Clinical Microbiology and Infection 27 (2021) 1422–1430

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Systematic review

Prognostic models for mortality after cardiac surgery in patients with infective endocarditis: a systematic review and aggregation of prediction models

Borja M. Fernandez-Felix [1, 2, *], Laura Varela Barca [3], Esther Garcia-Esquinas [2, 4, 5], Andrea Correa-Pérez [1, 6], Nuria Fernández-Hidalgo [7, 8], Alfonso Muriel [1, 2], Jesus Lopez-Alcalde [1, 2, 6, 9], Noelia Álvarez-Diaz [10], Jose I. Pijoan [2, 11, 12], Aida Ribera [2, 13], Enrique Navas Elorza [14], Patricia Muñoz [15], María del Carmen Fariñas [16], Miguel Ángel Goenaga [17], Javier Zamora [1, 2, 18]

[1] *Clinical Biostatistics Unit, Hospital Universitario Ramon y Cajal (IRYCIS), Madrid, Spain*
[2] *CIBER Epidemiology and Public Health (CIBERESP), Madrid, Spain*
[3] *Department of Cardiovascular Surgery, Fundacion Jimenez Diaz University Hospital, Madrid, Spain*
[4] *Department of Preventive Medicine and Public Health, School of Medicine, Universidad Autónoma de Madrid, Madrid, Spain*
[5] *IdiPaz (Hospital Universitario La Paz-Universidad Autónoma de Madrid), Madrid, Spain*
[6] *Faculty of Medicine, Universidad Francisco de Vitoria, Madrid, Spain*
[7] *Servei de Malalties Infeccioses, Hospital Universitari Vall d'Hebron, Barcelona, Spain*
[8] *Red Española de Investigación en Patología Infecciosa (REIPI), Instituto de Salud Carlos III, Madrid, Spain*
[9] *Institute for Complementary and Integrative Medicine, University Hospital Zurich and University of Zurich, Zurich, Switzerland*
[10] *Medical Library, Hospital Universitario Ramon y Cajal (IRYCIS), Madrid, Madrid, Spain*
[11] *Hospital Universitario Cruces/OSI EEC, Barakaldo, Spain*
[12] *Biocruces-Bizkaia Health Research Institute, Barakaldo, Spain*
[13] *Cardiovascular Epidemiology and Research Unit, Hospital Universitari Vall d'Hebron, Barcelona, Spain*
[14] *Department of Infectology, Hospital Universitario Ramon y Cajal (IRYCIS), Madrid, Spain*
[15] *Clinical Microbiology and Infectious Diseases Service, Hospital General Universitario Gregorio Marañón, Instituto de Investigación Sanitaria Gregorio Marañón, CIBER Enfermedades Respiratorias-CIBERES, Facultad de Medicina, Universidad Complutense de Madrid, Madrid, Spain*
[16] *Infectious Diseases Service, Hospital Universitario Marqués de Valdecilla-IDIVAL, Universidad de Cantabria, Santander, Spain*
[17] *Infectious Diseases Service, Hospital Universitario Donostia, IIS Biodonostia, OSI Donostialdea, San Sebastián, Spain*
[18] *WHO Collaborating Centre for Global Women's Health, Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, UK*

ARTICLE INFO

Article history:
Received 4 April 2021
Received in revised form
19 May 2021
Accepted 31 May 2021
Available online 2 July 2021

Editor: L. Scudeller

Keywords:
Aggregation
Infective endocarditis
Meta-model
Prognostic models
Systematic review
Validation

ABSTRACT

Background: There are several prognostic models to estimate the risk of mortality after surgery for active infective endocarditis (IE). However, these models incorporate different predictors and their performance is uncertain.
Objective: We systematically reviewed and critically appraised all available prediction models of postoperative mortality in patients undergoing surgery for IE, and aggregated them into a meta-model.
Data sources: We searched Medline and EMBASE databases from inception to June 2020.
Study eligibility criteria: We included studies that developed or updated a prognostic model of postoperative mortality in patient with IE.
Methods: We assessed the risk of bias of the models using PROBAST (Prediction model Risk Of Bias ASsessment Tool) and we aggregated them into an aggregate meta-model based on stacked regressions and optimized it for a nationwide registry of IE patients. The meta-model performance was assessed using bootstrap validation methods and adjusted for optimism.
Results: We identified 11 prognostic models for postoperative mortality. Eight models had a high risk of bias. The meta-model included weighted predictors from the remaining three models (EndoSCORE, specific ES-I and specific ES-II), which were not rated as high risk of bias and provided full model equations. Additionally, two variables (age and infectious agent) that had been modelled differently across studies, were estimated based on the nationwide registry. The performance of the meta-model

* Corresponding author: Borja M. Fernandez-Felix, Hospital Universitario Ramón y Cajal, Ctra de Colmenar Viejo, Km,9,100, 28034, Madrid, Spain.
E-mail address: borjamanuel86@gmail.com (B.M. Fernandez-Felix).

https://doi.org/10.1016/j.cmi.2021.05.051

was better than the original three models, with the corresponding performance measures: C-statistics 0.79 (95% CI 0.76–0.82), calibration slope 0.98 (95% CI 0.86–1.13) and calibration-in-the-large −0.05 (95% CI −0.20 to 0.11).
Conclusions: The meta-model outperformed published models and showed a robust predictive capacity for predicting the individualized risk of postoperative mortality in patients with IE.
Protocol registration: PROSPERO (registration number CRD42020192602). **Borja M. Fernandez-Felix, Clin Microbiol Infect 2021;27:1422**

Introduction

Infective endocarditis (IE) is an uncommon but severe disease with a high mortality rate. Its current estimated incidence is three to ten episodes per 100 000 person-years, and its in-hospital mortality rate ranges between 15% and 40% [1,2]. Management of IE is often complex and, the decision whether to perform surgery remains a challenge because of the high mortality rate associated with the procedure. For that reason, it is estimated than less than half of the patients with surgical indication finally undergo cardiac surgery [3]; which leads to a significantly decreased chance of survival [4]. In this context, there has been a great interest in modelling the prognosis of patients with IE to accurately estimate the risk of mortality in patients undergoing surgery for IE, and to help in the decision-making process.

