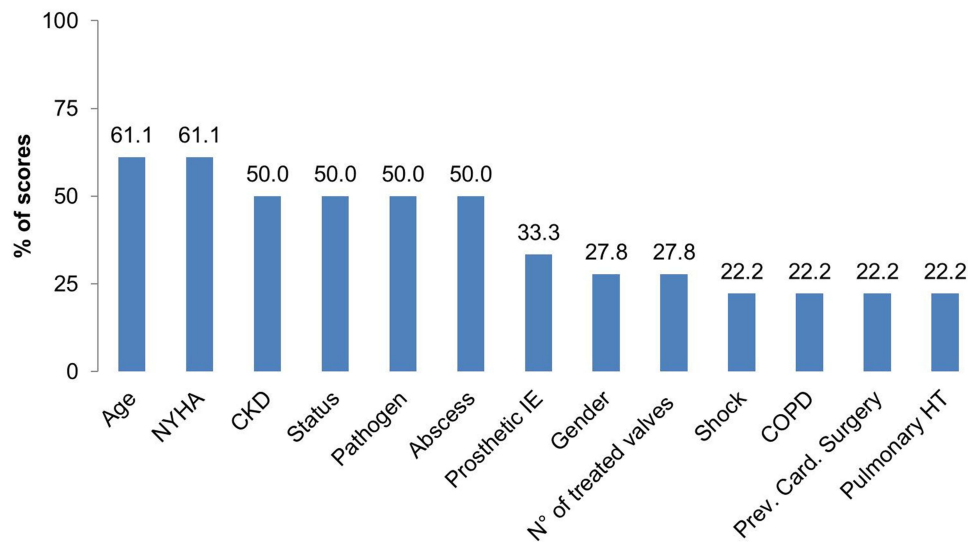


Figure 2 Clinical Features Most Frequently Used Across Endocarditis Risk Scores.

Abbreviations: CKD, Chronic kidney disease; COPD, Chronic obstructive pulmonary disease; HT, hypertension; IE, infective endocarditis; NYHA, New York Heart Association.

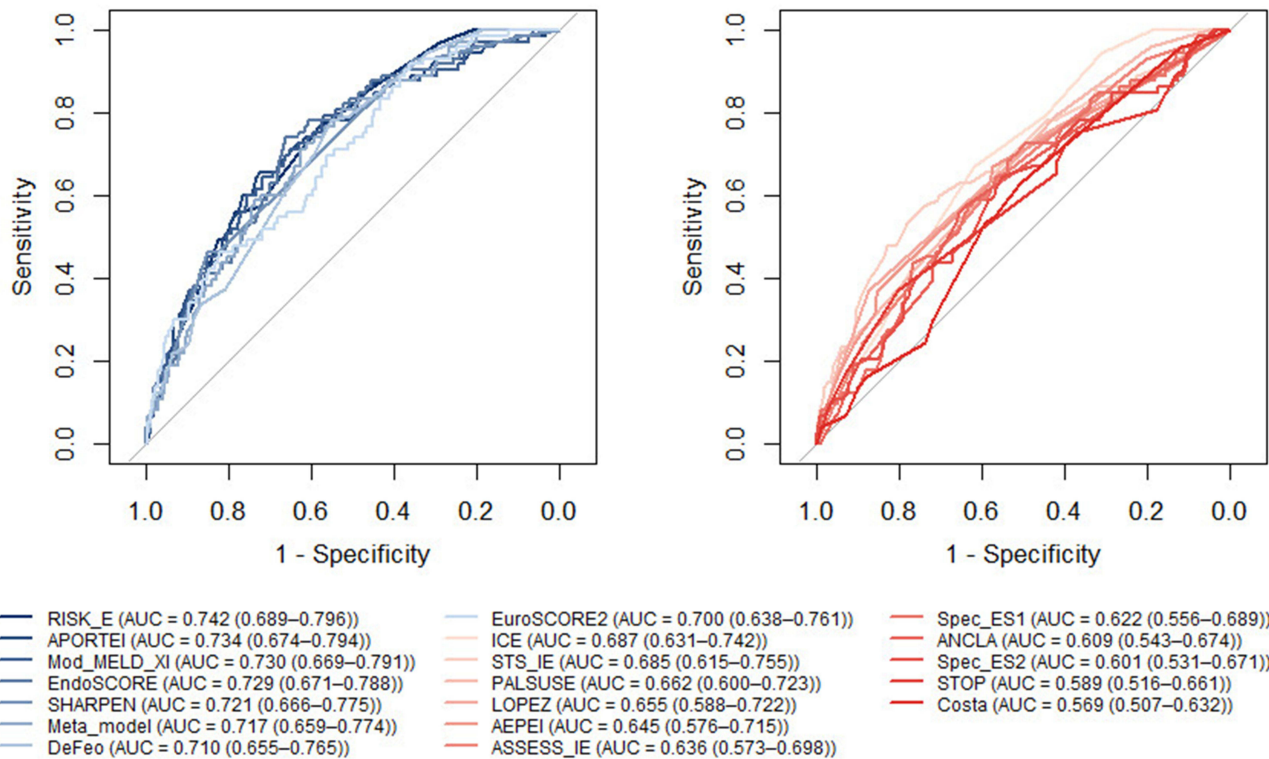


Figure 3 Scores ROC comparison: scores with AUC ≥0.7 (left) and scores with AUC <0.7 (right).

