Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Prediction of prolonged intensive care unit stay after coronary artery bypass grafting: development and internal validation of preoperative and 24-hour models in a Colombian cohort

David René Rodríguez Lima
Hospital Universitario Mayor, Mederi
Edith Elianna Rodríguez Aparicio
edith.rodriguez@urosario.edu.co

Hospital Universitario Mayor, Mederi
Diana Marcela Acosta Agámez
Universidad del Rosario
Naiby Yulieth Naiby Yulieth
Universidad del Rosario
María José Nieto Quintín
Universidad del Rosario
Brenda Yulieth González Rodríguez
Universidad del Rosario

Research Article

Keywords: Coronary artery bypass, Intensive care units, Length of stay, Clinical prediction model, EuroSCORE II, Internal validation, Decision curve analysis, Latin America

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11118804/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Prolonged intensive care unit (ICU) stay after coronary artery bypass grafting (CABG) drives morbidity and resource use. Existing prediction models were derived mainly in elective, high-income-country cohorts, and it is unclear whether EuroSCORE II identifies these patients. We developed and internally validated two models to predict ICU stay ≥ 7 days, one for use before surgery and one 24 hours after ICU admission and compared them with EuroSCORE II.

Methods

Retrospective cohort of adults undergoing isolated or combined CABG at a university hospital in Bogotá, Colombia (2,600 m above sea level), between 2019 and 2026. Predictors were pre-specified. Model 1 included ten preoperative predictors; Model 2 added cardiopulmonary bypass duration, intraoperative transfusion, cardiogenic shock, PaO_2/FiO_2 on ICU admission and first-day chest-tube drainage to the Model 1 linear predictor. Missing data were multiply imputed. Internal validation used bootstrapping, followed by uniform shrinkage. We assessed discrimination, calibration, decision curves and temporal transportability, and derived a point score.

Results

Of 915 patients, 131 (14.3%) stayed in the ICU for ≥ 7 days; 88.7% of operations were urgent in-hospital procedures. The area under the receiver operating characteristic curve (AUC) was 0.73 (95% CI 0.68–0.78) for Model 1 and 0.76 (95% CI 0.70–0.80) for Model 2 (optimism-corrected 0.70 and 0.73), compared with 0.68 (95% CI 0.63–0.73) for EuroSCORE II. Concomitant cardiac procedures (odds ratio 3.52), older age, reduced left ventricular ejection fraction, renal dysfunction and extracardiac vascular disease were the main preoperative predictors; cardiogenic shock and chest-tube drainage added information at 24 hours. Both models had greater net benefit than recalibrated EuroSCORE II at thresholds of 10–30%. The 24-hour points score separated low- (7.1%), intermediate- (18.5%) and high-risk (47.3%) groups. Baseline risk varied across calendar periods (observed-to-expected ratio for Model 2, 0.58–1.60).

Conclusions

Preoperative and early postoperative data predicted prolonged ICU stay after CABG with moderate discrimination. The 24-hour model discriminated better than EuroSCORE II, whereas the advantage of the preoperative model was modest. The models may support ICU capacity planning and the early targeting of preventive care, but they require external validation and local recalibration before clinical use.

Background

Cardiovascular diseases are the leading cause of death worldwide, accounting for an estimated 19.8 million deaths in 2022, most of them from ischemic heart disease and stroke [1]. Coronary artery bypass grafting (CABG) remains the standard of care for selected patients with coronary artery disease [2]. Management in the intensive care unit (ICU) during the immediate postoperative period is central to recovery and depends on the patient's preoperative condition, the characteristics of the operation and the response to cardiopulmonary bypass (CPB) [3]. ICU length of stay (LOS) after cardiac surgery is closely linked to healthcare costs and postoperative morbidity [4, 5], and patients with a prolonged stay have higher in-hospital and long-term mortality [6, 7].

Several models to predict prolonged ICU LOS after cardiac surgery have been published [4, 8–11]. Their performance varies across settings because case mix, organization of care and ICU discharge policies differ between institutions [4, 12, 13]. The risk factors most consistently associated with prolonged stay are older age, reduced left ventricular ejection fraction (LVEF), chronic obstructive pulmonary disease (COPD), atrial fibrillation, renal dysfunction, anemia, emergency status, cardiogenic shock and the complexity of the procedure [5, 9, 10, 14]. A prolonged stay is usually the consequence of surgical or medical complications such as bleeding, arrhythmias, low cardiac output, vasoplegia, acute kidney injury and infection [15–17].