Prognostic models are mathematical equations that relate multiple variables for a particular individual to the probability of postoperative mortality. In the last decade, several IE prognostic models using preoperative patient-related and IE-specific factors, have been proposed. Unfortunately, these models have not been implemented in guidelines or are rarely applied in clinical practice. The poor adoption of these models could be a consequence of a shared perception of their limited validity because they have usually been built in relatively small cohorts and they usually lack external validation. Consequently, researchers continue to develop new models using their own data without considering previous knowledge, which leads to a scenario with multiple prognostic models of dubious validity. Therefore, we aimed to systematically review and critically appraise all available prediction models for postoperative mortality after cardiac surgery in patients with IE. We also aimed to aggregate those models with low risk of bias into a meta-model based on stacked regressions.

Materials and methods

The protocol for this study was registered on PROSPERO (registration number CRD42020192602). We designed this systematic review according to the recent guidance [5,6], and reported its results following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [7] and TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) recommendations [8,9].

Literature search

We searched Medline through Ovid and Embase through Elsevier from inception to 1 June 2020. We conducted a literature search to identify all potential studies for inclusion, without any language or publication dates restriction. We used the methodological filter developed by Geersing et al. for prediction models research in MEDLINE [10], which was adapted for EMBASE. We added terms related to cardiac surgery and endocarditis. We further searched bibliographic references of included articles to identify other potential eligible studies. Complete search strings are shown in the Supplementary material (Appendix S1).

Eligibility criteria

We included original studies that developed prognostic models, with or without external validation, to predict the risk of postoperative mortality after cardiac surgery in patients with IE, as well as studies that updated previously published models. We accepted the authors' definition of postoperative mortality (either 30 days and/or in-hospital mortality), but excluded models that predicted mortality as part of a composite adverse outcome. Titles, abstracts and full texts were screened for eligibility in pairs by three reviewers independently (BMFF, LVB, ACP) using EPPI-Reviewer 4 [11] Discrepancies were resolved by consensus.

Data extraction

Data extraction of included articles was done by three reviewers independently (pairs from BMFF, LVB, ACP). Discrepancies were solved by consensus. Reviewers used a standardized data extraction form based on CHARMS (CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies) [6]. We extracted data on the following items: general information of the study, source of data, participants' characteristics, outcome definition and time of occurrence, candidate predictors and analysis methods (see Supplementary material, Appendix S2). When the completed model equation or relevant data were not provided, we contacted the corresponding authors to require this information.

Risk of bias assessment

We used a standardized form based on PROBAST (PRediction model risk of Bias ASsessment Tool) [12,13] to evaluate risk of bias (RoB) and applicability. We used the PROBAST definition of RoB. Concerns regarding the applicability of a primary study would arise when the population, predictors or outcomes of the study differed from those specified in our review question. RoB and applicability were assessed by two independent reviewers (pairs from BMFF, LVB, ACP). We evaluated the relevant items on the following domains: participants, predictors, outcome and analysis. Each domain was rated as a high, low or unclear RoB, and as providing high, low or unclear concerns regarding applicability. Any discrepancies were discussed between reviewers and resolved through discussion. The Supplementary material provides details on critical appraisal and applicability (Appendix S3).

GAMES registry

We used the nationwide GAMES (Grupo de Apoyo al Manejo de la Endocarditis infecciosa en España) [14] registry as the validation data set, to estimate existing models' weights for the meta-model development and its validation, and to externally validate the previously published models. Since January 2008, all consecutive episodes of IE in 34 Spanish hospitals were prospectively registered in GAMES using a standardized form. Regional and local ethics committees approved the study, and patients gave their informed consent in each centre. For the present study, we selected all the infective episodes (n = 1453) registered in the GAMES cohort involving adult patients (aged ≥18 years) who had undergone cardiac surgery with preoperative diagnosis of active IE. From these, 354 (24.4%) died after surgery (273 in the first 30 days and the remaining 81 during hospitalization). Assessment of predictors was done in an unblinded manner (i.e. with knowledge of the participant's outcome). Table S1 (see Supplementary material) shows the main descriptive characteristic of patients in the validation nationwide registry.

Statistical analyses

Model aggregation was based on stacked regressions [15]. This methodology allows the synthesis of models collated in a systematic review into a meta-model using a validation data set [16,17]. We did not consider for aggregation the models that did not report the full equation or the models that were classified as high RoB. Stacked regressions used the linear predictor of each model as a covariable in the meta-model, to subsequently create a linear combination of model predictions. That is, the original coefficients of each model are weighted by an independent parameter estimated in the meta-model, so that the models with worse performance in the validation data set are penalized more. When aggregation of the coefficients was not possible, either because the definition of the predictor from primary studies was too heterogeneous or because predictors had been modelled differently in the published models (for instance, a numerical variable treated as a continuous predictor in one model and being categorized at different cut-points in the others), these predictors were dropped, and were included in the meta-model as independent covariables to re-estimate their coefficients entirely from scratch based on the validation data set. Non-linear relationships for continuous predictors were tested using fractional polynomials [18].

Predictors with missing data in the validation data set were imputed under the missing at random assumption using multiple imputation with chained equations [19]. We included all predictors and the outcome in the imputation models to ensure compatibility (see Supplementary material, Appendix S4). Imputations checks were completed by looking at the distributions of imputed values to ensure plausibility. We generated ten multiple imputed data sets and all primary analyses were performed in each imputed data set. Pooled parameters were estimated both in the aggregation and validation processes using Rubin's rules [20].

The meta-model validation was assessed in terms of discrimination (i.e. through the use of the C-statistic, with values from 1 indicating perfect discrimination to 0.5 no discrimination) and calibration (i.e. through the calibration slope and calibration-in-the-large (CITL), with 1 and 0 as ideal values, respectively; as well as with calibration plots). Calibration plots represent the average predicted probability for risk groups categorized using deciles of predicted probability against observed proportion in each group, and fitting a Lowess smoother to show calibration across the entire range of predicted probabilities at the individual-level [21,22]. For the calibration plots we used the average predicted probabilities for individuals by pooling the imputed data sets using Rubin's rules [20]. Because the meta-model was optimized to the validation data set, we assessed its optimism-corrected performance measures by applying bootstrap validation with 500 replicates. As sensitivity analyses, we tested all model performance regardless of their critical appraisal. In addition, the meta-model performance was assessed only for 30-day mortality to investigate the meta-model robustness. To facilitate the use of the model, an online version of the prognostic tool was implemented in Evidencio (https://www.evidencio.com/). All analyses were performed using Stata software version 16 [23] (see Supplementary material, Appendix S5).

