Prognostic models for mortality after cardiac surgery in patients with infective endocarditis

Fernandez-Felix, Borja M. ; Barca, Laura Varela ; Garcia-Esquinas, Esther ; Correa-Perez, Andrea ; Fernandez-Hidalgo, Nuria ; Muriel, Alfonso; López-Alcalde, Jesús; Alvarez-Diaz, Noelia ; Pijoan, Jose I. ; Ribera, Aida ; Navas Elorza, Enrique ; Munoz, Patricia ; del Carmen Farinas, María ; Goenaga, Miguel Angel ; Zamora, Javier

DOI:
10.1016/j.cmi.2021.05.051

Document Version
Peer reviewed version

Citation for published version (Harvard):
Fernandez-Felix, BM, Barca, LV, Garcia-Esquinas, E, Correa-Perez, A, Fernandez-Hidalgo, N, Muriel, A, López-Alcalde, J, Alvarez-Diaz, N, Pijoan, JI, Ribera, A, Navas Elorza, E, Munoz, P, del Carmen Farinas, M, Goenaga, MA & Zamora, J 2021, 'Prognostic models for mortality after cardiac surgery in patients with infective endocarditis: a systematic review and aggregation of prediction models', *Clinical Microbiology and Infection*, vol. 27, no. 10, pp. 1422-1430. https://doi.org/10.1016/j.cmi.2021.05.051

CMI
CLINICAL MICROBIOLOGY
AND INFECTION

ESCMID

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

Journal:	*Clinical Microbiology and Infection*
Manuscript ID	CLM-21-21010.R1
Article Type:	Systematic Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Fernandez-Felix, Borja M; Hospital Universitario Ramon y Cajal, Clinical Biostatistics Unit. IRYCIS; CIBERESP, CIBER Epidemiology and Public Health Varela Barca, Laura; Fundación Jiménez Díaz, Department of Cardiovascular Surgery Garcia-Esquinas, Esther; CIBERESP, CIBER Epidemiology and Public Health; Universidad Autonoma de Madrid, Department of Preventive Medicine and Public Health. School of Medicine. IdiPaz Correa-Pérez, Andrea; Hospital Universitario Ramon y Cajal, Clinical Biostatistics Unit. IRYCIS; Universidad Francisco de Vitoria, Faculty of Medicine Fernández-Hidalgo, Nuria; Hospital Universitari Vall d'Hebron, Servei de Malalties Infeccioses; Instituto de Salud Carlos III, Red Española de Investigación en Patología Infecciosa (REIPI) Muriel, Alfonso; Universidad de Alcalá de Henares, Departamento de Enfermería y Fisioterapia; Hospital Universitario Ramon y Cajal, Clinical Biostatistics Unit. IRYCIS; CIBERESP, CIBER Epidemiology and Public Health Lopez-Alcalde, Jesus; Hospital Universitario Ramon y Cajal, Clinical Biostatistics Unit. IRYCIS; CIBERESP, CIBER Epidemiology and Public Health; Universidad Francisco de Vitoria, Faculty of Medicine; University Hospital Zurich and University of Zurich, Institute for Complementary and Integrative Medicine Álvarez-Díaz, Noelia; Hospital Universitario Ramon y Cajal, Medical Library Pijoan, Jose I; CIBERESP, CIBER Epidemiology and Public Health; Cruces University Hospital, OSI EEC; Biocruces-Bizkaia Health Research Institute Ribera, Aida; Hospital Universitari Vall d'Hebron, Cardiovascular Epidemiology and Research Unit; CIBERESP, CIBER Epidemiology and Public Health Navas, Enrique; Hospital Universitario Ramon y Cajal, Infectious Diseases Muñoz, Patricia; Hospital General Universitario Gregorio Marañón, Clinical Microbiology and Infectious Diseases ; Hospital General Universitario Gregorio Marañón-CIBERES, Clinical Microbiology and Infectious Diseases Fariñas, M Carnen; Hospital Universitario Marques de Valdecilla,

	Infectious Diseases Service. IDIVAL Goenaga, Miguel; Hospital Universitario de Donostia. IIS Biodonostia. OSI Donostialdea, Clinical Microbiology and Infectious Diseases Zamora , Javier ; Hospital Universitario Ramon y Cajal, Clinical Biostatistics Unit. IRYCIS; CIBERESP, CIBER Epidemiology and Public Health; University of Birmingham, WHO Collaborating Centre for Global Women's Health, Institute of Metabolism and Systems Research
Key Words:	Prognostic models, Systematic review, Meta-model, Aggregation, Validation, Infective Endocarditis
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 post-operative 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 post-operative 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 post-operative mortality. Eight models had a high risk of bias. The meta-model included weighted predictors from the remaining three models (i.e., EndoSCORE, specific ES-I and specific ES-II), which were not rated as high risk of bias and provided full model equation. Additionally, two variables (i.e., age and infectious agent) which had been modelized differently across studies, were estimated based on the nationwide registry. The performance of the meta-model was better than the original three models, with the corresponding performance measures: C-statistics 0.79 (95% CI 0.76 to 0.82), calibration slope 0.98 (95% CI 0.86 to 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 post-operative mortality in patients with IE. Protocol Registration: PROSPERO (registration number CRD42020192602)

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

Author list:

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], Aida Ribera[2,12], Enrique Navas Elorza[13], Patricia Muñoz[14], Mª Carmen Fariñas[15], M. Ángel Goenaga[16], Javier Zamora[1,2,17]

Affiliations:

[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; Biocruces-Bizkaia Health Research Institute, Barakaldo, Spain
[12] Cardiovascular Epidemiology and Research Unit, Hospital Universitari Vall d'Hebron, Barcelona, Spain
[13] Department of Infectology, Hospital Universitario Ramon y Cajal (IRYCIS), Madrid, Spain
[14] Clinical Microbiology and Infectious Diseases Service, Hospital General Universitario Gregorio Marañón, Madrid. Instituto de Investigación Sanitaria Gregorio Marañón. CIBER Enfermedades Respiratorias-CIBERES. Facultad de Medicina, Universidad Complutense de Madrid, Spain
[15] Infectious Diseases Service. Hospital Universitario Marqués de Valdecilla-IDIVAL. Universidad de Cantabria, Santander, Spain.
[16] Infectious Diseases Service. Hospital Universitario Donostia. IIS Biodonostia. OSI Donostialdea. San Sebastián, Spain
[17] WHO Collaborating Centre for Global Women's Health, Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, UK

Corresponding author:
Complete name: Borja Manuel Fernández Félix
Email address: borjamanuel86@gmail.com
ORCID: http://orcid.org/0000-0002-8798-019X
Postal address: Hospital Universitario Ramón y Cajal. Ctra. de Colmenar Viejo, Km. 9,100 (28034) MADRID
Phone number: +34 91 3368103
Fax number: +34 91 3369016

Category: Systematic review

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 post-operative 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 post-operative 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 post-operative mortality. Eight models had a high risk of bias. The meta-model included weighted predictors from the remaining three models (*i.e.*, EndoSCORE, specific ES-I and specific ES-II), which were not rated as high risk of bias and provided full model equation. Additionally, two variables (*i.e.*, age and infectious agent) which had been modelized differently across studies, were estimated based on the nationwide registry. The performance of the meta-model was better than the original three models, with the corresponding performance measures: C-statistics 0.79 (95% CI 0.76 to 0.82), calibration slope 0.98 (95% CI 0.86 to 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 post-operative mortality in patients with IE.

Protocol Registration: PROSPERO (registration number CRD42020192602)

Key words: Prognostic models, systematic review, meta-model, aggregation, validation, infective endocarditis.

Background

Infective endocarditis (IE) is an uncommon but severe disease with a high mortality rate. Its current estimated incidence is 3-10 episodes per 100.000 person-years, while 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 modeling 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 processes.

Prognostic models are mathematical equations that relates multiple variables for a particular individual to the probability of post-operative mortality. In the last decade, several IE prognostic models using preoperative patient´s-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 lack of external validation. Consequently, researchers carry on developing new models using their own data without considering prior 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 post-operative 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.

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 01/06/2020. We conducted a literature search to identify all potential studies for inclusion, without any language or publication dates restriction. We used the methodologic 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 **Supplementary Material: S1.**

Eligibility criteria

We included original studies that developed prognostic models, with or without external validation, to predict the risk of post-operative mortality after cardiac surgery in patients with IE, as well as studies that updated previously published models. We accepted the authors` definition of post-operative 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. (**Supplementary Material: S2)**. When the completed model equation or relevant data were not provided, we contacted the correspondence 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 (**Supplementary Material: S3**).

GAMES registry

We used the nationwide GAMES – Grupo de Apoyo al Manejo de la Endocarditis infecciosa en España – (14) registry as the validation dataset, 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 center. For

the present study, we selected all the infective episodes (n=1,453) 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). **Supplementary Material: Table S1** 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 dataset (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 risk of bias. Stacked regressions used the linear predictor of each model as a co-variable in the meta-model, to subsequently created 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 dataset 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 modeled 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 dataset. Non-linear relationships for continuous predictors were tested using fractional polynomials (18).

Predictors with missing data in the validation dataset 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. (**Supplementary Material: S4**). Imputations checks were completed by looking at the distributions of imputed values to ensure plausibility. We generated 10 multiple imputed datasets and all primary analyses were performed in each imputed dataset. 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 datasets using Rubin's rules (20). Because the meta-model was optimized to the validation dataset, 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-days 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).

Results

Search results and study selection

We retrieved 4,862 titles through our systematic search combining Medline and Embase. From these, 684 duplicate references were identified. Of 4,178 titles assessed by title and

abstract, 34 studies were retained for full text screening, and 2 additional studies were detected in the bibliographic references of these articles. Nine studies describing 11 prediction models met the inclusion criteria (**Figure 1 and Supplementary 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 center (24–26) and three from multiple centers (27–29)); two studies used data from multicenter registries (30,31); and one study used data from both a multicenter 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. (**Supplementary 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 (NYHA) classification of functional status (n=6), paravalvular complications (n = 6) and infection etiology (n = 5). The predictor definitions and the models' composition are shown in the **Supplementary Table S4 and Table S5**.

Model development and presentation

Sample sizes for models' development varied between 128 and 13,617 patients, and the number of events ranged from 21 to 1,117. 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 technique (n = 9), while logistic mixed effects (27) and logistic Generalized Estimating Equation (GEE) 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 (EPP) 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) (**Supplementary Table S6**). The method of predictors 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 inform 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 correspondence authors did not respond when were asked for further details) (**Supplementary 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' model (29) and 21 in the Gatti's 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, Supplementary Table S8 and Figure 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 EPP 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 risk of bias, 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 (Specific ES-I) and EuroSCORE-II (Specific ES-II) models were aggregated in the meta-model. The model developed by Di Mauro (EndoSCORE) (27) included 15 parameters, while the other two (Specific ES-I and Specific ES-II) developed by Fernández-Hidalgo (28), presented 10 and 9 parameters respectively, from the EuroSCORE models predictors (36,37) and IE-related factors (**Table 2 and Supplementary Table S7**). The dependent variable for the meta-model was mortality (either 30-days or in-hospital).