(80.1), modified MELD-XI (80.1), EndoSCORE (79.9), RISK-E (79.5). Hosmer–Lemeshow test showed bad calibration only for STOP.

Scores with slopes close to 1 (>0.7), indicating good calibration across the risk spectrum include: Specific ES-II, De Feo, PALSUSE, RISK-E, STOP, STS-IE, APORTEI; scores with slopes very low (near 0), indicating extreme overfitting, include: AEPEI, LOPEZ, modified MELD-XI, EndoScore, meta-mo scores for which the slope could not be estimated due to negligible variability in predicted probabilities include: ANCLA, ASSESS IE, Costa, ICE, SHARPEN, Specific ES-I; almost all scores showed intercepts close to 0, indicating good agreement between predicted and observed average

risk; scores with extreme negative intercepts, reflecting overestimation of risk, include: EuroSCORE II (−22.2) and APORTEI (−13.9).

Although machine learning-based scores have shown promising predictive performance, they were not included in this study due to their limited explainability, which may limit their clinical applicability; in particular, weights for each variable included in the scores were not reported in the original papers.

Discussion

In this large, single-center external validation study, the performance of 18 operative mortality risk scores in patients undergoing surgery for IE was systematically assessed. Our results confirm that IE-specific models, particularly the RISK-E,¹⁷ offer superior predictive performance compared to generic cardiac surgery scores. These findings underscore the importance of incorporating disease-specific variables in risk stratification tools for IE.

Our results align with previous validation efforts, including the multicenter work by Fernandez-Felix et al,¹⁵ which emphasized the benefit of aggregating models and optimizing them using large-scale registry data. Similarly, Agrawal et al highlighted the underutilization of existing models in clinical practice and the need for tailored tools.²³

Of note, the top-performing scores were based on heterogeneous sets of variables, with no consistent overlap, neither in terms of IE related variables; this lack of overlap was also observed among the lowest-performing scores, suggesting that neither strong nor weak predictive performance is necessarily linked to a specific subset of variables; more specifically, RISK-E,¹⁷ APORTEI⁷ and EndoSCORE¹¹ consider the type of pathogen and valvular complications, while modified MELD-XI¹⁴ and SHARPEN¹⁸ do not consider any IE specific aspects.

General scores such as EuroSCORE II, while widely used, lack of the granularity to capture the complexity of IE. Its tendency to overestimate risk may lead to undue therapeutic futility in high-risk patients. Conversely, well-calibrated scores like RISK-E¹⁷ can facilitate shared decision-making by providing more accurate risk estimates.

With the exception of RISK-E,¹⁷ which included a train/test split, and a few scores that reported external validation (eg, AEPEI,⁵ APORTEI,⁷ ICE¹²), the majority of models were developed on single, undivided datasets without validation. Despite high AUCs reported in some cases (eg, STOP:²⁰ 0.941; De Feo¹⁰ 0.910), these values likely reflect methodological biases, such as overfitting and lack of external testing, rather than true predictive performance.

When recalculated on our external cohort, most scores showed substantially lower AUCs, highlighting their limited generalizability and the importance of rigorous validation procedures in prognostic model development.

It should be emphasized that several prognostic models applied in cardiovascular surgery and critical care (such as AEPEI,⁵ the modified MELD-XI,¹⁴ and SHARPEN¹⁸) do not incorporate variables that are specific to infective endocarditis. While the absence of disease-oriented parameters does not necessarily result in inferior statistical performance, it inevitably reduces their specificity for IE and, consequently, their potential clinical applicability in this complex and heterogeneous condition.

Another important methodological consideration concerns the validation process. For example, the ICE¹² and ASSESS⁸ scores, which were originally derived by splitting training and test cohorts across different time periods, demonstrated unsatisfactory performance in our dataset. This discrepancy most likely reflects the influence of endogenous temporal factors—such as evolving patient demographics, shifting diagnostic strategies, and advances in surgical or antimicrobial therapy—that were not adequately captured by the models. Such findings highlight the inherent limitations of risk scores when applied outside the temporal or institutional context in which they were developed.