Results

Search results and study selection

We retrieved 4862 titles through our systematic search combining Medline and Embase. From these, 684 duplicate references were identified. Of 4178 titles assessed by title and abstract, 34 studies were retained for full-text screening, and two additional studies were detected in the bibliographic references of these articles. Nine studies describing 11 prediction models met the inclusion criteria (Fig. 1 and see Supplementary material, Table S2).

Source of data and participants

All included prognostic model development studies were published between 2011 and 2018. Six used data from a study cohort (three of them from a single centre [24–26] and three from multiple centres [27–29]); two studies used data from multicentre registries [30,31]; and one study used data from both a multicentre cohort and a local clinical registry [32]. Eight studies used data from patients in Europe (Spain, Italy, France or Portugal) and one from patients in North America. Participants were recruited between 1980 and 2015 (see Supplementary material, Table S3).

Outcomes

Three models were developed to predict any death occurring before discharge or within 30 days of surgery [24,26,30], five models to predict any death occurring before discharge [25,29,31,32], and the remaining three as death within 30 days of surgery [27,28]. The incidence of deaths varied between 8.2% and 29.2% (Table 1).

Predictors

The number of candidate predictors considered in the models ranged from 15 to 57 and included patient-, clinical-, surgery- and IE-related factors. The number of parameters retained in the final models ranged from 2 to 15 (Table 1). The most common factors were critical preoperative state (n = 9), renal failure (n = 7), age (n = 6), New York Heart Association classification of functional status (n = 6), paravalvular complications (n = 6) and infection aetiology (n = 5). The predictor definitions and the models' composition are shown in the Supplementary material (Tables S4 and S5).

Model development and presentation

Sample sizes for model development varied between 128 and 13 617 patients, and the number of events ranged from 21 to 1117. Only two models from the same study adequately informed the handling of missing data [28], and these used complete data analyses. Logistic regression analysis was the most common modelling

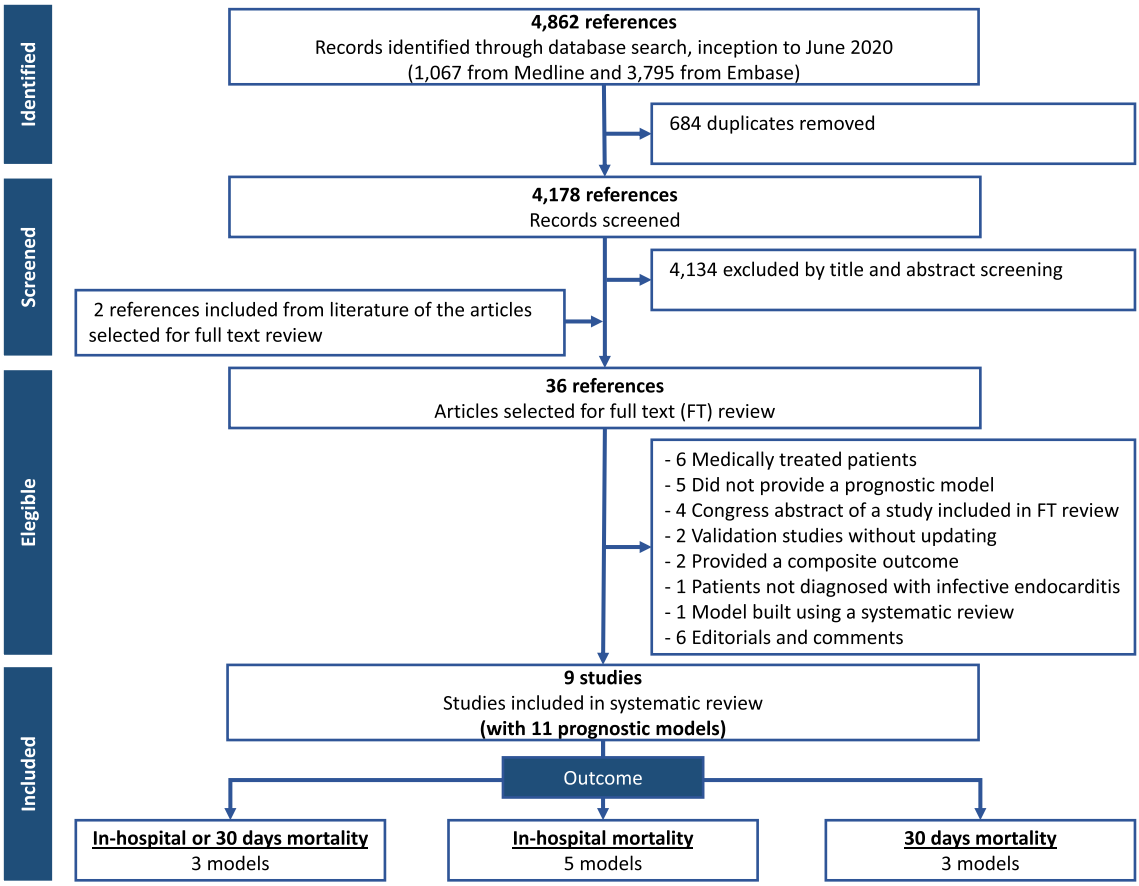

Fig. 1. PRISMA flowchart of study inclusions and exclusions.

technique ($n = 9$), while logistic mixed effects [27] and logistic Generalized Estimating Equation models [30] were only used in one model development each. Nine models used univariable analyses to select the candidate predictors. In nine out of eleven models the number of events per parameter assessed for inclusion in the final model was lower than the minimum required for development of a new prediction model, based on the sample size estimation proposed by Riley et al. [33,34] (see Supplementary material, Table S6). The method of predictor selection during multivariable modelling was backward selection in three models [25,32], stepwise selection in two models [29,31], and an automatic algorithm based on Akaike information criteria in multiple bootstrap samples in the other two models, with predictors selected in at least 70% of the bootstrapped samples being included in the final model [28]. Four models did not provide information about the method used to select predictors (Table 1).

In seven out of 11 models the authors omitted the complete model equation (in five of them corresponding authors did not respond when were asked for further details) (see Supplementary material, Table S7). Nine models were presented as a scoring system, and two of them included nomograms.

Model performance

The model performance was assessed in terms of discrimination through the C-statistic in all models. Nevertheless calibration was often wrongly assessed using the Hosmer–Lemeshow test [35] in six models. Only three models [26,28] used calibration slopes and CITL. Eight models were internally validated: three models were evaluated by bootstrapping with correction for optimism [27,28], one was assessed through the 0.632 bootstrap method [25], two used temporal split samples [32] and two used random split samples [29,30]. Three models only estimated the apparent performance [24,26,31]. Three models were externally validated in the same development study using very small sample sizes, with only 18 events in the Olmos et al. model [29] and 21 in the Gatti et al. models [32]. Clinical utility of the models was never assessed.