To construct the meta-model, we first calculated the linear predictors (LP) from EndoSCORE, Specific ES-I and Specific ES-II for each observation in the validation dataset, after dropping the parameters for age and infection etiology because these variables were modelized differently in the different studies. Subsequently, we adjusted the meta-model using a logistic regression model, which incorporated the LPs as co-variables, to estimate the models' weights for aggregation, as well as the predictors for age (treated as continuous) and infection etiology (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 are patient-related factors (i.e. age, gender, renal failure, prior cardiac surgery, chronic pulmonary disease, pulmonary hypertension and left ventricular ejection fraction), clinical presentation-related factors (i.e. critical preoperative state, New York Heart Association (NYHA) classification of functional status), surgery-related factors (i.e. presence of paravalvular complications (abscess and/or fistulae), urgency of procedure and number of treated valves/prostheses) and finally IE-related factors (i.e. valve location and infection etiology) (**Supplementary Table S5**). We have developed an online calculator to allow a simple and effective use of the meta-model. The magnitude of the associations of the predictive factors with mortality is shown in **Table 2** and the complete meta-model equation in **Supplementary 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 to 0.788) for EndoSCORE, 0.758 (95% CI 0.731 to 0.786) for Specific ES-I, and 0.762 (95% CI 0.735 to 0.789) for Specific ES-II. The optimism adjusted C-statistic for the meta-model was 0.79 (95% CI 0.76 to 0.82) (**Figure 2**). Calibration slopes were < 1 for all published models: 0.80 (95% CI 0.69 to 0.92) for

EndoScore, 0.82 (95% CI 0.70 to 0.94) for Specific ES-I, and 0.76 (95% CI 0.65 to 0.87) for Specific ES-II. CITL was 0.58 (95% CI 0.44 to 0.71) for EndoSCORE and 0.62 (95% CI 0.48 to 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 to 1.13) for the slope and -0.05 (95% CI -0.20 to 0.11) for CITL (**Figure 2**). The calibration plots for the three previously published models and the meta-model are shown in **Figure 3**.

Sensitivity analysis showed that the meta-model had better overall performance than all published models regardless of their quality assessment (**Supplementary Figure S2**). Moreover, even though the meta-model was not fitted for the 30-days mortality outcome, it outperformed the three models used for model aggregation. (**Supplementary Figure S3**)

Discussion

Summary of findings

In this systematic review of prediction models for post-operative mortality in patients with infective endocarditis, we identified and critically appraised 11 models developed in 9 studies. The predicted outcome varied between studies (in-hospital, 30-days or both in-hospital or 30-days mortality). Of the eleven prognostic models, only two had low RoB and one unclear; 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 healthcare decisions in practice (34).

Four out of the 11 published models reported the full model equation required for a models' 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 (NYHA) classification of functional status presence of paravalvular complications (abscess and/or fistulae), urgency of procedure, number of treated valves/prostheses, valve location and infection etiology. It showed better performance than the original models. We investigated the internal validity of the meta-model using bootstrap validation, and the results indicate 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 likely 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 post-operative mortality in patients with infective endocarditis, 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 prior knowledge optimally and outperformed previously published models.

Our study has some limitations. The outcome definition in the validation dataset was either 30-days or in-hospital post-operative mortality, and the outcome definition in the three models used for aggregation was 30-days mortality. Despite this difference a sensitivity analysis showed that the meta-model outperformed all published models when we explored its performance for the 30-days 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 to 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 these incomplete models were classified as high risk of bias 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 risk of bias. 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 risk of bias 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 LVEF, NYHA, renal failure have also 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 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 a great value helping endocarditis teams to manage this complex disease and lead toward 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-Cotrufo et al 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 infective endocarditis, 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 post-operative mortality in patients with infective endocarditis. 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.

Conflict of interest

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and the authors have declared that no competing interests exist.

Funding

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.

Acknowledgements

We are grateful to the team involved with the collection of the GAMES registry data (**Supplementary Material: S6**).

Authors contributions

Conceptualization: BMFF, LVB, EGE, JLA, AM, JIP, AR, JZ; Search strategies: BMFF, NAD, JLA; Data extraction and Critical appraisal: BMFF, LVB, ACP; Methodology: BMFF, EGE, AM, JZ; Software, Formal analysis: BMFF; Validation: AM, JZ; Data adquisition/curation: BMFF, ENE, PM, MCF, MAG: Writing - Original draft: BMFF, EGE, JZ; Visualization: BMFF, LVB, NFH; Supervision: EGE, JZ; Writing – Review & Editing: All authors.

Bibliography

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 Mar 9;169(5):463.

2. Thuny F, Grisoli D, Collart F, Habib G, Raoult D. Management of infective endocarditis: challenges and perspectives. The Lancet. 2012 Mar;379(9819):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 Mar 7;37(10):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 Jan 13;131(2):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 Jan 5;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 Oct 14;11(10):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 Jul 21;6(7):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 Jan 7;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. Annals of Internal Medicine. 2015 Jan 6;162(1):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. Smalheiser NR, editor. PLoS ONE. 2012 Feb 29;7(2):e32844.

11. Thomas, J., Graziosi, S., Brunton, J., Ghouze, Z., O'Driscoll, P., & Bond, M. (2020). EPPI-Reviewer: advanced software for systematic reviews, maps and evidence synthesis. EPPI-Centre Software. London: UCL Social Research Institute.

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 Jan 1;170(1):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 Jan 1;170(1):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 (Baltimore). 2015 Oct;94(43):e1816.

15. Breiman L. Stacked regressions. Mach Learn. 1996 Jul;24(1):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. Statist Med. 2014 Jun 30;33(14):2341–62.

17. Martin GP, Mamas MA, Peek N, Buchan I, Sperrin M. A multiple-model generalisation of updating clinical prediction models. Statistics in Medicine. 2018 Apr 15;37(8):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. 303 p. (Wiley series in probability and statistics).

19. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Statist Med. 2011 Feb 20;30(4):377–99.

20. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: Wiley; 1987.

21. Riley RD, Windt D van der, 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 [cited 2020 Apr 28]. Available from: https://doi.org/10.1007/978-3-030-16399-0

23. StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC.

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. The Scientific World Journal. 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 Aug;45(4):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 Feb;22(2):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. International Journal of Cardiology. 2017 Aug;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. Clinical Microbiology and Infection. 2018 Oct;24(10):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 Sep;103(18):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. The Journal of Thoracic and Cardiovascular Surgery. 2011 Jan;141(1):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. International Journal of Cardiology. 2014 Jul;175(1):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. JAHA [Internet]. 2017 Jul [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. Statistics in Medicine. 2019 Mar 30;38(7):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 Mar 18;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 Dec;17(1):230.

36. Nashef SAM, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). European Journal of Cardio-Thoracic Surgery. 1999 Jul;16(1):9–13.

37. Nashef SAM, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, et al. EuroSCORE II. European Journal of Cardio-Thoracic Surgery. 2012 Apr 1;41(4):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†. European Journal of Cardio-Thoracic Surgery. 2018 May 1;53(5):1049–54.

39. Pivatto Júnior F, Bellagamba CC de A, 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. ABC Cardiol [Internet]. 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. Scandinavian Cardiovascular Journal. 2019 May 4;53(3):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 Mar;30(2):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 Dec;47(6):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. European Journal of Cardio-Thoracic Surgery. 2020 Apr 1;57(4):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. Statistics in Medicine. 2012 Oct 15;31(23):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 Jul 1;102(13):992–4.

46. Efron B, Morris C. Stein's Paradox in Statistics. Scientific American. 1977;236(5):119–27.

47. van Houwelingen JC. Shrinkage and Penalized Likelihood as Methods to Improve Predictive Accuracy. Statistica Neerland. 2001 Mar;55(1):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 Oct;12(10):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 Jun 15;40(13):3066–84.

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.

1

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; -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; 0.48)	1.19 (0.90; 1.62)
LVEF (%)	-0.03			-0.013 (-0.02; -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)
Etiology[c]	-	-	-		
- *Staphylococcus* spp.				0.64 (0.35; 0.94)	1.90 (1.42; 2.56)
- Fungi				0.61 (-0.46; 1.40)	1.84 (0.63; 4.06)

LVEF: left ventricular ejection fraction; NYHA class: New York Health Association classification of functional status; OR: Odds ratio

[a] Weights used to create the meta-model: EndoScore = 0.433; Sp. ES-I = 0.131; Sp. ES-II = 0.379

Stacked regression:

$$\ln\left(\frac{p}{1-p}\right) = -1.861 + 0.433 \times LP^{\dagger}_{DM} + 0.131 \times LP^{\dagger}_{FH-I} + 0.379 \times LP^{\dagger}_{FH-II} + 0.045 \times Age + 0.64 \times Staphyloco + 0.61 \times Fungi$$

Where, p is the probability of post-operative mortality and $LP_i^{\dagger}$ is the linear predictor for each model selected for aggregation dropping the parameters from age and infection etiology; 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 x (-2.60) + 0.131 x (-3.13) + 0.379 x (-4.21) = -5.00, and for instance, the stacked coefficient for renal failure = 0.433 x (0.50) + 0.131 x (0.46) + 0.379 x (0) = 0.277

[b] Age was categorized in Di Mauro 2017 and treated as continuous in Fernández-Hidalgo 2018

[c] Etiology was categorized in different ways in each existing model.

Figure 1. PRISMA flowchart of study inclusions and exclusions.