Most existing models were derived in high-income countries, in cohorts dominated by elective surgery. Their transportability to Latin American centers, where many patients undergo CABG urgently during the same admission as an acute coronary syndrome, is unknown. In our institution, EuroSCORE II showed poor discrimination and calibration for in-hospital mortality after CABG [18], and its ability to identify patients who will need a long ICU LOS has not been evaluated. In addition, a prediction model is only useful if it is tied to a defined moment of use. A preoperative estimate can inform scheduling, bed allocation and patient counselling, whereas an estimate made early after ICU admission can inform capacity planning and the intensity of early postoperative care. Events that occur later in the ICU course, such as postoperative complications, cannot inform either decision.

The aim of this study was to develop and internally validate two nested models to predict ICU LOS of 7 days or longer after CABG, one using information available before surgery (Model 1) and one using information available 24 hours after ICU admission (Model 2); to compare their performance with that of EuroSCORE II; and to derive a simplified point score for bedside use.

Methods

Study design, setting and reporting

We conducted a retrospective cohort study using the institutional cardiac surgery registry of Hospital Universitario Mayor–Méderi, a university hospital in Bogotá, Colombia, located at 2,600 m above sea level. The registry records demographic, clinical, operative and postoperative data and in-hospital

outcomes for all adult cardiac surgery procedures in a REDCap database [19]. By institutional protocol, all patients are admitted to the ICU after cardiac surgery. Most patients undergoing CABG at our center are referred as inpatients after an acute coronary syndrome and are operated on during the same hospital admission; according to EuroSCORE II definitions, these operations are classified as urgent [20]. The study is reported in accordance with the TRIPOD + AI statement [21] (Additional file 2).

Participants

We included all adults who underwent CABG, isolated or combined with another cardiac procedure (valve, ascending aorta or other), between January 2019 and February 2026. We excluded patients who died in the operating room or recovery area before ICU admission, patients with missing ICU LOS, and records in which ICU LOS exceeded total hospital LOS, which we considered implausible data-entry errors (those with complete predictor data were re-included in a sensitivity analysis). In-hospital vital status was available for all patients.

Outcome

The outcome was prolonged ICU stay, defined as an ICU LOS of 7 days or longer after surgery. ICU LOS was recorded in whole days from ICU admission after surgery to ICU discharge or death. There is no consensus definition of prolonged ICU stay, and published thresholds range from 24 hours to 14 days [5, 12]. We chose 7 days a priori because shorter thresholds (24–72 hours) mainly reflect local discharge policies and bed scheduling [4, 22], whereas a stay of 7 days or longer identifies patients with major complications and high resource use, has been used in large derivation and validation cohorts, including CABG-specific cohorts [6, 9, 14], and is less dependent on local practice. In the primary analysis, patients who died in the ICU before day 7 were classified according to their ICU stay (not prolonged), reflecting a resource-use perspective. Because death precludes a prolonged stay, we examined this choice in sensitivity analyses.

Predictors and time of prediction

Candidate predictors were pre-specified on the basis of previous models and systematic reviews [5, 8–10, 14], clinical plausibility, and availability in the registry with an unambiguous time of measurement. No univariable screening or automated selection was used. Definitions, timing and rationale are given in Additional file 1 (Table S1).

Model 1 (preoperative) is intended for use after the heart-team decision to operate, to support scheduling, bed allocation and counselling. Its ten predictors were age, sex, LVEF, serum creatinine, COPD, extracardiac vascular disease (peripheral arterial disease or previous stroke), atrial fibrillation or flutter, critical preoperative state (EuroSCORE II definition, including preoperative intra-aortic balloon pump), concomitant cardiac procedure (any procedure in addition to CABG) and preoperative

hemoglobin. Operative priority was not included because 88.7% of operations were urgent; in our setting this variable reflects the referral pathway rather than clinical acuity, and acuity is captured by the critical preoperative state.

Model 2 (24 hours after ICU admission) is intended to support ICU capacity planning and the targeting of early preventive care. It combined the linear predictor of Model 1 with five variables available by the end of the first postoperative day: CPB duration (0 minutes for off-pump surgery), intraoperative red-cell transfusion, cardiogenic shock (registry definition: shock present within 24 hours before surgery or on ICU admission; because it includes the state on ICU admission, it was assigned to Model 2), PaO_2/FiO_2 ratio on ICU admission, and chest-tube drainage during the first postoperative day.

Postoperative complications (renal failure, reoperation, stroke, myocardial infarction and infection) were not used as predictors. Their timing is not recorded in the registry, and they are intermediate events on the pathway to a prolonged stay rather than information available at the time of prediction; they are reported as outcomes. Vasoactive support, lactate and duration of mechanical ventilation were not recorded with a time stamp and could not be considered.