Risk of bias

The RoB was high in eight models, unclear in one [27] and low in the remaining two [28] (Table 1, see Supplementary material, Table S8 and Fig. S1). Two of the eight models with high RoB scored at 'high risk' in the participants domain. Eight models scored at 'high risk' in the analysis domain. Most of the models had small sample sizes and even the number of events per parameter was close to 1 in several models, increasing the risk of overfitting [34]. Many studies decided model predictors based on univariable analysis, three reported only the apparent performance and two used random splitting validation. The calibration was sub-optimally assessed in all models classified as high RoB, with most of them using the Hosmer–Lemeshow test.

Derivation of the meta-model

The eight models with high RoB were excluded from the statistical synthesis so that only the EndoScore, Specifics EuroSCORE-I

Table 1
Models' characteristics

Author, Year Model name	Modelling method	Sample size	Events n (%)	Predictors		EPCP/ EPFP	Selection of candidate predictors	Selection of final predictors	Type of validation	Performance measures		Critical appraisal			
				Cand.	Final							P	Pr	O	A
In-hospital or 30 days mortality															
De Feo, 2012 [24] De Feo score	Logistic regression	440	40 (9.1)	19	6	2.1/ 6.7	Univariable (p-value < 0.05)	n.a.	Int: Apparent Ext: n.a.	Disc: C = 0.88 (0.82;0.93) Cal: HL Test	RoB.	−	?	+	−
											App.	−	+	+	
Gaca, 2011 [30] STS Score	Logistic GEE regression	13,617	1,117 (8.2)	38	13	29.4/ 85.9	Univariable and previous STS model variables	n.a.	Int: Random Split (D:70%/V:30%) Ext: n.a.	Disc: C = 0.76 Cal: Calibration plot	RoB.	−	+	+	−
											App.	+	+	+	
Madeira 2016 [26] -	Logistic regression	128	21 (16.4)	15	2	1.4/ 10.5	Univariable	n.a.	Int: Apparent Ext: n.a.	Disc: C = 0.87 (0.79;0.94) Cal: Slope; CITL	RoB.	?	+	+	−
											App.	?	+	+	
In-hospital mortality															
Gatti 2017a [32] AEPEI score	Logistic regression	361	56 (15.5)	57	5	1.0/ 11.2	Univariable (p-value < 0.1)	Backward	Int: 0.632 Bootstrap Ext: (n=161; e=21)	Disc: C = 0.72 (0.64;0.78) Cal: HL Test	RoB.	+	+	+	−
											App.	+	?	+	
Gatti 2017a [32] Alternate AEPEI score	Logistic regression	361	56 (15.5)	57	3	1.0/ 11.2	Univariable (p-value < 0.1)	Backward	Int: 0.632 Bootstrap Ext: (n=161; e=21)	Disc: C = 0.69 (0.61;0.76) Cal: HL Test	RoB.	+	+	+	−
											App.	+	+	+	
Gatti 2017b [25] ANCLA score	Logistic regression	138	28 (20.3)	56	5	0.5/ 5.6	Univariable (p-value < 0.1)	Backward	Int: 0.632 Bootstrap Ext: n.a.	Disc: C = 0.83 (0.75;0.89) Cal: HL Test	RoB.	+	+	+	−
											App.	+	+	+	
Martínez-Sellés 2014 [31] PALSUSE	Logistic regression	437	106 (24.3)	n.a.	7	n.a./ 15.1	Univariable (p-value < 0.1)	Stepwise	Int: Apparent Ext: n.a.	Disc: C = 0.84 (0.79;0.88) Cal: HL Test	RoB.	+	+	+	−
											App.	+	+	+	
Olmos 2017 [29] RISK-E	Logistic regression	424	124 (29.2)	37	8	3.4/ 15.5	Univariable (p-value < 0.1) and clinically relevant	Stepwise	Int: Random Split (D:66%/V:33%) Ext: (n=204; e=18)	Disc: C = 0.76 (0.64;0.88) Cal: HL Test; Calibration plot	RoB.	+	+	+	−
											App.	+	+	+	
30 days mortality															
Di Mauro 2017 [27] EndoSCORE	Logistic mixed effect regression	2,715	298 (11.0)	32	15	9.3/ 19.9	Univariable (p-value < 0.2)	n.a.	Internal: Bootstrap External: n.a.	Disc: C = 0.85 (0.84;0.86) Cal: CITL and slope vs. the ideal values	RoB.	?	+	+	?
											App.	?	+	+	
Fernández-Hidalgo 2018 [28] Specific ES-I	Logistic regression	779	208 (26.7)	26	10	8.0/ 20.8	Variables in ES-I and specific IE risk factor	Bootstrap	Int: Bootstrap Ext: n.a.	Disc: C = 0.77 (0.74;0.81) Cal: Slope = 0.93 CITL = -0.06	RoB.	+	+	+	+
											App.	+	?	+	
Fernández-Hidalgo 2018 [28] Specific ES-II	Logistic regression	779	208 (26.7)	27	9	7.7/ 23.1	Variables in ES-II and specific IE risk factor	Bootstrap	Int: Bootstrap Ext: n.a.	Disc: C = 0.77 (0.73;0.81) Cal: Slope = 0.93 CITL = -0.05	RoB.	+	+	+	+
											App.	+	+	+	

STS, Society of Thoracic Surgeons; AEPEI, Association pour l'Etude et la Prevention de l'Endocadite Infectieuse; ANCLA, Anemia, NYHA class IV, critical state, large intracardiac destruction, and surgery on thoracic aorta; PALSUSE, prosthetic valve, age≥70, large intracardiac destruction, Staphylococcus spp, urgent surgery, sex [female], EuroSCORE≥10; RISK-E, Risk-Endocarditis; ES, EuroSCORE; GEE, Generalized Estimating Equation; n, number of events; Cand, number of candidate predictors assessed; EPCP, events per candidate predictor; EPFP, events per final predictor; Critical appraisal domains (P, participants; Pr, predictors; O, outcome; A, analysis); n.a., not available; Int, Internal validation (D, development cohort; V, validation cohort); Ext, external validation (*n*, sample size; e, number of events); Disc, Discrimination; Cal, calibration; HL, Hosmer-Lemeshow; CITL, calibration-in-the-large; RoB, Risk of Bias; App, applicability. +, Low RoB or low concern for applicability; −, High RoB or high concern for applicability; ?, Unclear RoB or applicability.