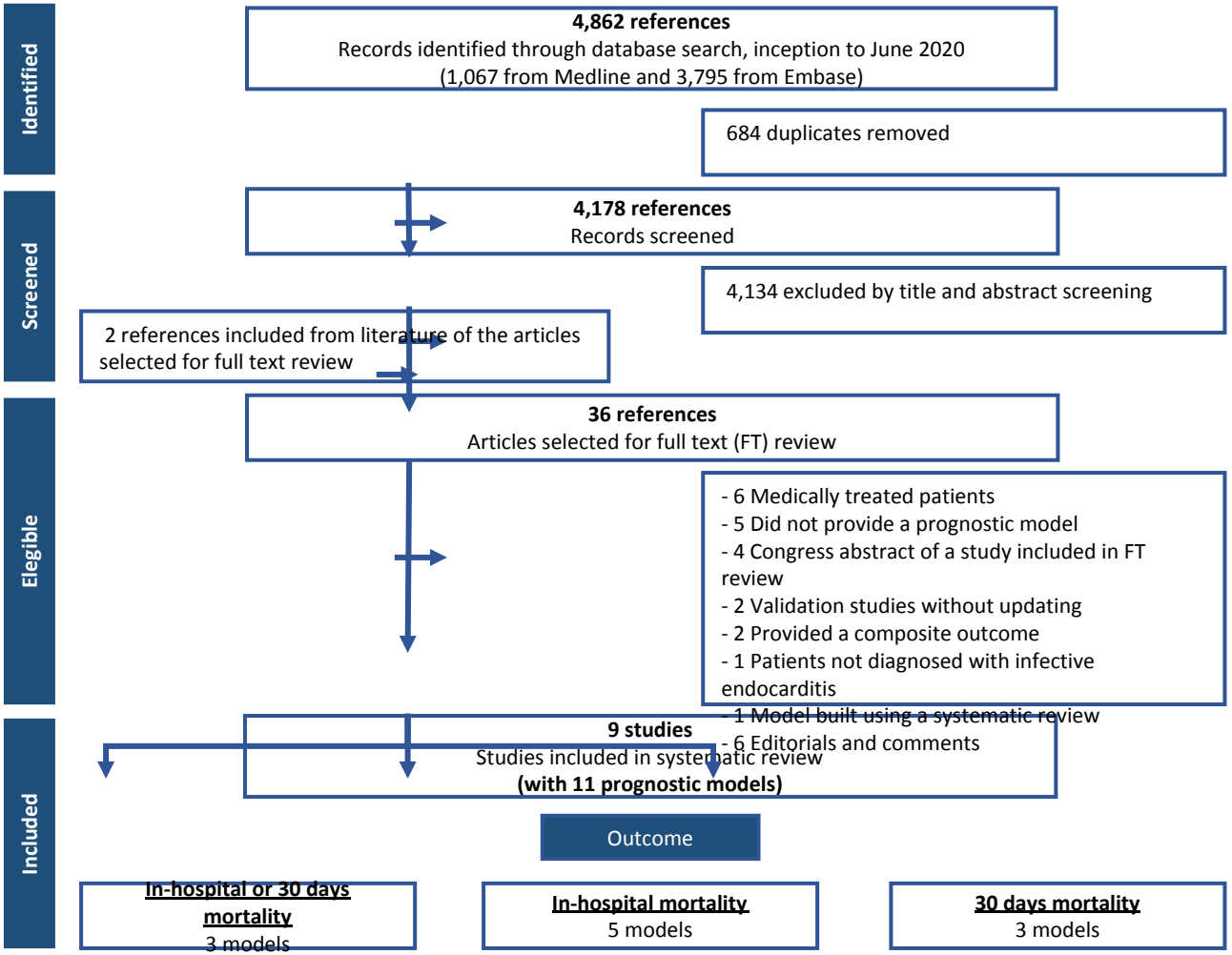
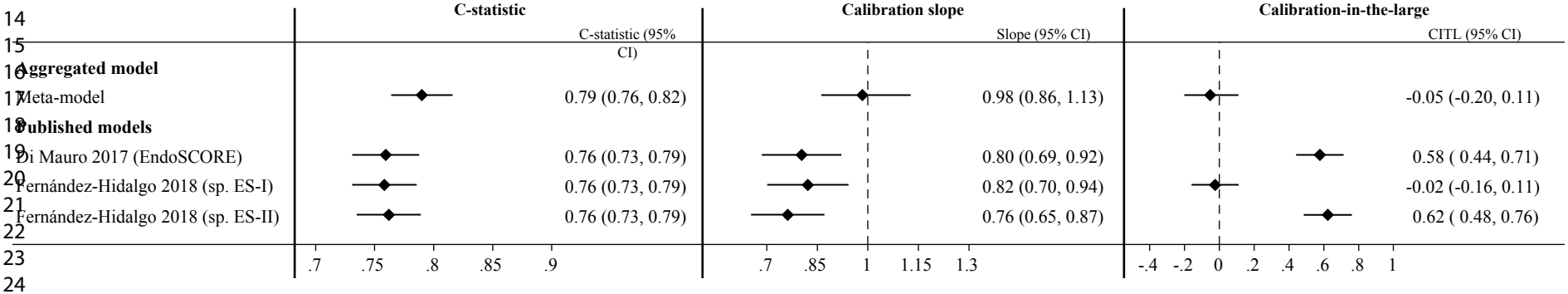
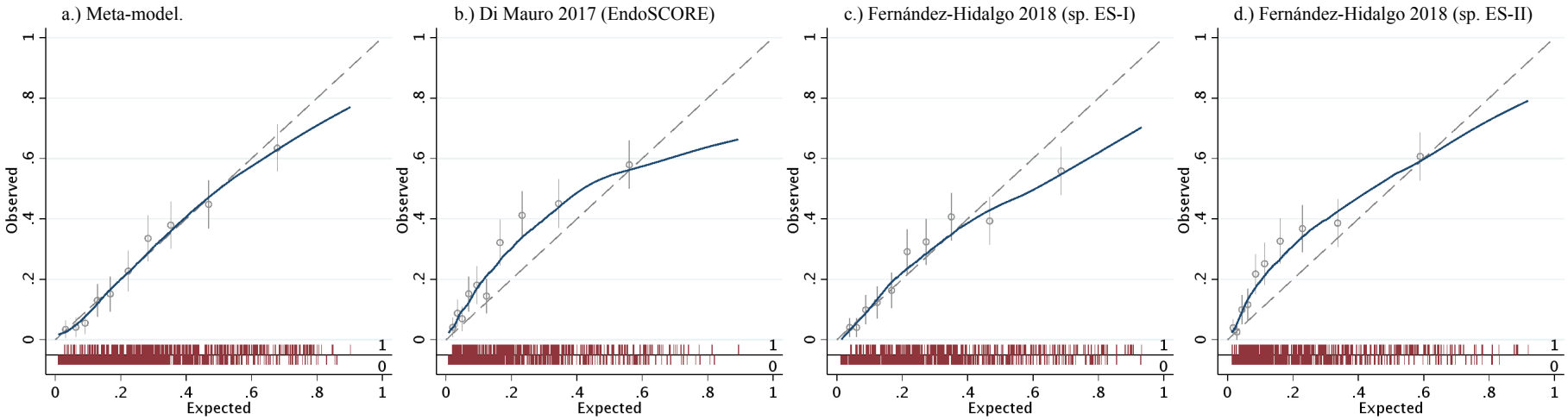


Figure 2. Bootstrap internal validation of the meta-model and external validation of existing models selected for aggregation



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

Figure 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.

Supplementary material

S1: Search strategies....2
Ovid (Medline)....2
Embase (Elsevier)....3
S2: Data extraction....4
S3: Critical appraisal and applicability....5
S4: Data imputation....5
S5: Statistical software....6
Table S1: Characteristics of patients included in the validation dataset (GAMES registry)....6
Table S2: Studies excluded and motive of exclusion....8
Table S3: Characteristics of the primary studies....10
Table S4: Definition of the predictors....11
Table S5. Model compositions and percentage of missing data in GAMES registry....14
Table S6: Minimum sample size for development of a new multivariable prediction model....15
Table S7: Prognostic models equation....16
Box S1: meta-model equation and example of use....17
Table S8: Critical appraisal using PROBAST....18
Figure S1: Summary of risk of bias and applicability of the studies....20
Figure S2: Validation of all models regardless of critical appraisal....21
Figure S3: Validation of the meta-model and existing models selected for aggregation for 30-days mortality outcome....22
S6: Members of GAMES group....23

S1: Search strategies

The following exact search was used (search date 01/06/2020):

Ovid (Medline)

1. exp Endocarditis/
2. endocarditi*.tw.
3. 1 or 2
4. Cardiac Surgical Procedures/
5. (cardiac and (surger* or procedure*)).tw.
6. 4 or 5
7. 3 and 6
8. Validat$.af.
9. Predict$.ti.
10. Rule$.af.
11. 8 or 9 or 10
12. (Predict$ and (Outcome$ or Risk$ or Model$)).af.
13. ((History or Variable$ or Criteria or Scor$ or Characteristic$ or Finding$ or Factor$) and (Predict$ or Model$ or Decision$ or Identif$ or Prognos$)).af.
14. Decision$.af.
15. Logistic Models/
16. Model$.af.
17. Clinical$.af.
18. 15 or 16 or 17
19. 14 and 18
20. (Prognostic and (History or Variable$ or Criteria or Scor$ or Characteristic$ or Finding$ or Factor$ or Model$)).af.
21. 11 or 12 or 13 or 19 or 20
22. exp ROC Curve/
23. stratification.af.
24. discrimination.af.
25. discriminate.af.
26. c-statistic.af.
27. c statistic.af.
28. "Area under the curve".af.
29. AUC.af.
30. calibration.af.
31. indices.af.
32. algorithm.af.
33. multivariable.af.
34. 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33
35. 21 or 34
36. 7 and 35

Embase (Elsevier)

#1 'endocarditis'/exp

#2 endocardit*:ab,ti

#3 #1 OR #2

#4 'heart surgery'/exp

#5 cardiac:ab,ti AND (surger*:ab,ti OR procedure*:ab,ti)

#6 #4 OR #5

#7 #3 AND #6

#8 validat*:ab,ti

#9 predict*:ti

#10 rule*:ab,ti

#11 #8 OR #9 OR #10

#12 predict*:ab,ti AND (outcome*:ab,ti OR risk*:ab,ti OR model*:ab,ti)

#13 (history:ab,ti OR variable*:ab,ti OR criteria:ab,ti OR scor*:ab,ti OR characteristic*:ab,ti OR finding*:ab,ti OR factor*:ab,ti) AND (predict*:ab,ti OR model*:ab,ti OR decision*:ab,ti OR identif*:ab,ti OR prognos*:ab,ti)

#14 decision*:ab,ti

#15 'statistical model'/exp

#16 model*:ab,ti

#17 clinical*:ab,ti

#18 #15 OR#16 OR #17

#19 #14 AND #18

#20 prognostic:ab,ti AND (history:ab,ti OR variable*:ab,ti OR criteria:ab,ti OR scor*:ab,ti OR characteristic*:ab,ti OR finding*:ab,ti OR factor*:ab,ti OR model*:ab,ti)

#21 #11 OR #12 OR #13 OR #19 OR #20

#22 'receiver operating characteristic'/exp

#23 stratification:ab,ti

#24 discrimination:ab,ti

#25 discriminate:ab,ti

#26 'c-statistic':ab,ti

#27 'c statistic':ab,ti

#28 'area under the curve':ab,ti

#29 auc:ab,ti

#30 calibration:ab,ti

#31 indices:ab,ti

#32 algorithm:ab,ti

#33 multivariable:ab,ti

#34 #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33

#35 #21 OR #34

#36 #7 AND #35 #37 #7 AND #35 AND ([embase]/lim OR [pubmed-not-medline]/lim)

S2: Data extraction

Information on the following items was extracted using a standardized form based on CHARMS (CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies):

1. Study information: Author, year, journal and model´s name.
2. Source of data.
3. Participants: Recruitment method and dates; study setting; study regions and number of centers involved; inclusion and exclusion criteria; patient´s age (mean and standard deviation or median and interquartile range); number and percentage of native valve endocarditis; number, percentage and type (i.e. aortic, mitral, pulmonary or tricuspid) of valves affected.
4. Outcome: Definition and timing of occurrence.
5. Predictors: Number of candidate predictors; type of predictors; definition; and timing of measurement (preoperative or intraoperative)
6. Analysis:
    a. Sample size: Number of participants, events and events per predictor/parameter (EPP).
    b. Missing data: Number of participants with any missing value and methods used to handle missing data.
    c. Model development: Modelling method; method for selection of candidate predictors; method for selection of predictors during multivariable modelling
    d. Model performance: Discrimination and calibration measures.
    e. Model evaluation: Type of validation (apparent, internal or external) and optimism adjustment.
    f. Model results: Number of predictors included in the final model; presentation (e.g. coefficients and confidence interval); inclusion of model´s constant; alternative presentation of the final model.

S3: Critical appraisal and applicability

Model were assessed to risk of bias using a standardized form based on the PROBAST on the following domains: Participants; Predictors; Outcome; Analysis.

The signalling questions were answered for each domain with one out of these options ('yes', 'probably yes', 'probably no', 'no', 'no information'); where 'yes' means the absence of a potential bias. We rated domain-level 'Risk of bias' assessments as:

- Low risk of bias: if the criterion is adequately fulfilled in the study, i.e. the study is at a low risk of bias for the given domain.
- High risk of bias: if the criterion is not fulfilled in the study, i.e. the study is at high risk of bias for the given domain.
- Unclear risk of bias: if the study report does not provide enough information to allow for a clear judgement or if the risk of bias is unknown for one of the domains listed above.

The applicability judgement of the model to the research question occurs per following domains: Participants, Predictors and Outcome. The possible responses were: 'low concern regarding applicability', 'high concern regarding applicability' and 'unclear concern regarding applicability' (equivalent to the categories for risk of bias).

If risk of bias or applicability were high in at least one of the domains, overall risk of bias or applicability was judged high. If at least one of the answers was “No” or “Probably no,” the judgment could still be low risk of bias, in this case specific reasons were provided. The complete information about of the 'Risk of bias' and 'Applicability' assessment of the authors is shown in **Supplementary Table S8 and Figure S1**.

S4: Data imputation

We used linear regression imputation for continuous variables, truncated regression imputation for continuous variable with a restricted range, logistic regression imputation for binary data, multinomial logistic regression imputation for unordered categorical data and ordered logistic regression imputation for ordered categorical data.

S5: Statistical software

The analyses were conducted in Stata version 16 using mi command for multiple imputation, `mfpmi` command for estimation meta-model coefficients using logistic regression modelling in presence of multiple imputation datasets, `roctab` and `logistic` command for C-statistics, slope calibration and calibration-in-the-large calculations. These commands were combined in a syntax (available from the corresponding author upon reasonable request) to obtain bootstrap confidence intervals and performance measures adjusted for optimism. `Forestplot` and `pmcalplot` commands were used for figures.