Sample size

The sample size was fixed by the registry. Using the criteria of Riley et al. [23], with 915 patients, an outcome prevalence of 14.3% and an anticipated area under the curve (AUC) of 0.72 (Cox–Snell R^2 0.077), the data supported about 8 predictor parameters for an expected shrinkage of at least 0.90. Model 1 used 12 parameters (ten predictors, two of them with a nonlinear term), for an expected shrinkage of 0.86. We therefore estimated shrinkage by bootstrapping and applied it to the final coefficients. The criterion for precise estimation of the overall risk (at least 189 patients) was met.

Missing data

Missing values (Additional file 1, Table S2) were imputed by multivariate imputation by chained equations with predictive mean matching (20 imputed datasets, 20 iterations) [24]. The imputation model included all predictors, the outcome, ICU length of stay, in-hospital death, year of surgery, EuroSCORE II and auxiliary variables (hematocrit, body mass index, New York Heart Association class, hypertension, insulin-treated diabetes, recent myocardial infarction, dialysis, cross-clamp time, operative duration and lowest hemoglobin in the first 24 hours). The outcome was not imputed. A complete-case analysis was performed as a sensitivity analysis.

Model development

Both models were fitted by maximum-likelihood logistic regression with all pre-specified predictors. Serum creatinine and chest-tube drainage were $\log_2$-transformed because of skewness. Continuous

predictors were modelled linearly unless a likelihood-ratio test comparing a restricted cubic spline with three knots (10th, 50th and 90th percentiles) against a linear term had $p < 0.05$; this applied to LVEF ($p = 0.040$) and creatinine ($p = 0.029$), whereas the other continuous predictors were modelled linearly ($p \geq 0.36$; Additional file 1, Table S3). Interaction terms were not considered, to limit the number of parameters. Model 2 was built as an extension of Model 1: The Model 1 linear predictor entered as a single term together with the five early postoperative variables [25]. This keeps the preoperative weights fixed and limits the number of new parameters. Coefficients were pooled across imputed datasets with Rubin's rules.

Internal validation and shrinkage

Internal validation used 200 bootstrap resamples in each imputed dataset (4,000 in total), repeating model estimation in every resample: fitting Model 1, computing its linear predictor and fitting Model 2 [26–28]. The functional forms and knot positions were fixed beforehand, and imputation was performed once, before bootstrapping [28]. Optimism was the mean difference between performance in the bootstrap sample and performance in the original data, and optimism-corrected performance was apparent performance minus optimism. The optimism-corrected calibration slope was used as a uniform shrinkage factor: all coefficients except the intercept were multiplied by it, and the intercept was re-estimated [25]. The shrunken models are the final models.

Performance measures

Discrimination was measured with the AUC; 95% confidence intervals (CIs) were obtained with the DeLong method and pooled across imputations on the logit scale. Calibration was assessed with calibration-in-the-large (CITL), the calibration slope, and flexible (loess) calibration curves with bootstrap 95% confidence bands, summarized by the integrated calibration index (ICI) and the 90th percentile of absolute error (E90) [29, 30]. The Hosmer–Lemeshow test was not used because it has low power and does not describe the direction or size of miscalibration [29]. Overall performance was measured with the Brier score and the scaled Brier score. Calibration curves, ICI, reclassification, decision curves and classification measures were computed on the apparent predictions of the final shrunken models, averaged across imputations, and may therefore be slightly optimistic.

Comparison with EuroSCORE II

EuroSCORE II was calculated with the published algorithm [20], correcting three coding issues documented in the registry (the order of operative-priority codes, pulmonary hypertension severity and neurological dysfunction). AUCs were compared with the DeLong test for paired curves, applied to predictions averaged across imputations [31]; because Model 2 extends Model 1, their comparison used a pooled Wald test of the five added terms. Because EuroSCORE II estimates the risk of death, its probabilities are not calibrated for ICU stay. For calibration, reclassification and decision-curve analyses we therefore used a logistically recalibrated EuroSCORE II (intercept and slope re-estimated in our data).

We computed the category-free net reclassification improvement (NRI) and the integrated discrimination improvement (IDI), with 2,000 bootstrap CIs [32].

Clinical utility

Decision-curve analysis compared the net benefit of each model with that of treating all or no patients as high risk across threshold probabilities of 2–45% [33, 34]. We also report sensitivity, specificity, predictive values and likelihood ratios at predicted-risk thresholds of 10%, 20% and 30%.