(Specific ES-I) and EuroSCORE-II (Specific ES-II) models were aggregated in the meta-model. The model developed by Di Mauro et al. (EndoSCORE) [27] included 15 parameters, whereas the other two (Specific ES-I and Specific ES-II) developed by Fernández-Hidalgo et al. [28], presented ten and nine parameters, respectively, from the EuroSCORE models predictors [36,37] and IE-related factors (Table 2 and see Supplementary material, Table S7). The dependent variable for the meta-model was mortality (either 30-day or in-hospital).

To construct the meta-model, we first calculated the linear predictors from EndoSCORE, Specific ES-I and Specific ES-II for each observation in the validation data set, after dropping the parameters for age and infection aetiology because these variables were modelled differently in the different studies. Subsequently, we adjusted the meta-model using a logistic regression model, which incorporated the linear predictors as co-variables, to estimate the models' weights for aggregation, as well as the predictors for age (treated as continuous) and infection aetiology (categorized into three groups: *Staphylococcus* spp., fungi and other microorganisms) to re-estimate the coefficients from scratch. The meta-model included the predictors considered in at least one of the three original models. These were patient-related factors (age, gender, renal failure, prior cardiac surgery, chronic pulmonary disease, pulmonary hypertension and left ventricular ejection fraction), clinical presentation-related factors (critical preoperative state, New York Heart Association classification of functional status), surgery-related factors (presence of paravalvular complications (abscess and/or fistulae), urgency of procedure and number of treated valves/prostheses) and finally IE-related factors (valve location and infection aetiology) (see Supplementary material, Table S5). We have developed an online calculator to allow a simple and effective use of the meta-model (https://www.evidencio.com/models/show/2498). The magnitude of the associations of the predictive factors with mortality is shown in Table 2 and the complete meta-model equation is given in the Supplementary material (Box S1).

Validation of the models

The three prediction models considered for aggregation and the meta-model were validated in the GAMES registry. The C-statistics and their 95% confidence intervals (95% CI) for the published models were: 0.759 (95% CI 0.731–0.788) for EndoSCORE, 0.758 (95% CI 0.731–0.786) for Specific ES-I and 0.762 (95% CI 0.735–0.789) for Specific ES-II. The optimism adjusted C-statistic for the meta-model was 0.79 (95% CI 0.76–0.82) (Fig. 2). Calibration slopes were <1 for all published models: 0.80 (95% CI 0.69–0.92) for EndoScore, 0.82 (95% CI 0.70–0.94) for Specific ES-I and 0.76 (95% CI 0.65–0.87) for Specific ES-II. CITL was 0.58 (95% CI 0.44–0.71) for EndoSCORE, 0.62 (95% CI 0.48–0.76) for Specific ES-II and −0.02 (95% CI −0.16 to 0.11) for Specific ES-I. Optimism adjusted calibration measures for the meta-model were 0.98 (95% CI 0.86–1.13) for the slope and −0.05 (95% CI −0.20 to 0.11) for CITL (Fig. 2). The calibration plots for the three previously published models and the meta-model are shown in Fig. 3.

Sensitivity analysis showed that the meta-model had better overall performance than all published models regardless of their quality assessment (see Supplementary material, Fig. S2). Moreover, even though the meta-model was not fitted for the 30-day

Table 2
Coefficients and odds ratios of the meta-model and the prediction models used for aggregation

Predictors	Original models			Aggregated model	
	EndoSCORE Di Mauro 2017	Sp. ES-I Fernández-Hidalgo 2018	Sp. ES-II Fernández-Hidalgo 2018	Meta-model[a]	
				Coefficient (95% CI)	OR (95% CI)
Intercept	−2.60	−3.13	−4.21	−5.00 (−5.97 to −4.00)	—
Gender (female)	0.51			0.22 (0.14–0.31)	1.25 (1.15–1.36)
Age[b] (years)	—	—	—	0.045 (0.03–0.06)	1.05 (1.03–1.06)
Renal failure	0.50	0.46		0.28 (0.17–0.41)	1.32 (1.19–1.51)
Prior cardiac surgery		1.10	0.96	0.51 (0.36–0.69)	1.67 (1.43–1.99)
Chronic pulmonary disease	0.68			0.29 (0.19–0.41)	1.34 (1.21–1.51)
Pulmonary hypertension		1.27		0.17 (−0.11 to 0.48)	1.19 (0.90–1.62)
LVEF (%)	−0.03			−0.013 (−0.02 to −0.01)	0.99 (0.98–0.99)
Critical preoperative state	1.46	1.12	1.02	1.17 (0.97–1.40)	3.22 (2.64–4.06)
NYHA class. (>I)		0.70	0.62	0.33 (0.23–0.44)	1.39 (1.26–1.55)
Abscess	1.09			0.47 (0.30–0.65)	1.60 (1.35–1.92)
Fistulae		1.22	1.14	0.59 (0.42–0.79)	1.80 (1.52–2.20)
Priority of procedure					
Urgent status			1.16	0.44 (0.16–0.68)	1.55 (1.17–1.97)
Emergency status		0.81	1.95	0.85 (0.53–1.17)	2.34 (1.70–3.22)
Number of valves treated					
Two valves treated	0.50			0.22 (0.14–0.30)	1.25 (1.15–1.35)
Three valves treated	1.50			0.65 (0.41–0.90)	1.92 (1.51–2.46)
Valve location (Mitral)		0.37	0.38	0.19 (0.14–0.25)	1.21 (1.15–1.28)
Aetiology[c]	—	—	—		
Staphylococcus spp.				0.64 (0.35–0.94)	1.90 (1.42–2.56)
Fungi				0.61 (−0.46 to 1.40)	1.84 (0.63–4.06)

Abbreviations: LVEF, left ventricular ejection fraction; NYHA class, New York Health Association classification of functional status.
Stacked regression: $\ln(\frac{p}{1-p}) = -1.861 + 0.433 \times LP^{\dagger}_{DM} + 0.131 \times LP^{\dagger}_{FH-I} + 0.379 \times LP^{\dagger}_{FH-II} + 0.045 \times Age + 0.64 \times$ *Staphylococcus* spp. $+ 0.61 \times$ Fungi
Where, p is the probability of postoperative mortality and $LP^{\dagger}_{i}$ is the linear predictor for each model selected for aggregation dropping the parameters from age and infection aetiology; DM (Di Mauro model [EndoSCORE]); FH-I (Fernández-Hidalgo model [sp. ES-I]); FH-II (Fernández-Hidalgo model [sp. ES-II]). Consequently, stacked intercept $= -1.861 + 0.433 \times (-2.60) + 0.131 \times (-3.13) + 0.379 \times (-4.21) = -5.00$, and for instance, the stacked coefficient for renal failure $= 0.433 \times (0.50) + 0.131 \times (0.46) + 0.379 \times 0) = 0.277$.
[a] Weights used to create the meta-model: EndoScore = 0.433; Sp. ES-I = 0.131; Sp. ES-II = 0.379.
[b] Age was categorized in Di Mauro et al. [27] and treated as continuous in Fernández-Hidalgo et al. [28].
[c] Aetiology was categorized in different ways in each existing model.