Table S1: Characteristics of patients included in the validation dataset (GAMES registry)

	Mortality		Missing data
	No (n=1,099)	Yes (n=354)	
	n (%)	n (%)	n
Patient related-factors			
Age (years), mean(sd)	62.0 (13.4)	68.9 (10.0)	-
Female	275 (25.1%)	112 (31.8%)	6
Chronic pulmonary disease	179 (18.3%)	83 (26.9%)	165
Diabetes	248 (22.6%)	131 (37.0%)	2
Hypertension	546 (49.8%)	238 (67.4%)	4
Pulmonary hypertension	58 (5.3%)	27 (7.6%)	-
Creatinine (mg/dl.), mean(sd)	1.1 (0.9)	1.4 (1.1)	56
Prior CABG	56 (5.1%)	31 (8.8%)	4
Prior valvular surgery	356 (32.5%)	168 (47.6%)	6
LVEF (%), mean(sd)	59.8 (11.0)	58.0 (12.0)	366
Clinical presentation related-factors			
Septic shock	85 (7.7%)	102 (28.8%)	
NYHA			25
• I	883 (81.5%)	241 (69.9%)	
• II	158 (14.6%)	68 (19.7%)	
• III	30 (2.8%)	27 (7.8%)	
• IV	12 (1.1%)	9 (2.6%)	
Preoperative status			23
• Elective	746 (69.1%)	180 (51.3%)	
• Urgent	265 (24.6%)	115 (32.8%)	
• Emergent	68 (6.3%)	56 (16.0%)	
Valves affected			-
• 0	13 (1.2%)	3 (0.8%)	
• 1	913 (83.1%)	288 (81.4%)	
• 2	169 (15.4%)	60 (16.9%)	
• 3	4 (0.4%)	3 (0.8%)	
Surgery-related factors			

Abscess	284 (26.0%)	124 (35.3%)	8
Fistula	34 (3.1%)	23 (6.5%)	-
Dehiscence	117 (10.7%)	63 (17.8%)	2
Weight of intervention			-
• Single non-CABG	867 (78.9%)	273 (77.1%)	
• 2 procedures	225 (20.5%)	74 (20.9%)	
• 3 procedures	7 (0.6%)	7 (2.0%)	
Surgery in aorta	24 (2.2%)	11 (3.1%)	-
IE-related factors			
Type of valve			15
• Natural	754 (69.4%)	186 (52.8%)	
• Prosthetic	332 (30.6%)	166 (47.2%)	
Valve location			
• No valve treated	13 (1.2%)	3 (0.8%)	
• Aortic	547 (49.8%)	164 (46.3%)	
• Mitral	350 (31.8%)	121 (34.2%)	
• Pulmonary	2 (0.2%)	0 (0.0%)	
• Tricuspid	14 (1.3%)	3 (0.8%)	
• Multiple	173 (15.7%)	63 (17.8%)	
Infection etiology			52
• *Staphylococcus* spp.	367 (34.7%)	190 (55.2%)	
- coagulase-negative staphylococci	208 (57%)	92 (48%)	
- S. *aureus*	159 (43%)	98 (52%)	
MSSA	115	75	
MIRSA	0	2	
MRSA	23	12	
Unknown	21	9	
• *Pseudomonas* spp.	3 (0.3%)	4 (1.2%)	
• Fungal disease	20 (1.9%)	10 (2.9%)	
• *Streptococcus* spp.	363 (34.3%)	70 (20.3%)	
• Other microorganisms	304 (28.8%)	70 (20.3%)	

n: number of patients; sd: standard deviation; CABG: coronary artery bypass grafting; LVEF: left ventricular ejection fraction; NYHA: New York Heart Association; MSSA: methicillin sensitivity S. *aureus*; MIRSA: methicillin intermediate resistant S. *aureus*; MRSA: methicillin resistant S. *aureus*

Table S2: Studies excluded and motive of exclusion

DOI / PMID	Reference
Medically treated patients	
PMID: 3893114	Alsip S G, Blackstone E H, Kirklin J W, Cobbs C G. 1985. "Indications for cardiac surgery in patients with active infective endocarditis". *The American journal of medicine* 78(6B):138-48.
PMID: 2759756	Woo K S, Lam Y M, Kwok H T, Tse L K. K, Vallance-Owen J. 1989. "Prognostic index in prediction of mortality from infective endocarditis". *International Journal of Cardiology* 24(1):47-54.
	Kjaergaard J, Rasmussen R, Bruun N, Hassager C. 2009. "Vegetation length or area: Which is the better predictor of outcome in infective endocarditis?". *International Journal of Antimicrobial Agents* 33:S27-S28.
10.1177/2048872615574706	Guimaraes ” Baseline predictors of in-hospital mortality in patients with infective endocarditis”. Abstracts for the Cardiac Society of Australia and New Zealand Annual Scientific Meeting and the International Society for Heart Research Australasian Section Annual Scientific Meeting. 2016. *Heart Lung and Circulation* 25:.
10.1016/j.recesp.2020.04.010	García-Granja P E, López J, Vilacosta I, Sarriá C, Domínguez F, Ladrón R, et al.. 2020. "Predictive model of in-hospital mortality in left-sided infective endocarditis". *Revista Espanola de Cardiologia* :.
10.1590/s0102-76382007000200007	Costa MA, Wollmann DR Jr, Campos AC, Cunha CL, Carvalho RG, Andrade DF, et al.. 2007. "Risk index for death by infective endocarditis: a multivariate logistic model.". *Revista brasileira de cirurgia cardiovascular : orgao oficial da Sociedade Brasileira de Cirurgia Cardiovascular* 22(2):192-200.
Did not provide a prognostic model	
10.1177/2048872616663431	Garcia Granja, P E, Lopez J, Ladron R, Vilacosta I, Olmos C, Ortiz Bautista, et al.. 2016. "Influence of valve culture in prognosis of leftsided infective endocarditis". *European Heart Journal: Acute Cardiovascular Care* 5:384-385.
10.1093/ejcts/ezv223	Patrat-Delon Solene, Rouxel Adrien, Gacouin Arnaud, Revest Matthieu, Flecher Erwan, Fouquet Olivier, et al.. 2016. "EuroSCORE II underestimates mortality after cardiac surgery for infective endocarditis". *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery* 49(3):944-51.
10.1016/j.jescts.2017.02.004	Elmasry A, Omran A M, Elprince A, Elameen S, Mansy M M, Mahlab A S. 2017. "Predictors of in-hospital mortality in surgically treated valvular infective endocarditis cases at National Heart institute, Egypt". *Journal of the Egyptian Society of Cardio-Thoracic Surgery* 25(1):35-44.
10.1177/0218492318798258	Nagy Mohamad, Alkady Hesham, Abo Senna, Waleed, Abdelhay Soliman. 2018. "Predictors of surgical outcome in isolated prosthetic mitral valve endocarditis". *Asian cardiovascular & thoracic annals* 26(7):517-523.
10.1016/j.repc.2019.08.009	Guiomar N, Vaz-da-Silva M, Mbala D, Sousa-Pinto B, Monteiro J P, Ponce P,et al.. 2020. "Cardiac surgery in infective endocarditis and predictors of in-hospital mortality". *Revista Portuguesa de Cardiologia* :.
Congress abstract of a study included in full text review	
Original study ref: 10.1016/j.ijcard.2014.04.266	Martinez-Selles M, Munoz P, Arnaiz A, Moreno M, Galvez J, Rodriguez-Roda J, et al.. 2014. "Valve surgery in active infective endocarditis: A simple score to predict in-hospital prognosis". *European Heart Journal* 35:756.
Original study ref: 10.1093/icvts/ivv304	Madeira S, Santos M, Rodrigues R, Tralhao A, Mesquita J, Carmo J, et al.. 2015. "Assessment of operative mortality risk in patients with active infective endocarditis undergoing cardiac surgery: Performance of the EuroScore I and II logistic models". *European Heart Journal* 36:268.
Original study ref: 10.1136/heartjnl-2016-311093	Olmos C, Vilacosta I, Fernandez C, Tirado G, Freitas-Ferraz A, Lopez J, et al.. 2015. "Development and validation of a risk score for cardiac surgery in infective endocarditis". *European Heart Journal* 36:374.
Original study ref: 10.1007/s00380-014-0472-0	Wang T K. M, Oh T, Voss J, Kang N, Pemberton J. 2013. "Comparison and implications of contemporary risk scores for predicting mortality and morbidity after surgery for active infective endocarditis". European Heart Journal 34:502.
Validation studies without updating	

10.1007/s00380-014-0472-0	Wang Tom Kai Ming, Oh Timothy, Voss Jamie, Gamble Greg, Kang Nicholas, Pemberton James. 2015. "Comparison of contemporary risk scores for predicting outcomes after surgery for active infective endocarditis". Heart and vessels 30(2):227-34.
10.1080/14017431.2019.1610188	Gatti Giuseppe, Sponga Sandro, Peghin Maddalena, Givone Filippo, Ferrara Veronica, Benussi Bernardo, et al.. 2019. "Risk scores and surgery for infective endocarditis: in search of a good predictive score". Scandinavian cardiovascular journal : SCJ 53(3):117-124
Provided a composite outcome	
10.1001/jama.289.15.1933	Hasbun R, Vikram H R, Barakat L A, Buenconsejo J, Quagliarello V J. 2003. "Complicated Left-Sided Native Valve Endocarditis in Adults: Risk Classification for Mortality". *Journal of the American Medical Association* 289(15):1933-1940.
10.1136/hrt.2010.200295	Lopez Javier, Fernandez-Hidalgo Nuria, Revilla Ana, Vilacosta Isidre, Tornos Pilar, Almirante Benito, et al.. 2011. "Internal and external validation of a model to predict adverse outcomes in patients with left-sided infective endocarditis". *Heart (British Cardiac Society)* 97(14):1138-42.
Patients not diagnosed with infective endocarditis	
10.1016/j.ejcts.2011.01.002	Akar Ahmet Ruchan, Kurtcephe Murat. Sener Erol, Alhan Cem, Durdu Serkan, Kunt Ayse Gul, et al.. 2011. "Validation of the EuroSCORE risk models in Turkish adult cardiac surgical population". *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery* 40(3):730-5.
Editorials and Coments	
Editorial: 10.1001/jama.289.15.1991	Granowitz E V, Longworth D L. 2003. "Risk Stratification and Bedside Prognostication in Infective Endocarditis". Journal of the American Medical Association 289(15):1991-1993.
Editorial: 10.36660/abc.20200070	Martins A B. B, Lamas C D. C. 2020. "Prognostic scores for mortality in cardiac surgery for infective endocarditis". Arquivos Brasileiros de Cardiologia 114(3):525-529.
Editorial: 10.1053/j.jvca.2018.02.005	Stein Erica, Andritsos Michael. 2018. "Risk Stratification and Optimization of Cardiac Surgical Patients With Infective Endocarditis: Does It Matter?". Journal of cardiothoracic and vascular anesthesia 32(6):2537-2539.
Editorial: 10.21037/jtd.2019.09.69	Tattevin Pierre, Fillatre Pierre, Tchamgoue Serge, Lesouhaitier Mathieu, Nesseler Nicolas, Tadie Jean-Marc. 2019. "Should we include microorganisms in scores to predict outcome in candidates for cardiac surgery during the acute phase of endocarditis?". Journal of thoracic disease 11(10):E158-E162.
Comment: 10.2169/internalmedicine.3579-19	Toyoda S, Saito F, Inoue T. 2020. "Authors' reply: How to construct novel criteria for predicting complication with infectious endocarditis". Internal Medicine 59(1):147-148.
Comment: 10.1016/j.ijcard.2015.08.167	Wang T K. M. 2016. "Risk scores for endocarditis surgery: Callout for reporting logistic models". International Journal of Cardiology 202:960.
Model built using a systematic review	
10.1093/ejcts/ezz328	Varela Barca, L, Fernández-Felix B M, Navas Elorza E, Mestres C A, Muñoz P, Cuerpo-Caballero G, et al.. 2020. "Prognostic assessment of valvular surgery in active infective endocarditis: Multicentric nationwide validation of a new score developed from a meta-analysis". *European Journal of Cardio-thoracic Surgery* 57(4):724-731.