Similarly, the STS-IE score²¹ did not report a test-set area under the curve (AUC) in the original publication and also showed suboptimal discriminative ability in our external evaluation. Taken together, these results underscore the need for caution when extrapolating the reported performance of existing risk scores to independent populations. They also highlight the critical importance of robust external validation in determining the true clinical utility and generalizability of such prognostic tools.

Although several publications and narrative reviews^{24–26} have summarized the development of IE-related prognostic scores, and some models have undergone partial validation in selected cohorts, to the best of our knowledge, the present work represents the first comprehensive external validation of all 18 prognostic scores currently available for infective endocarditis.

A major strength of our study is its broad scope: in contrast to prior investigations that focused only on a limited subset of IE risk models, we systematically evaluated the entire spectrum of published scores within a large, independent external cohort. This comprehensive approach, not only permits a balanced comparison of predictive performance, but also provides a more nuanced

and clinically meaningful assessment of their potential role in guiding decision-making and benchmarking outcomes in patients with IE.

Limitations

Although this study is retrospective and based on a single-center cohort, this characteristic—often viewed as a limitation—does not undermine the validity of our work. In the context of external validation, using data from an institution different from those in which the original scores were developed is an advantage rather than a weakness, as it provides a rigorous test of their true generalizability. Differences in clinical practice, microbiological profiles, and surgical thresholds between centers are expected and, in fact, essential to evaluate whether these widely used scores can be reliable across diverse real-world settings. Additionally, some variables required for model reconstruction were not uniformly available. Another important consideration is that our analysis focused exclusively on operative, short-term mortality, in line with the original works. Long-term outcomes—including recurrent infection, heart failure, late valve-related events, and overall survival—were not assessed. For a disease as complex as infective endocarditis, these longer-term trajectories are highly relevant to patient counseling and therapeutic decision-making, and our results should therefore be interpreted as pertaining only to the immediate perioperative period.

Finally, the study period spans more than a decade, during which diagnostic modalities, antimicrobial strategies, and perioperative management evolved considerably. These temporal changes may have affected the stability of associations between predictors and outcomes and could partly explain the variability in performance of historically derived models. Although such evolution reflects real-world practice and is unavoidable in longitudinal cohorts, it remains a potential source of bias that should be considered when extrapolating our findings to contemporary or future settings.

Conclusions

In this large external validation study, we comprehensively assessed the performance of contemporary IE-specific and general cardiac surgery risk scores in patients undergoing surgery for infective endocarditis. Our findings demonstrate substantial variability in discriminative ability and calibration across models, with IE-specific tools—most notably RISK-E—showing more consistent performance than generic scores. These results highlight the importance of rigorous external validation to determine the true clinical utility and generalizability of prognostic models for IE. Further prospective, multicenter studies are needed to confirm these findings and to support the refinement or recalibration of existing tools.

Data Sharing Statement

The data supporting the findings of this study are not publicly available, and no additional data will be shared.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Disclosure

The authors report no conflict of interest.

References

1. Baddour LM, Wilson WR, Bayer AS, 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):1435–1486. doi:10.1161/CIR.0000000000000296
2. Chen H, Zhan Y, Kimin zhang K, et al. The global, regional, and national burden and trends of infective endocarditis from 1990 to 2019: results from the global burden of disease study 2019. *Front Med*. 2022;9(9):774224. doi:10.3389/fmed.2022.774224
3. Chu VH, Park LP, Athan E, et al. Association between surgical indications, operative risk, and clinical outcome in infective endocarditis: a prospective study from the International collaboration on endocarditis. *Circulation*. 2015;131(2):131–140. doi:10.1161/CIRCULATIONAHA.114.012461