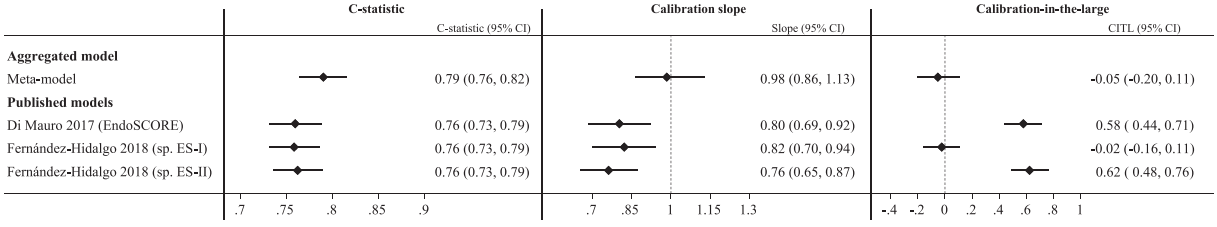

Fig. 2. Bootstrap internal validation of the meta-model and external validation of existing models selected for aggregation Dashed lines indicate lines of perfect calibration slope (1) and calibration-in-the-large (0). Black diamonds indicate point estimates and horizontal lines indicate 95% CIs. CITL: Calibration-in-the-large.

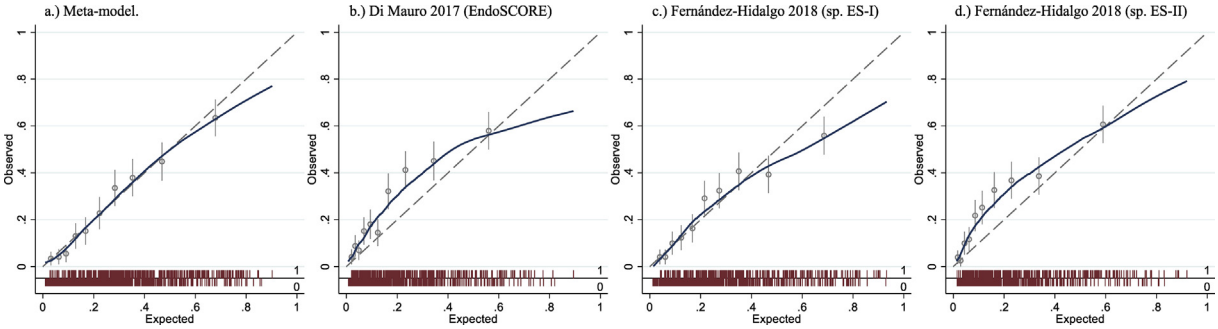

Fig. 3. Calibration plots of the meta-model and of the prediction models selected for aggregation. Dashed lines represent perfect calibration, grey circles and bars indicate average risks and their confidence interval by deciles of the risk spectrum, dark blue lines indicate the Lowess smoother assessment of the calibration at the individual level, and red spike plots show the distribution of events and non-events.

mortality outcome, it outperformed the three models used for model aggregation (see Supplementary material, Fig. S3).

Discussion

Summary of findings

In this systematic review of prediction models for postoperative mortality in patients with IE, we identified and critically appraised 11 models developed in nine studies. The predicted outcome varied between studies (in-hospital, 30-day or both in-hospital and 30-day mortality). Of the 11 prognostic models, only two had low RoB and one had unclear RoB; the remaining eight models had high RoB mainly owing to poor statistical methods used, which suggests that their predictive performance when used in practice is probably lower than that reported. The sample sizes used to develop the models were limited and this is a well-known problem that leads to inaccurate predictions and consequently incorrect health-care decisions in practice [34].

Four out of the 11 published models reported the full model equation required for a model's aggregation and a complete independent external validation as recommended by reporting guidelines [8,9]. Two models' equations were recovered after request to the corresponding authors. Three models that were flagged as low or unclear RoB were aggregated to build the meta-model. Our meta-model included as predictors age, gender, renal failure, prior cardiac surgery, chronic pulmonary disease, pulmonary hypertension, left ventricular ejection fraction, critical preoperative state, New York Heart Association classification of functional status presence of paravalvular complications (abscess and/or fistulae), urgency of procedure, number of treated valves/prostheses, valve location and infection aetiology. It showed better performance than the original models. We investigated the internal validity of the meta-model using bootstrap validation, and the results indicated there was no substantial over-optimism and that the validation sample was sufficiently large to combine and update the published models. Therefore, the meta-model is probably less prone to over-optimism and more generalizable to new patient populations or settings, because it was built from the evidence of several patient cohorts and optimized to a nationwide registry.

Strengths and limitations

To our knowledge, this is the first systematic review with specific focus on prediction models of postoperative mortality in patients with IE, with a thorough evaluation of the RoB, and using an external validation cohort to build a meta-model. We only combined the prediction models with low or unclear RoB and adjusted them to a new patient population. We used multiple imputation of predictors to avoid loss of useful information. The resulting meta-model incorporated previous knowledge optimally and outperformed previously published models.