92

Table S3: Characteristics of the primary studies.

Author, Year	Enrolment period	Study setting	Study design	Study region (Centers)	Age Mean (sd) or median (Q_1;Q_3)	Native valve (%)	Valves affected
In-hospital or 30 days mortality							
De Feo, 2012	1980 - 2009	Cardiac surgery centers	Retrospective cohort	Italy (1)	49 (16)	100	All
Gaca, 2011	2002 - 2008	Cardiac surgery centers	Registry (STS ACSD)	North America (Unclear)	55 (45;66)	NI	All
Madeira, 2016	2007 - 2014	Cardiac surgery centers	Retrospective cohort	Portugal (1)	60 (47;70)	73.4	All
In-hospital mortality							
Gatti, 2017a	2000-2015 (Italy) 2008 (France)	Cardiac surgery centers	Retrospective cohort and registry (AEPEI)	Italy (1) France (7)	59.1 (15.4)	78.9	All
Gatti, 2017b	1999 - 2015	Cardiac surgery centers	Retrospective cohort	Italy (1)	60.6 (8.5)	74.6	All
Martínez-Sellés, 2014	2008 - 2010	Cardiac surgery centers	Registry (GAMES)	Spain (26)	61.4 (15.5)	61.1	All
Olmos, 2017	1996 - 2014	Cardiac surgery centers	Retrospective cohort	Spain (3)	62 (14)	61.1	A/M
30 days mortality							
Di Mauro, 2017	2000 - 2015	Cardiac surgery centers	Retrospective cohort	Italy (26)	59.6 (15.1)	81.8	All
Fernández-Hidalgo, 2018	2000 - 2011	Cardiac surgery centers	Retrospective cohort	Spain (9)	58 (15.1)	NI	All

Sd: Standard deviation; Q_1: First quartil; Q_3: Thrird quartil: STS ACSD: The Society of Thoracic Surgeons Adult Cardiac Surgery Database; AEPEI: Association pour l'Etude et la Prevention de l'Endocadite Infectieuse; GAMES: Grupo de Apoyo al Manejo de la Endocarditis infecciosa en España; A: Aortic valve; M: Mitral valve; NI: No information.

Table S4: Definition of the predictors

Preoperative patient-related factors	
Age	**Di Mauro 2017; De Feo 2012; Fernández-Hidalgo 2018; Martínez-Sellés 2014; Olmos 2017; GAMES registry.**
Gender	**Di Mauro 2017; Martínez-Sellés 2014; GAMES registry.**
Renal failure	**Di Mauro 2017.** Creatinine ≥ 2 mg/dl. **Gaca 2011.** Documented history of renal failure and/or history of creatinine > 2 mg/dl. Prior renal transplant patients not included as pre-op renal failure unless since transplantation creatinine creatine values had been > 2.0 mg/dl. **De Feo 2012; GAMES registry.** Creatinine > 2 mg/dl. **Gatti 2017a.** eGFR <50 mL/min/1.73 m2. The creatinine clearance rate calculated according to the Cockcroft–Gault formula was used to estimate GFR. **Fernández-Hidalgo 2018.** Serum creatinine >200 mmol/l preoperatively. **Olmos 2017.** Renal failure was defined as GFR <60 mL/min/1.73 m2.
Body max index	**Gatti 2017a.**
Chronic pulmonary disease	**Di Mauro 2017.** Long term use of bronchodilators or steroids for lung disease. **Gaca 2011; GAMES registry.** Chronic lung disease.
Diabetes Mellitus	**Gaca 2011.** History of IDDM or NIDDM diabetes mellitus. Patients placed on a pre-operative diabetic pathway of Insulin drip but at admission were controlled with none, diet or oral method are not coded as insulin dependent.
Hypertension	**Gaca 2011.** Diagnosis of hypertension, documented by one of the following: a. Documented history of hypertension diagnosed and treated with medication, diet and/or exercise. b. Prior documentation of systolic blood pressure >140 mmHg or diastolic blood pressure > 90 mmHg for patients without diabetes or chronic kidney disease, or prior documentation of systolic blood pressure >130 mmHg or diastolic blood pressure > 80 mmHg on at least 2 occasions for patients with diabetes or chronic kidney disease. c. Currently on pharmacologic therapy to control hypertension.
Pulmonary hypertension	**Gatti 2017a.** Systolic pulmonary artery pressure > 55mmHg. **Fernández-Hidalgo 2018; GAMES registry.** Systolic pulmonary artery pressure > 60 mmHg.
Anemia	**Gatti 2017b.** Haemoglobin <12 g/dl for women and <13 g/dl for men.
Thrombocytopaenia	**Olmos 2017.** Platelet count <150.000/mL.
Left ventricular ejection fraction	**Di Mauro 2017; GAMES registry.** Percentage of left ventricular ejection fraction.
Arrhythmia	**Gaca 2011.** History of preoperative arrhythmia (sustained ventricular tachycardia, ventricular fibrillation, atrial fibrillation, atrial flutter, third degree heart block) treated with any of the following modalities: ablation therapy, AICD, pacemaker, pharmacological treatment or electrocardioversion.
Prior cardiac surgery	**Gaca 2011.** Prior CABG or prior valve surgery (i.e. previous surgical replacement and/or surgical repair of a cardiac valve, including percutaneous valve procedures). **Fernández-Hidalgo 2018; GAMES registry.** One or more previous major cardiac operations involving opening the pericardium.
Clinical presentation-related factors	
Critical preoperative state	**Di Mauro 2017; Gatti 2017a; Gatti 2017b; Fernández-Hidalgo 2018; GAMES registry.** Any one or more of the following: ventricular tachycardia or fibrillation or aborted sudden death, preoperative cardiac massage, preoperative ventilation before arrival in the anesthetic room, preoperative inotropic support, intra-aortic balloon counter pulsation or preoperative acute renal failure (anuria or oliguria,10 ml/h). **Gaca 2011.** Patient placed on IABP or received IV inotropic agents within 48 hours preceding surgery.

	De Feo 2012. (Ventilatory support in original paper) Patients admitted to the Cardiac Surgery Department on mechanical ventilation (intubated) or requiring ventilatory support by noninvasive ventilation during preoperative stay (generally for poor hemodynamic conditions and/or pulmonary edema). **Olmos 2017. (**Cardiogenic shock in original paper) Systolic pressure <90 mmHg and tissue hypoperfusion due to myocardial dysfunction, despite adequate preload, and accompanied by low cardiac index and high pulmonary wedge pressure.
NYHA functional class	**De Feo 2012; Gatti 2017a; Gatti 2017b; Fernández-Hidalgo 2018; GAMES registry.** NYHA classification for dyspnea: I: no symptoms on moderate exertion; II: symptoms on moderate exertion; III: symptoms on light exertion; IV: symptoms at rest.
Septic shock	**Olmos 2017.** Acute circulatory failure in sepsis, with persistent systolic pressure <90 mmHg despite adequate volume resuscitation.
EuroSCORE I	**Martínez-Sellés 2014.** European system for cardiac operative risk evaluation I. Nashef 1999.
EuroSCORE II	**Madeira 2016.** European system for cardiac operative risk evaluation II. Nashef 2011.
Surgery-related factors	
Paravalvular complications	**De Feo 2012.** Presence of either an annular abscess or aortocavitary fistula. **Di Mauro 2017.** Presence of an abscess. **Fernández-Hidalgo 2018.** Presence of a fistula. **Martínez-Sellés 2014.** (Substantial intracardiac destruction in original paper) Abscesses present or echocardiography findings suggestive of invasive infection (communication between chambers, wall dissection or large valvular dehiscence). **Olmos 2017.** Presence of abscess, pseudoaneurysm, fistula or prosthetic dehiscence. **GAMES registry.** purulent cavity with necrosis and capacity to invade adjacent structures.
Urgency of procedure	**Gaca 2011.** Urgent status: procedure required during the same hospitalization to minimize chance of further clinical deterioration; Emergency status: patient requiring emergency operations will have ongoing, refractory (difficult, complicated, and/or unmanageable) unrelenting cardiac compromise, with or without hemodynamic instability, and not responsive to any form of therapy except cardiac surgery. An emergency operation is one in which there should be no delay in providing operative intervention. **Fernández-Hidalgo 2018.** Urgent status: patients not electively admitted for operation but who require surgery on the current admission for medical reasons and cannot be discharged without a definitive procedure; Emergency status: operation before the beginning of the next working day after decision to operate. **Martínez-Sellés 2014.** Definition not available. **GAMES registry.** Urgent surgery: surgery required within 24 h of its indication; Emergency surgery: surgery required on the day of admission.
Number of treated valves/prostheses	**Di Mauro 2017; Gaca 2011; GAMES registry.** Number of treated valves/prostheses.
Weight of intervention	**Gatti 2017b.** Surgery on thoracic aorta.
IE-related factors	
Infection etiology	Pathogen isolated on blood or specimen culture. **Di Mauro 2017.** *Pseudomonas aeruginosa*; *Staphylococcus aureus*; Fungi; Other microorganisms. **Fernández-Hidalgo 2018; Martínez-Sellés 2014.** *Staphylococcus* spp. **Olmos 2017**. *Staphylococcus aureus* or fungi.

	GAMES registry. *Staphylococcus* spp. (coagulase-negative staphylococci or S. *aureus*); *Pseudomonas* spp.; Fungal disease; *Streptococcus* spp.; Other microorganisms.
Type of valve	**Madeira 2016; Olmos 2017.** Not available. **Martínez-Sellés 2014.** Prosthetic valve IE was defined as infection occurring on any type of non-native tissue or mechanical device.
Active endocarditis	**Gaca 2011** Type of endocarditis the patient has. If the patient is currently being treated for endocarditis, the disease is considered active. If no antibiotic medication (other than prophylactic medication) is being given at the time of surgery, then the infection is considered treated.
Valve location	**Fernández-Hidalgo 2018.** Infection location (aortic, mitral, other). **Games registry.** Infection location (aortic, mitral, pulmonary, tricuspid).
Positivity of latest pre-op. blood culture	**De Feo 2012.** Operation without possibility of previous attainment of negative cultures by antibiotic therapy (latest culture had always been performed within 5 to 7 days preoperatively).

eGFR: estimated glomerular filtration rate; **GFR**: glomerular filtration rate; **IDDM**: insulin-dependent diabetes mellitus; **NIDDM**: non-insulin-dependent; AICD: CABG: coronary artery bypass grafting; IABP: Intra-Aortic Balloon Pump; NYHA: New York Heart Association; **GAMES**: Grupo de Apoyo al Manejo de la Endocarditis infecciosa en España.

96

Table S5. Model compositions and percentage of missing data in GAMES registry.