4. Delgado V, Marsan NA, de Waha S, et al. ESC scientific document group 2023 ESC guidelines for the management of endocarditis. *Eur Heart J*. 2023;44(39):3948–4042. doi:10.1093/eurheartj/ehad193
5. Gatti G, Perrotti A, Obadia JF, et al. Simple scoring system to predict in-hospital mortality after surgery for infective endocarditis. *J Am Heart Assoc*. 2017;6(7):e004806. doi:10.1161/JAHA.116.004806
6. Gatti G, Benussi B, Gripsi F, et al. A risk factor analysis for in-hospital mortality after surgery for infective endocarditis and a proposal of a new predictive scoring system. *Infection*. 2017;45(4):413–423. doi:10.1007/s15010-016-0977-9
7. Barca LV, Fernández-Felix BM, Elorza EN, et al. Prognostic assessment of valvular surgery in active infective endocarditis: multicentric nationwide validation of a new score developed from a meta-analysis. *Eur J Cardiothorac Surg*. 2020;57(4):724–731. doi:10.1093/ejcts/ezz328
8. Wei X, Ran P, Nong Y, et al. ASSESS-IE: a novel risk score for patients with infective endocarditis. *J Cardiovasc Transl Res*. 2024;17(3):695–704. doi:10.1007/s12265-023-10456-9
9. da Costa MAC, Wollmann DR, Campos ACL, et al. Risk index for death by infective endocarditis: a multivariate logistic model. *Rev Bras Cir Cardiovasc*. 2007;22(2):192–200. doi:10.1590/s0102-76382007000200007
10. De Feo M, Cotrufo M, Carozza A, et al. The need for a specific risk prediction system in native valve infective endocarditis surgery. *Scientific World J*. 2012;2012(2012):307571. doi:10.1100/2012/307571
11. Di Mauro M, Actis Dato GM, Barili F, et al. A predictive model for early mortality after surgical treatment of heart valve or prosthesis infective endocarditis. The EndoSCORE. *Int J Cardiol*. 2017;241:97–102. doi:10.1016/j.ijcard.2017.03.148
12. Park LP, Chu VH, Peterson G, et al. Validated risk score for predicting 6-month mortality in infective endocarditis. *J Am Heart Assoc*. 2016;5(4):e003016. doi:10.1161/JAHA.115.003016
13. López J, Fernández-Hidalgo N, Revilla A, et al. Internal and external validation of a model to predict adverse outcomes in patients with left-sided infective endocarditis. *Heart*. 2011;97(14):1138–1142. doi:10.1136/hrt.2010.200295
14. He PC, Wei XB, Luo SN, et al. Risk prediction in infective endocarditis by modified MELD-XI score. *Eur J Clin Microbiol Infect Dis*. 2018;37(7):1243–1250. doi:10.1007/s10096-018-3240-8
15. Fernandez-Felix BM, Barca LV, Garcia-Esquinas E, et al. Prognostic models for mortality after cardiac surgery in patients with infective endocarditis: a systematic review and aggregation of prediction models. *Clin Microbiol Infect*. 2021;27(10):1422–1430. doi:10.1016/j.cmi.2021.05.051
16. Martínez-Sellés M, Muñoz P, Arnáiz A, et al. Valve surgery in active infective endocarditis: a simple score to predict in-hospital prognosis. *Int J Cardiol*. 2014;175(1):133–137. doi:10.1016/j.ijcard.2014.04.266
17. Olmos C, Vilacosta I, Habib G, et al. Risk score for cardiac surgery in active left-sided infective endocarditis. *Heart*. 2017;103(18):1435–1442. doi:10.1136/heartjnl-2016-311093
18. Chee QZ, Tan YQB, Ngiam JN, et al. The SHARPEN clinical risk score predicts mortality in patients with infective endocarditis: an 11-year study. *Int J Cardiol*. 2015;15(191):273–276. doi:10.1016/j.ijcard.2015.04.236
19. Fernández-Hidalgo N, Ferreria-González I, Marsal JR, et al. A pragmatic approach for mortality prediction after surgery in infective endocarditis: optimizing and refining EuroSCORE. *Clin Microbiol Infect*. 2018;24(10):1102.e7–1102.e15. doi:10.1016/j.cmi.2018.01.019
20. Habertheuer A, Geirsson A, Gleason T, et al. STratification risk analysis in OPERative management (STOP score) for drug-induced endocarditis. *J Card Surg*. 2021;36(7):2442–2451. doi:10.1111/jocs.15570
21. Gaca JG, Sheng S, Daneshmand MA, et al. Outcomes for endocarditis surgery in North America: a simplified risk scoring system. *J Thorac Cardiovasc Surg*. 2011;141(1):98–106.e1–2. doi:10.1016/j.jtcvs.2010.09.016
22. Luo L, Huang SQ, Liu C, et al. Machine learning-based risk model for predicting early mortality after surgery for infective endocarditis. *J Am Heart Assoc*. 2022;11(11):e025433. doi:10.1161/JAHA.122.025433
23. Agrawal A, Arockiam AD, Jamil Y, et al. Contemporary risk models for infective endocarditis surgery: a narrative review. *Ther Adv Cardiovasc Dis*. 2023;17:1–18. doi:10.1177/17539447231193291
24. Rizzo V, Salmasi MY, Sabetai M, et al. Infective endocarditis: do we have an effective risk score model? A systematic review. *Front Cardiovasc Med*. 20(10):1093363.
25. Cho CSY, Lim K, Siu ICH, et al. Infective endocarditis risk scores: a narrative review. *J Thorac Dis*. 2025;17(4):2662–2678. doi:10.21037/jtd-2024-2041
26. El Hatimi SE, Rpeiding ML, Selton-Suty C, et al. Predictive performance of surgical mortality risk scores in infective endocarditis. *Eur J Cardiothorac Surg*. 2024;65(1):ezad433. doi:10.1093/ejcts/ezad433

Clinical Epidemiology

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

Dovepress
Taylor & Francis Group