Our study has some limitations. The outcome definition in the validation data set was either 30-day or in-hospital postoperative mortality, and the outcome definition in the three models used for aggregation was 30-day mortality. Despite this difference a sensitivity analysis showed that the meta-model outperformed all published models when we explored its performance for the 30-day mortality. Two out of the three models considered for aggregation were developed in the same cohort. This circumstance increases the probability that the same predictors were included in both models and, therefore, it could magnify their associations with the outcome in the meta-model. However, we think that the impact of this magnification is limited because the weight of the ES-I model is relatively small compared with the other two models. Unfortunately, although we identified 11 prediction models in our systematic review, we were only able to validate the models for which the complete model equation was available. All of these incomplete models were classified as high RoB and were consequently excluded from the analysis. We cannot rule out the presence of publication bias in our review. Unpublished studies are likely to be of poor quality (small, overfitted, and with poor predictive performance). Therefore, it is very likely that they would have been excluded from our meta-model due to their high RoB. So the impact of this bias is expected to be low. Although the definition of predictors in GAMES registry was standardized, these could differ from definitions of published studies.

Comparison to existing studies

Most studies to develop new prediction models are based on small sample sizes and the modelling strategies are excessively driven by available data without considering the previous knowledge, leading to inefficient models. Other authors carried out external validation studies but none of them made a critical appraisal [38–41]. In a previous study, Varela et al. developed a prognostic model based on a systematic review of factors related to in-hospital mortality. The model was built using a series of univariate meta-analyses that pooled adjusted and unadjusted estimates altogether without taking into consideration the correlation among these factors. These pooled univariate estimates were then transformed into risk points to create a risk score [42,43]. Our proposal includes more factors and our analysis included only estimates from low RoB studies. All estimates are from multivariate adjusted models and the weight each model has to build the meta-model is determined by their predictive performance in a validation cohort. This statistical methodology is in concordance with current recommendations [16,44].

Implications for practice

The decision whether to perform surgery for IE remains a challenge in clinical practice and it should come after a careful balance between the procedural risk and its estimated benefit. Critical preoperative state and priority of the procedure (urgent or emergency) are the most salient risk factors included in our meta-model. Patients with depressed left ventricukar ejection fraction, New York Heart Association classification or renal failure also have worse prognosis. In addition, the aggressiveness of the IE infection as well as the technical difficulties of the surgery also implied higher risk of mortality. We expect a worse outcome in patients with IE caused by *Staphylococcus* spp. or fungi or in patients with paravalvular abscesses, fistulae or previous cardiac surgery because in these patients the surgery is challenging. Although risk scores for predicting mortality do not offer help in terms of establishing the burdens of surgical futility, they add great value by helping endocarditis teams to manage this complex disease and lead to more personalized assistance based on individual patient characteristics. Moreover, the meta-model can be used to determine the case-mix of surgical hospitals and compare their performance adjusted for their case-mix.

Although in the 2015 IE guidelines [45] the score created by De Feo et al. [24] for native IE is the only one recommended, it would be expected to change with the creation of several new IE-specific scores and the generation of a meta-model that outperformed existing models.

The explanatory interpretation of the meta-model coefficients should be made with caution because coefficients have been shrunk, and therefore could be affected by the Stein's paradox [46]. Shrinkage of the multivariable regression coefficients introduces a

bias towards the null, but at the same time, properly shrinking coefficients ensures better predictions [47].

Challenges and opportunities

Further external validation studies are necessary to confirm the improvement in predictive ability of the meta-model. We will develop an online calculator to allow a simple and effective use of the meta-model. Given the low incidence of IE, sufficiently large sample sizes for the adequate development of new predictive models are difficult to come by. We encourage authors to make their data available in order to allow building model based on available data [48,49].

Conclusions

The meta-model is a robust prognostic model to calculate the individualized risk of postoperative mortality in patients with IE. It was developed based on the previous evidence using aggregation methods of the existing models identified from a systematic review and after critical being appraised. The meta-model outperformed existing models; therefore, this preoperative tool can help guide individually tailored choices made by patients and clinicians.

Transparency declaration

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and the authors declare that they have no conflicts of interest.

CIBER (Biomedical Research Network in Epidemiology and Public Health) has partially supported the realization of this work (Grant number: ESP20G42X1). This public funding body had no role in the study design, the collection, analysis and interpretation of the data, the writing of the report nor the decision to submit the paper for publication.

Authors contributions

BMFF, LVB, EGE, JLA, AM, JIP, AR and JZ contributed to the conceptualization; search strategies were by BMFF, NAD and JLA; data extraction and critical appraisal were by BMFF, LVB and ACP; methodology was performed by BMFF, EGE, AM and JZ; software and formal analysis were the responsibility of BMFF; validation was by AM and JZ; data adquisition/curation were by BMFF, ENE, PM, MCF and MAG. The original draft was written by BMFF, EGE and JZ; visualization was by BMFF, LVB and NFH; and supervision was by EGE and JZ. All authors contributed to reviewing and editing the article.

Acknowledgements

We are grateful to the team involved with the collection of the GAMES registry data (see Supplementary material, Appendix S6).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cmi.2021.05.051.