	De Feo, 2012	Gaca, 2011	Madeira, 2016	Gatti, 2017a (Original)	Gatti, 2017a (Alternate)	Gatti, 2017b	Martínez-Sellés, 2014	Olmos, 2017	Di Mauro, 2017 (EndoSCORE)	Fernández-Hidalgo, 2018 (sp. ES-I)	Fernández-Hidalgo, 2018 (sp. ES-II)	Meta-model	Percentage of missing data in GAMES
Patient-related factor													
Renal failure	■	■		■	■			■	■	■		■	4%
Age (years)	■						■	■	■	■	■	■	0%
Prior cardiac surgery		■								■	■	■	1%
Gender							■		■			■	<1%
Chronic pulmonary disease		■							■			■	11%
Pulmonary hypertension				■						■		■	0%
Anemia						■							100%
BMI (kg/m)				■									29%
Diabetes Mellitus		■											<1%
Hypertension		■											<1%
Arrhythmia		■											<1%
Left ventricular ejection fraction (%)									■			■	7%
Thrombocytopaenia								■					100%
Clinical presentation-related factors													
Critical preoperative state	■	■		■	■	■		■	■	■	■	■	2%
NYHA functional class	■			■	■	■				■	■	■	2%
Septic shock								■					0%
EuroSCORE I							■						19%
EuroSCORE II			■										37%
Surgery-related factors													
Paravalvular complications	■						■	■	■	■	■	■	4%
Urgency of procedure		■					■			■	■	■	2%
Number of treated valves/ prostheses		■							■			■	0%
Weight of intervention						■							0%
IE-related factors													
Infection etiology							■	■	■	■	■	■	4%
Type of valve			■				■	■					1%
Valve location										■	■	■	0%
Active endocarditis		■											0%
Positivity of latest pre-op. blood culture	■												0%

RoB: Risk of Bias; GAMES: Grupo de Apoyo al Manejo de la Endocarditis infecciosa en España; BMI: body mass index; NYHA: New York Heart Association; IE: infective endocarditis; pre-op: pre-operative.
Dark cells indicate that the predictor was included in the model.

Table S6: Minimum sample size for development of a new multivariable prediction model.

Author, Year	Available data			Minimum Sample Size[a]/EPP required for development of a new multivariable prediction model		
	Events n (%)	Candidate predictors	Sample size/EPP	Explained variability scenarios		
				10%	20%	30%
De Feo, 2012	40 (9.1)	19	440 / 2.1	3,651 / 17.5	1,777 / 8.5	1,152 / 5.5
Gaca, 2011	1,117 (8.2)	38	13,617 / 29.4	7,709 / 16.6	3,757 / 8.1	2,439 / 5.3
Madeira, 2016	21 (16.4)	15	128 / 1.4	2,211 / 24.2	1,067 / 11.7	685 / 7.5
Gatti, 2017a (Original & Alternate)	56 (15.5)	57	361 / 1.0	8,589 / 23.4	4,147 / 11.3	2,664 / 7.2
Gatti, 2017b	28 (20.3)	56	138 / 0.5	7,649 / 27.7	3,679 / 13.3	2,353 / 8.5
Martínez-Sellés, 2014	106 (24.3)	NI	437 / NI	n.a.	n.a.	n.a.
Olmos, 2017	124 (29.2)	37	424 / 3.4	4,562 / 36.0	2,185 / 17.2	1,390 / 11.0
Di Mauro, 2017	298 (11.0)	32	2,715 / 9.3	5,600 /19.2	2,718 / 9.3	1,756 / 6.0
Fernández-Hidalgo, 2018 (Sp. ES-I)	208 (26.7)	26	779 / 8.0	3,277 / 33.6	1,571 / 16.1	1,001 / 10.3
Fernández-Hidalgo 2018 (Sp. ES-II)	208 (26.7)	27	779 / 7.7	3,403 / 33.6	1,631 / 16.3	1,039 / 10.3

Sp. ES-I: specific EuroSCORE I; Sp. ES-II: specific EuroSCORE II; n: number of events; EPP: events per parameter; NI: not informed; n.a.: not applicable.

[a] We calculated the minimum sample size required for the development of a new multivariable prediction model using the criteria proposed by Riley et al. (1). We used the number of candidate predictors and mortality rates from the original paper, and we considered three different scenarios for the variability explained by the model (10%, 20% or 30%). Prediction models with C-statistics between 0.7 and 0.8 typically have R-squared values between 10 and 20% (2) and were models which reported C-statistic close to 0.9. For a mortality proportion of 0.2. the max(R^2_{CS}) is 0.63 (1), therefore for the 10% explained variability scenario R^2_{CS} = 0.63*0.10 = 0.063.
R^2_{CS}: Cox-Snell R-squared

We used pmsampsize stata command developed by Riley R. and Ensor J.

1. 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. Statistics in Medicine. 2019 Mar 30;38(7):1276–96.

2. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating [Internet]. 2019 [cited 2020 Apr 28]. Available from: https://doi.org/10.1007/978-3-030-16399-0

Table S7: Prognostic models equation

De Feo 2012	(No constant) 0.041xage + 1.076 (if renal failure) + 1.777 (if NYHA class IV) + 2.281 (if critical preoperative state) + 1.093 (if positivity of latest pre-op. blood culture) + 1.110 (if paravalvular complications)
Di Mauro 2017	-2.60 + 0.46 (if age 60-70y) + 0.88 (if age 70-80y) + 1.53 (if age>80y) + 0.51 (if female) – 0.03xLVEF + 0.50 (if renal failure) + 0.68 (if chronic pulmonary disease) + 1.46 (if critical preoperative state) + 0.50 (if two valves/prostheses treated) + 1.50 (if three valves/prostheses treated) + 1.09 (if paravalvular complications) + 1.46 (if *Pseudomonas aeruginosa*) + 1.24 (if *Staphylococcus aureus*) + 1.66 (if fungi) + 0.60 (if other microorganisms)
Fernández-Hidalgo 2018	**Specific ES-I:** -3.132 + 1.101 (if prior cardiac surgery) + 1.121 (if critical preoperative state) + 0.464 (if renal failure) + 0.702 (if NYHA class > 1) + 0.059x(age-60) (if age > 60y) + 0.806 (if emergency status) + 1.220 (if paravalvular complications) + 0.528 (if *Staphylococcus* spp.) – 1.268 (if pulmonary hypertension) + 0.374 (if mitral location) **Specific ES-II:** -4.210 + 0.964 (if prior cardiac surgery) + 1.024 (if critical preoperative state) + 0.617 (if NYHA class > 1) + 0.062x(age-60) (if age > 60y) + 1.950 (if emergency status) + 1.157 (if urgent status) + 1.141 (if paravalvular complications) + 0.531 (if *Staphylococcus* spp.) + 0.383 (if mitral location)
Gaca 2011	(No constant) 0.490 (if Prior CABG) + 0.422 (if urgent status) + 1.153 (if cardiogenic shock) + 0.672 (if critical preoperative state) + 0.602 (if multiple valve procedure) + 0.471 (if prior valve surgery) + 0.547 (if IDDM) + 0.431 (if NIDDM) + 0.342 (if hypertension) + 0.344 (if chronic pulmonary disease) + 0.695 (if active endocarditis) + 0.827 (if renal failure) + 0.504 (if arrhythmia)
Gatti 2017a	**Original:** -3.065 + 0.58 (if BMI > 27kg/m^2) + 1.26 (if renal failure) + 0.75 (if NYHA class IV) + 0.58 (if pulmonary hypertension) + 0.86 (if critical preoperative state) **Alternate:** -1.411 + 1.32 (if renal failure) + 0.75 (if NYHA class IV) + 0.85 (if critical preoperative state)
Gatti 2017b	**Preoperative:** (No constant) 2.40 (if anemia) + 0.96 (if NYHA class IV) + 1.60 (if critical preoperative state) + 1.86 (if paravalvular complications) + 2.02 (if surgery on thoracic aorta)
Madeira 2016	(No constant) 1.932 (if prosthetic valve IE) + 0.081xEuroSCORE-II
Martínez-Sellés 2014	(No constant) 0.030xage + 0.790 (if prosthetic valve IE) + 0.640 (if paravalvular complications) + 0.740 (if female) + 0.690 (if urgent status) + 0.830 (if *Staphylococcus* spp.) + 0.02xEuroSCORE-I
Olmos 2017	-3.358 + 0.916 (if age 52-63y) + 1.336 (if age 64-72y) + 1.362 (if age≥73y) + 0.645 (if prosthetic endocarditis) + 0.903 (if *Staphylococcus aureus* or fungi) + 0.702 (if septic shock) + 0.655 (if thrombocytopenia) + 0.542 (if renal failure) + 1.486 (if cardiogenic shock) + 0.541 (if paravalvular complications)

NYHA: New York Heart Association; LVEF: Left ventricular ejection fraction; CABG: Coronary artery bypass graft; IDDM: insulin-dependent diabetes mellitus; NIDDM: non-insulin-dependent diabetes mellitus; BMI: Body mass index: IE: Infective endocarditis.

Box S1: meta-model equation and example of use

The equation of the meta-model to estimate probability of mortality in patient with infective endocarditis is as follows:

$$P(mortality) = \frac{\exp(Y)}{1 + \exp(Y)}$$

where Y = −5.00 + 0.22 [if female] + 0.045 * age + 0.28 [if renal failure] + 0.51 [if prior cardiac surgery] + 0.29 [if chronic pulmonary disease] + 0.17 [if pulmonary hypertension] - (0.013 * LVEF) + 1.17 [if critical preoperative state] + 0.33 [if NYHA>I] + 0.43 [if abscess] + 0.59 [if fistulae] + 0.44 [if urgent status] + 0.85 [if emergency status] + 0.22 [if two valves treated] + 0.65 [if three valves treated] + 0.19 [if mitral location] + 0.64 [if *Staphylococcus spp.*] + 0.61 [if Fungi]

Example:

A 60-year-old woman with renal failure and pulmonary hypertension, with a left ventricular ejection fraction of 60%, NYHA-II, with paravalvular abscess. The preoperative condition is not critical, but the patient must undergo urgent surgery. Infective endocarditis is located in the aortic valve and was caused by *Staphylococcus spp*.

Y = −5.00 + 0.22 [female] + 0.045*60 + 0.28 [renal failure] + 0.17 [pulmonary hypertension] - (0.013*60) + 0.33 [NYHA=II] + 0.43 [abscess] + 0.44 [urgent surgery] + 0.64 [Staphylococcus spp.] = -0.57

$$P(mortality) = \frac{\exp(-0.57)}{1 + \exp(-0.57)} \approx 36\%$$

LVEF: left ventricular ejection fraction; NYHA. New York Hearth Assotiation

1

Table S8: Critical appraisal using PROBAST.

Domain	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Model information											
Author, Year and Model name	De Feo, 2012	Gaca, 2011, STSS score	Madeira, 2016	Gatti, 2017a, AEPEI original	Gatti, 2017a, AEPEI alternate	Gatti, 2017b, ANCLA	Martínez-Sellés, 2014, PALUSE	Olmos, 2017, RISK-E	Di Mauro, 2017, EndoSCORE	Fernández-Hidalgo, 2018, sp.ES-I	Fernández-Hidalgo, 2018, sp. ES-II
1. Participants											
Risk of Bias	**High**	**High**	**Unclear**	**Low**	**Low**	**Low**	**Low**	**Low**	**Unclear**	**Low**	**Low**
Applicability	**High**	**Low**	**Unclear**	**Low**	**Low**	**Low**	**Low**	**Low**	**Unclear**	**Low**	**Low**
1.1 Were appropriate data sources used?											
	PY	PY	PY	PY	PY	PY	PY	PY	PY	PY	PY
1.2 Were all inclusions and exclusions of participants appropriate?											
	N	PN	NI	PY	PY	PY	PY	PY	NI	PY	PY
Observations:	**De Feo, 2012:** The model was developed in a subgroup of patients. These participants represent a selected lower (or higher) risk sample of the original. **Gaca, 2011:** Excluded complete sites if data were missing in some variables, likely to have introduced bias but less important than excluding individual participants. **Madeira, 2016; Di Mauro, 2017:** No information about recruitment methods and exclusion criteria. **Gatti, 2017a:** The model was developed using only data from 2008 in France, because the data collection was particularly exhaustive, comprehensive, and accurate, we did not consider it could introduce bias.										
2. Predictors											
Risk of Bias	**Unclear**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**
Applicability	**Low**	**Low**	**Low**	**Unclear**	**Low**	**Low**	**Low**	**Low**	**Low**	**Unclear**	**Low**
2.1 Were predictors defined and assessed in a similar way for all participants?											
	Y	Y	Y	PY	PY	Y	Y	Y	Y	N	N
2.2 Were predictor assessments made without knowledge of outcome data?											
	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI
2.3 Are all predictors available at the time the model is intended to be used?											
	Y	Y	Y	PN	Y	Y	Y	Y	Y	Y	Y
Observations:	No author informed if predictor assessments was make without knowledge of outcome data, although we didn't penalized RoB if predictors assessed had an objective interpretation. **De Feo, 2012:** There were predictors assessed with subjective interpretation. **Gatti, 2017a (original); Fernández-Hidalgo. 2017 (ES-I):** Systolic pulmonary artery pressure predictor could be hard to recovery. **Fernández-Hidalgo. 2017:** Databases were not homogeneous, but authors did an effort to homogenize it, we did not penalized the RoB.										
3. Outcome											
Risk of Bias	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**
Applicability	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**	**Low**
3.1 Was the outcome determined appropriately?											
	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
3.2 Was a pre-specified or standard outcome definition used?											
	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
3.3 Were predictors excluded from the outcome definition?											
	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?											

Domain	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Model information											
Author, Year and Model name	De Feo, 2012	Gaca, 2011, STSS score	Madeira, 2016	Gatti, 2017a, AEPEI original	Gatti, 2017a, AEPEI alternate	Gatti, 2017b, ANCLA	Martínez-Sellés, 2014, PALUSE	Olmos, 2017, RISK-E	Di Mauro, 2017, EndoSCORE	Fernández-Hidalgo, 2018, sp.ES-I	Fernández-Hidalgo, 2018, sp. ES-II
	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
3.5 Was the outcome determined without knowledge of predictor information?											
	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI
3.6 Was the time interval between predictor assessment and outcome determination appropriate?											
	PY	PY	PY	PY	PY	PY	PY	PY	PY	PY	PY
Observations:	When the outcome is a hard variable which do not required interpretation such as mortality, previous knowledge of predictor information does not introduce RoB.										
4. Analysis											
Risk of Bias	**High**	**High**	**High**	**High**	**High**	**High**	**High**	**High**	**Unclear**	**Low**	**Low**
4.1 Were there a reasonable number of participants with the outcome?											
	N	Y	N	N	N	N	NI	N	PN	PN	PN
4.2 Were continuous and categorical predictors handled appropriately?											
	N	PY	PN	N	N	N	PN	N	PY	PY	PY
4.3 Were all enrolled participants included in the analysis?											
	PN	PN	PY	NI	PN	PN	PY	N	PY	PN	PN
4.4 Were participants with missing data handled appropriately?											
	NI	PN	NI	NI	NI	NI	NI	NI	NI	NI	NI
4.5 Was selection of predictors based on univariable analysis avoided?											
	N	N	N	N	N	N	N	N	N	Y	Y
4.6 Were complexities in the data accounted for appropriately?											
	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
4.7 Were relevant model performance measures evaluated appropriately?											
	N	PN	N	Y	PN	PN	N	Y	Y	Y	Y
4.8 Were model overfitting and optimism in model performance accounted for?											
	N	N	N	N	N	N	N	N	Y	Y	Y
4.9 Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?											
	PN	Y	Y	Y	Y	Y	PN	Y	Y	Y	Y
Observations:	**De Feo, 2012:** Very small number of events per parameter (EPP), continuous predictors not handled appropriately, probably using complete data and only apparent validation available. **Gaca, 2011:** Large EPP (aprox. 30), but predictors selected based on univariable analysis, random splitting validation (D:70% and V:30%) and no inform how missing data were handled. **Madeira, 2016:** Very small EPP and apparent validation. **Gatti 2017a; Gatti, 2017b:** Very small EPP, predictors selected based on univariable analysis and continuous predictors categorized based on the best discriminative performance. **Martínez-Sellés, 2014:** EPP not available, no informed about missing data, continuous predictors dichotomized and apparent performance. **Olmos, 2017:** Very small EPP and random splitting validation (D:70% and V:30%) and did not inform neither missing data nor continuous predictors were handled. **Di Mauro, 2017:** Predictors were selected based on univariable analysis ($p<0.2$). Although EPP was sufficiently large and model performance was optimism adjusted, unfortunately calibration measures were tested but not reported, thus we rated it as unclear RoB. **Fernández-Hidalgo, 2018:** EPP was slightly lower than required but were not univariable selection and model performance was optimism adjusted by bootstrap validation. The complete data analysis were not worrying because only 4 (0.5%) patients were excluded. We did not penalized RoB.										

Y: Yes; PY: Probably yes; N: No; PN: Probably no; NI: No information

Figure S1: Summary of risk of bias and applicability of the studies

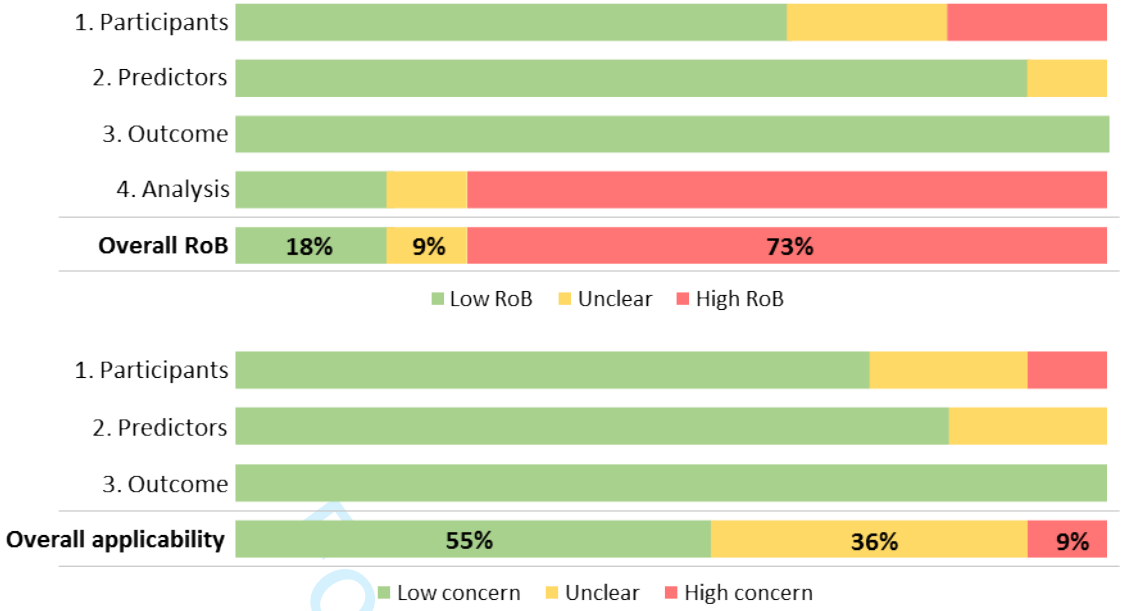
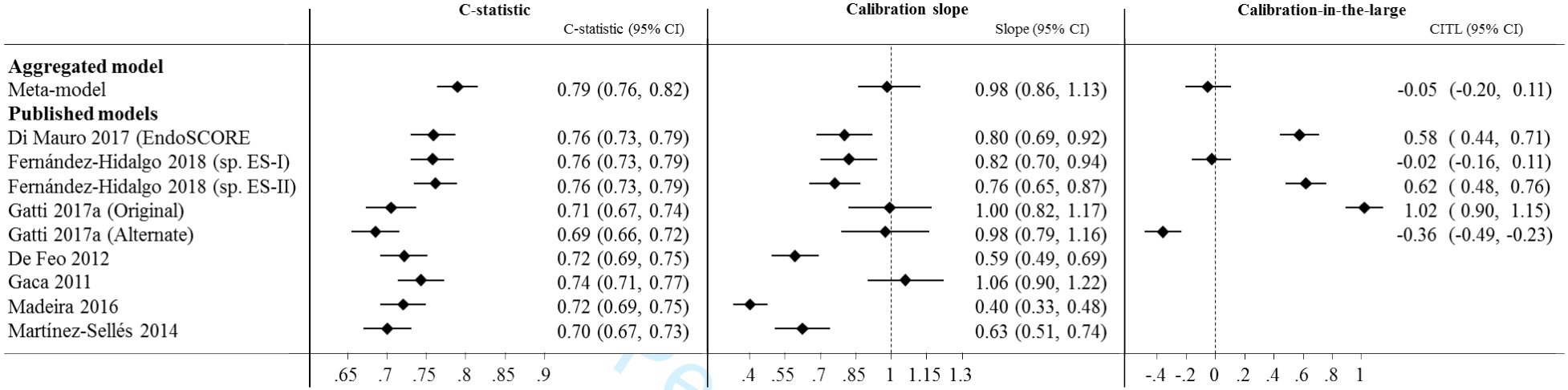
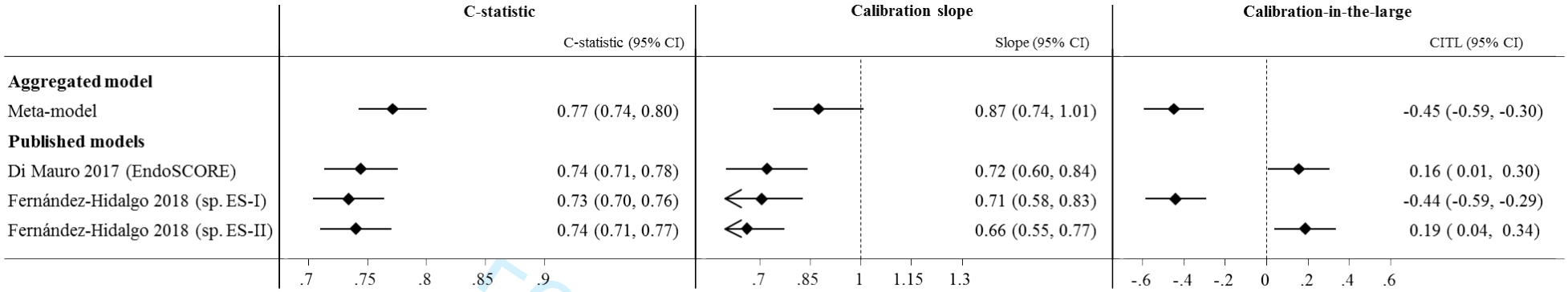


Figure S2: Validation of all models regardless of critical appraisal.



Dashed lines indicate lines of perfect calibration slope (1) and calibration-in-the-large (0). Black diamonds indicate point estimates and horizontal bars indicate 95% CIs.

Figure S3: Validation of the meta-model and existing models selected for aggregation for 30-days mortality outcome.



Dashed lines indicate lines of perfect calibration slope (1) and calibration-in-the-large (0). Black diamonds indicate point estimates and horizontal bars indicate 95% CIs.