References

[1] Murdoch DR. Clinical presentation, etiology, and outcome of infective endocarditis in the 21st century: the international collaboration on endocarditis–prospective cohort study. Arch Intern Med 2009;169:463.
[2] Thuny F, Grisoli D, Collart F, Habib G, Raoult D. Management of infective endocarditis: challenges and perspectives. Lancet 2012;379:965–75.
[3] Iung B, Doco-Lecompte T, Chocron S, Strady C, Delahaye F, Le Moing V, et al. Cardiac surgery during the acute phase of infective endocarditis: discrepancies between European Society of Cardiology guidelines and practices. Eur Heart J 2016;37:840–8.
[4] Chu VH, Park LP, Athan E, Delahaye F, Freiberger T, Lamas C, 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;131:131–40.
[5] Debray TPA, Damen JAAG, Snell KIE, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ 2017:i6460.
[6] Moons KGM, de Groot JAH, 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.
[7] Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6:e1000097.
[8] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015;350:g7594.
[9] Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1.
[10] Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search Filters for finding prognostic and diagnostic prediction studies in medline to enhance systematic reviews. PLoS ONE 2012;7:e32844.
[11] Thomas J, Graziosi S, Brunton J, Ghouze Z, O'Driscoll P, Bond M. EPPI-Reviewer: advanced software for systematic reviews, maps and evidence synthesis. EPPI-Centre Software. London: UCL Social Research Institute; 2020.
[12] Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med 2019;170:W1.
[13] 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.
[14] Muñoz P, Kestler M, De Alarcon A, Miro JM, Bermejo J, Rodríguez-Abella H, et al. Current epidemiology and outcome of infective endocarditis: a multicenter, prospective, cohort study. Medicine (Balt) 2015;94:e1816.
[15] Breiman L. Stacked regressions. Mach Learn 1996;24:49–64.
[16] Debray TPA, Koffijberg H, Nieboer D, Vergouwe Y, Steyerberg EW, Moons KGM. Meta-analysis and aggregation of multiple published prediction models: meta-analysis and aggregation of multiple published prediction models. Stat Med 2014;33:2341–62.
[17] Martin GP, Mamas MA, Peek N, Buchan I, Sperrin M. A multiple-model generalisation of updating clinical prediction models. Stat Med 2018;37:1343–58.
[18] Royston P, Sauerbrei W. Multivariable model-building: a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables. Chichester, England ; Hoboken, NJ: John Wiley; 2008. p. 303 (Wiley series in probability and statistics).
[19] White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med 2011;30:377–99.
[20] Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.
[21] Riley RD, van der Windt D, Croft P, Moons KGM. Prognosis research in healthcare: concepts, methods and impact. 2019.
[22] Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating [Internet]. 2019. https://doi.org/10.1007/978-3-030-16399-0 [cited 2020 Apr 28]. Available from:.
[23] StataCorp. Stata statistical software: Release 16. College Station, TX: StataCorp LLC; 2019.
[24] De Feo M, Cotrufo M, Carozza A, De Santo LS, Amendolara F, Giordano S, et al. The need for a specific risk prediction system in native valve infective endocarditis surgery. Sci World J 2012;2012:1–8.
[25] Gatti G, Benussi B, Gripshi F, Della Mattia A, Proclemer A, Cannatà A, 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;45:413–23.
[26] Madeira S, Rodrigues R, Tralhão A, Santos M, Almeida C, Marques M, et al. Assessment of perioperative mortality risk in patients with infective endocarditis undergoing cardiac surgery: performance of the EuroSCORE I and II logistic models. Interact CardioVasc Thorac Surg 2016;22:141–8.
[27] Di Mauro M, Dato GMA, Barili F, Gelsomino S, Santè P, Corte AD, 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.
[28] Fernández-Hidalgo N, Ferreria-González I, Marsal JR, Ribera A, Aznar ML, de Alarcón A, et al. A pragmatic approach for mortality prediction after surgery in infective endocarditis: optimizing and refining EuroSCORE. Clin Microbiol Infect 2018;24:1102.e7–1102.e15.
[29] Olmos C, Vilacosta I, Habib G, Maroto L, Fernández C, López J, et al. Risk score for cardiac surgery in active left-sided infective endocarditis. Heart 2017;103:1435–42.

[30] Gaca JG, Sheng S, Daneshmand MA, O'Brien S, Rankin JS, Brennan JM, et al. Outcomes for endocarditis surgery in North America: a simplified risk scoring system. J Thor Cardiovasc Surg 2011;141:98–106.e2.
[31] Martínez-Sellés M, Muñoz P, Arnáiz A, Moreno M, Gálvez J, Rodríguez-Roda J, et al. Valve surgery in active infective endocarditis: a simple score to predict in-hospital prognosis. Int J Cardiol 2014;175:133–7.
[32] Gatti G, Perrotti A, Obadia J, Duval X, Iung B, Alla F, et al. Simple scoring system to predict in-hospital mortality after surgery for infective endocarditis [Internet] JAHA 2017 [cited 2020 Dec 28];6(7). Available from: https://www.ahajournals.org/doi/10.1161/JAHA.116.004806.
[33] Riley RD, Snell KI, Ensor J, Burke DL, Harrell Jr FE, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med 2019;38:1276–96.
[34] Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ 2020:m441.
[35] On behalf of Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative, Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med 2019;17:230.
[36] Nashef SAM, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). Eur J Cardio-Thor Surg 1999;16:9–13.
[37] Nashef SAM, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, et al. EuroSCORE II. Eur J Cardio-Thor Surg 2012;41:734–45.
[38] Varela L, López-Menéndez J, Redondo A, Fajardo ER, Miguelena J, Centella T, et al. Mortality risk prediction in infective endocarditis surgery: reliability analysis of specific scores†. Eur J Cardio-Thor Surg 2018;53:1049–54.
[39] Pivatto Júnior F, de A Bellagamba CC, Pianca EG, Fernandes FS, Butzke M, Busato SB, et al. Análise de Escores de Risco para Predição de Mortalidade em Pacientes Submetidos à Cirurgia Cardíaca por Endocardite [Internet] ABC Cardiol 2020 [cited 2021 Mar 5]; Available from: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0066-782X2020000300518.
[40] Gatti G, Sponga S, Peghin M, Givone F, Ferrara V, Benussi B, et al. Risk scores and surgery for infective endocarditis: in search of a good predictive score. Scand Cardiovasc J 2019;53:117–24.
[41] Wang TKM, Oh T, Voss J, Gamble G, Kang N, Pemberton J. Comparison of contemporary risk scores for predicting outcomes after surgery for active infective endocarditis. Heart Vessels 2015;30:227–34.
[42] Varela Barca L, Navas Elorza E, Fernández-Hidalgo N, Moya Mur JL, Muriel García A, Fernández-Felix BM, et al. Prognostic factors of mortality after surgery in infective endocarditis: systematic review and meta-analysis. Infection 2019;47:879–95.
[43] Varela Barca L, Fernández-Felix BM, Navas Elorza E, Mestres CA, Muñoz P, Cuerpo-Caballero G, 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 Cardio-Thor Surg 2020;57:724–31.
[44] Debray TPA, Koffijberg H, Vergouwe Y, Moons KGM, Steyerberg EW. Aggregating published prediction models with individual participant data: a comparison of different approaches. Stat Med 2012;31:2697–712.
[45] Habib G, Lancellotti P, Iung B. 2015 ESC Guidelines on the management of infective endocarditis: a big step forward for an old disease. Heart 2016;102:992–4.
[46] Efron B, Morris C. Stein's Paradox in statistics. Sci Am 1977;236:119–27.
[47] van Houwelingen JC. Shrinkage and penalized likelihood as methods to improve predictive accuracy. Stat Neerland 2001;55:17–34.
[48] Debray TPA, Riley RD, Rovers MM, Reitsma JB. Moons KGM, Cochrane IPD Meta-analysis Methods group. Individual participant data (IPD) meta-analyses of diagnostic and prognostic modeling studies: guidance on their use. PLoS Med 2015;12:e1001886.
[49] Ensor J, Snell KIE, Debray TPA, Lambert PC, Look MP, Mamas MA, et al. Individual participant data meta-analysis for external validation, recalibration, and updating of a flexible parametric prognostic model. Stat Med 2021;40:3066–84.