S6: Members of GAMES group

Hospital Costa del Sol, (Marbella): Fernando Fernández Sánchez, Mariam Noureddine, Gabriel Rosas, Javier de la Torre Lima; **Hospital Universitario de Cruces**, (Bilbao): Elena Bereciartua, Roberto Blanco, María Victoria Boado, Marta Campaña Lázaro, Alejandro Crespo, Laura Guio Carrión, Mikel Del Álamo Martínez de Lagos, Gorane Euba Ugarte, Josune Goikoetxea, Marta Ibarrola Hierro, José Ramón Iruretagoyena, Josu Irurzun Zuazabal, Leire López-Soria, Miguel Montejo, Javier Nieto, David Rodrigo, Regino Rodríguez, Yolanda Vitoria, Roberto Voces; **Hospital Universitario Virgen de la Victoria**, (Málaga): Mª Victoria García López, Radka Ivanova Georgieva, Guillermo Ojeda, Isabel Rodríguez Bailón, Josefa Ruiz Morales; **Hospital Universitario Donostia-Poliklínica Gipuzkoa-IIS Biodonostia**, (San Sebastián): Harkaitz Azkune Galparsoro, Elisa Berritu Boronat, Mª Jesús Bustinduy Odriozola, Cristina del Bosque Martín, Tomás Echeverría, Alberto Eizaguirre Yarza, Ana Fuentes, Miguel Ángel Goenaga, Muskilda Goyeneche del Río, Ángela Granda Bauza, José Antonio Iribarren, Xabier Kortajarena Urkola, José Ignacio Pérez-Moreiras López, Ainhoa Rengel Jiménez, Karlos Reviejo, Alberto Sáez Berbejillo, Elou Sánchez Haza, Rosa Sebastián Alda, Itziar Solla Ruiz, Irati Unamuno Ugartemendia, Diego Vicente Anza, Iñaki Villanueva Benito, Mar Zabalo Arrieta; **Hospital General Universitario de Alicante**, (Alicante): Rafael Carrasco, Vicente Climent, Patricio Llamas, Esperanza Merino, Joaquín Plazas, Sergio Reus; **Complejo Hospitalario Universitario A Coruña**, (A Coruña): Nemesio Álvarez, José María Bravo-Ferrer, Laura Castelo, José Cuenca, Pedro Llinares, Enrique Miguez Rey, María Rodríguez Mayo, Efrén Sánchez, Dolores Sousa Regueiro; **Complejo Hospitalario Universitario de Huelva**, (Huelva): Francisco Javier Martínez; **Hospital Universitario de Canarias**, (Canarias): Mª del Mar Alonso, Beatriz Castro, Teresa Delgado Melian, Javier Fernández Sarabia, Dácil García Rosado, Julia González González, Juan Lacalzada, Lissete Lorenzo de la Peña, Alina Pérez Ramírez, Pablo Prada Arrondo, Fermín Rodríguez Moreno; **Hospital Regional Universitario de Málaga**, (Málaga): Antonio Plata Ciezar, José Mª Reguera Iglesias; **Hospital Universitario Central Asturias**, (Oviedo): Víctor Asensi Álvarez, Carlos Costas, Jesús de la Hera, Jonnathan Fernández Suárez, Lisardo Iglesias Fraile, Víctor León Arguero, José López Menéndez, Pilar Mencia Bajo, Carlos Morales, Alfonso Moreno Torrico, Carmen Palomo, Begoña Paya Martínez, Ángeles Rodríguez Esteban, Raquel Rodríguez García, Mauricio Telenti Asensio; **Hospital Clínic-IDIBAPS, Universidad de Barcelona**, (Barcelona): Manuel Almela, Juan Ambrosioni, Manuel Azqueta, Mercè Brunet, Marta Bodro, Ramón Cartañá, Carlos Falces, Guillermina Fita, David Fuster, Cristina García de la Mària, Delia García-Pares, Marta Hernández-Meneses, Jaume Llopis Pérez, Francesc Marco, José M. Miró, Asunción Moreno, David Nicolás, Salvador Ninot, Eduardo Quintana, Carlos Paré, Daniel Pereda, Juan M. Pericás, José L. Pomar, José Ramírez, Irene Rovira, Elena Sandoval, Marta Sitges, Dolors Soy, Adrián Téllez, José M. Tolosana, Bárbara Vidal, Jordi Vila; **Hospital General Universitario Gregorio Marañón**, (Madrid): Iván Adán, Juan Carlos Alonso, Ana Álvarez-Uría, Javier Bermejo, Emilio Bouza, Gregorio Cuerpo Caballero, Antonia Delgado Montero, Ana González Mansilla, Mª Eugenia García Leoni, Esther Gargallo, Víctor González Ramallo, Martha Kestler Hernández, Amaia Mari Hualde, Marina Machado, Mercedes Marín, Manuel Martínez-Sellés, Patricia Muñoz, María Olmedo, Álvaro Pedraz, Blanca Pinilla, Ángel Pinto, Cristina Rincón, Hugo Rodríguez-Abella, Marta Rodríguez-Créixems, Antonio Segado, Neera Toledo, Maricela Valerio, Pilar Vázquez, Eduardo Verde Moreno; **Hospital Universitario La Paz**, (Madrid): Isabel Antorrena, Belén Loeches, Mar Moreno, Ulises Ramírez, Verónica Rial Bastón, María Romero, Sandra Rosillo; **Hospital Universitario Marqués de Valdecilla**, (Santander): **Hospital Universitario Marqués de Valdecilla**, (Santander): Jesús Agüero Balbín, Cristina Amado, Carlos Armiñanzas Castillo, Ana Arnaiz, Francisco Arnaiz de las Revillas, Manuel Cobo Belaustegui, María Carmen Fariñas, Concepción Fariñas-Álvarez, Marta Fernández Sampedro, Iván García, Claudia González Rico, Laura Gutierrez-Fernandez , Manuel Gutiérrez-Cuadra, José Gutiérrez Díez, Marcos Pajarón, José Antonio Parra, Ramón Teira, Jesús Zarauza; **Hospital Universitario Puerta de Hierro**, (Madrid): Jorge Calderón Parra, Marta Cobo, Fernando Domínguez, Pablo García Pavía, Ana Fernández Cruz, Antonio Ramos-Martínez, Isabel Sánchez Romero; **Hospital Universitario Ramón y Cajal**, (Madrid): Tomasa Centella, José Manuel

Hermida, José Luis Moya, Pilar Martín-Dávila, Enrique Navas, Enrique Oliva, Alejandro del Río, Jorge Rodríguez-Roda Stuart, Soledad Ruiz; **Hospital Universitario Virgen de las Nieves**, (Granada): Carmen Hidalgo Tenorio; **Hospital Universitario Virgen Macarena**, (Sevilla): Manuel Almendro Delia, Omar Araji, José Miguel Barquero, Román Calvo Jambrina, Marina de Cueto, Juan Gálvez Acebal, Irene Méndez, Isabel Morales, Luis Eduardo López-Cortés; **Hospital Universitario Virgen del Rocío**, (Sevilla): Arístides de Alarcón, Encarnación Gutiérrez-Carretero, José Antonio Lepe, José López-Haldón, Rafael Luque-Márquez, Guillermo Marín, Antonio Ortiz-Carrellán, Eladio Sánchez-Domínguez; **Hospital San Pedro**, (Logroño): Luis Javier Alonso, Pedro Azcárate, José Manuel Azcona Gutiérrez, José Ramón Blanco, Antonio Cabrera Villegas, Lara García-Álvarez, Concepción García García, José Antonio Oteo; **Hospital de la Santa Creu i Sant Pau**, (Barcelona): Natividad de Benito, Mercé Gurguí, Cristina Pacho, Roser Pericas, Guillem Pons; **Complejo Hospitalario Universitario de Santiago de Compostela**, (A Coruña): M. Álvarez, A. L. Fernández, Amparo Martínez, A. Prieto, Benito Regueiro, E. Tijeira, Marino Vega; **Hospital Santiago Apóstol**, (Vitoria): Andrés Canut Blasco, José Cordo Mollar, Juan Carlos Gainzarain Arana, Oscar García Uriarte, Alejandro Martín López, Zuriñe Ortiz de Zárate, José Antonio Urturi Matos; **Hospital SAS Línea de la Concepción**, (Cádiz): Sánchez-Porto Antonio, Úbeda Iglesias Alejandro; **Hospital Clínico Universitario Virgen de la Arrixaca** (Murcia): José Mª Arribas Leal, Elisa García Vázquez, Alicia Hernández Torres, Ana Blázquez, Gonzalo de la Morena Valenzuela; **Hospital de Txagorritxu**, (Vitoria): Ángel Alonso, Javier Aramburu, Felicitas Elena Calvo, Anai Moreno Rodríguez, Paola Tarabini-Castellani; **Hospital Virgen de la Salud,** (Toledo): Eva Heredero Gálvez, Carolina Maicas Bellido, José Largo Pau, Mª Antonia Sepúlveda, Pilar Toledano Sierra, Sadaf Zafar Iqbal-Mirza; **Hospital Rafael Méndez,** (Lorca-Murcia):, Eva Cascales Alcolea, Ivan Keituqwa Yañez, Julián Navarro Martínez, Ana Peláez Ballesta; **Hospital Universitario San Cecilio** (Granada): Eduardo Moreno Escobar, Alejandro Peña Monje, Valme Sánchez Cabrera, David Vinuesa García; **Hospital Son Llátzer** (Palma de Mallorca): María Arrizabalaga Asenjo, Carmen Cifuentes Luna, Juana Núñez Morcillo, Mª Cruz Pérez Seco, Aroa Villoslada Gelabert; **Hospital Universitario Miguel Servet** (Zaragoza): Carmen Aured Guallar, Nuria Fernández Abad, Pilar García Mangas, Marta Matamala Adell, Mª Pilar Palacián Ruiz, Juan Carlos Porres; **Hospital General Universitario Santa Lucía** (Cartagena): Begoña Alcaraz Vidal, Nazaret Cobos Trigueros, María Jesús Del Amor Espín, José Antonio Giner Caro, Roberto Jiménez Sánchez, Amaya Jimeno Almazán, Alejandro Ortín Freire, Monserrat Viqueira González; **Hospital Universitario Son Espases** (Palma de Mallorca): Pere Pericás Ramis, Mª Ángels Ribas Blanco, Enrique Ruiz de Gopegui Bordes, Laura Vidal Bonet; **Complejo Hospitalario Universitario de Albacete** (Albacete): Mª Carmen Bellón Munera, Elena Escribano Garaizabal, Antonia Tercero Martínez, Juan Carlos Segura Luque; **Hospital Universitario Terrassa:** Cristina Badía, Lucía Boix Palop, Mariona Xercavins, Sónia Ibars. **Hospital Universitario Dr. Negrín** (Gran Canaria): Xerach Bosch, Eloy Gómez Nebreda, Ibalia Horcajada Herrera, Irene Menduiña Gallego, Imanol Pulido; **Complejo Hospitalario Universitario Insular Materno Infantil** (Las Palmas de Gran Canaria): Héctor Marrero Santiago, Isabel de Miguel Martínez, Elena Pisos Álamo. **Hospital Universitario 12 de Octubre** (Madrid): Eva Mª Aguilar Blanco, Mercedes Catalán González, María Angélica Corres Peiretti, Andrea Eixerés Esteve, Laura Domínguez Pérez, Santiago de Cossío Tejido, Francisco Galván Román, José Antonio García Robles, Francisco López Medrano, Mª Jesús López Gude, Mª Ángeles Orellana Miguel, Patrick Pilkington, Yolanda Revilla Ostalaza, Juan Ruiz Morales, Sebastián Ruiz Solís, Ana Sabín Collado, Marcos Sánchez Fernández, Javier Solera Rallo, Jorge Solís Martín. **Hospital Universitari de Bellvitge (L'Hospitalet de Llobregat):** Guillermo Cuervo, Francesc Escrihuela-Vidal, Jordi Carratalà, Inmaculada Grau, Sara Grillo, Carmen Ardanuy, Dámaris Berbel, José Carlos Sánchez Salado, Oriol Alegre, Alejandro Ruiz Majoral, Fabrizio Sbraga, Arnau Blasco, Laura Gracia Sánchez, Iván Sánchez-Rodríguez. **Hospital Universitario Fundación Jiménez Díaz** (Madrid): Beatriz Álvarez, Alfonso Cabello Úbeda, Ricardo Fernández Roblas, Miguel Ángel Navas Lobato, Ana María Pello. **Hospital Basurto** (Bilbao): Mireia de la Peña Triguero, Ruth Esther Figueroa Cerón, Lara Ruiz Gómez. **Hospital del Mar** (Barcelona): Mireia Ble, Juan Pablo Horcajada Gallego, Antonio José Ginel, Inmaculada López, Alexandra Mas, Antoni Mestres, Lluís

Molina, Ramón Serrat, Núria Ribas, Francisca Sánchez, Ana Silverio, Marina Suárez, Luisa Sorlí, Lluís Recasens, Manuel Taurón.