Point score

To provide a bedside tool, we converted the models into integer point scores. Continuous predictors were categorized at established clinical thresholds rather than at data-driven cut-offs, which overfit and discard information [35]: age (< 60, 60–69, 70–79, ≥ 80 years), LVEF (≥ 51%, 31–50%, ≤ 30%, based on the EuroSCORE II categories with the two lowest merged), serum creatinine (≤ 1.2, 1.21–2.0, > 2.0 mg/dL), anemia (hemoglobin < 14.3 g/dL in men and < 13.3 g/dL in women, the World Health Organization thresholds adjusted for an altitude of 2,500–2,999 m [36]), CPB duration > 120 minutes, PaO_2/FiO_2 <200 and chest-tube drainage of 500–999 mL and ≥ 1,000 mL. Regression coefficients were converted to points (1 point = 0.35 log-odds units) [37]. For the 24-hour score, the coefficients of the early variables, estimated together with the preoperative score, were divided by the coefficient of the preoperative score, so that the added points are on the same scale. The predicted probability for each total was obtained by logistic regression of the outcome on the total score. Risk strata were defined as predicted probabilities of about < 10% (low), 10–30% (intermediate) and ≥ 30% (high). The score derivation, including categorization and rounding, was internally validated with 400 bootstrap resamples.

Sensitivity analyses and temporal transportability

We refitted both models (50 bootstrap resamples per imputed dataset) in the following analyses: complete cases only; exclusion of patients who died in the ICU before day 7; a composite outcome of ICU stay ≥ 7 days or in-hospital death; alternative thresholds of ≥ 5 and ≥ 10 days; isolated CABG only; re-inclusion, in the complete-case dataset, of records with implausible ICU stay that had complete predictor data (9 of 10), classified as prolonged; and exclusion of 2020–2021, the years most affected by the COVID-19 pandemic [38]. We also fitted a single-stage model with all candidate predictors using the least absolute shrinkage and selection operator (LASSO), with the penalty chosen by 10-fold cross-validation and the tuning repeated within 100 bootstrap resamples in each of five imputed datasets [39]. To assess temporal transportability, we performed internal–external cross-validation by calendar period (2019–2020, 2021–2022 and 2023–2026): models were developed in two periods and validated in the third [26].

Influential observations

Influential observations were examined with Cook's distance, and the most influential records were reviewed for data errors. No observation was removed on the basis of influence.

Descriptive statistics and software

Continuous variables are reported as median [interquartile range, IQR] and categorical variables as number (%), with percentages calculated within each outcome group. Groups were compared with the Mann–Whitney U test and the chi-square or Fisher's exact test. Two-sided $p < 0.05$ was considered statistically significant. Analyses were performed in R version 4.3.3 with the packages mice, pROC, rms and glmnet. The analysis code is provided in Additional file 4.

Results

Participants

Of 1,487 procedures in the registry, 932 were CABG. After excluding 4 patients who died before ICU admission, 3 with missing ICU stay and 10 with implausible ICU stay, 915 patients formed the analytic cohort (Fig. 1). The median age was 67 years [IQR 60–73] and 25.7% were women. Surgery was urgent in 88.7% of patients and elective in 10.9%; 161 (17.6%) underwent a concomitant procedure and 109 (11.9%) off-pump surgery. The median EuroSCORE II was 3.37% [IQR 2.48–5.22%].

The median ICU stay was 3 days [IQR 2–5], and the 90th percentile was 8 days (Fig. 2). 131 patients (14.3%) stayed in the ICU for 7 days or longer; their median ICU stay was 9 days [IQR 8–13] and their median hospital stay 28 [20–38] days. In-hospital mortality was 5.7% (52 patients), and all deaths occurred in the ICU; 36 patients (3.9%) died before day 7 and were classified as not having a prolonged stay.

Patients with a prolonged stay were older and more often had reduced LVEF, chronic kidney disease or dialysis, peripheral arterial disease, atrial fibrillation, pulmonary hypertension, a critical preoperative state and a concomitant procedure; they also had longer CPB, more cardiogenic shock, lower PaO_2/FiO_2 and more chest-tube drainage (Table 1). Postoperative complications were several times more frequent in this group (Table 2), including reoperation (26.0% vs 5.2%), renal failure (7.6% vs 1.1%), stroke (4.6% vs 0.3%), reintubation (13.7% vs 2.6%) and sepsis (7.6% vs 1.4%). No predictor had more than 11% missing values, and 731 patients (79.9%) had complete data.

Table 1
Baseline, operative and early postoperative characteristics by ICU length of stay

Variable	All (n = 915)	ICU stay ≥ 7 days (n = 131)	ICU stay < 7 days (n = 784)	p value	Missing, n
Demographics					
Age, years	67 [60–73]	69 [62–75]	66 [60–73]	0.005	
Female sex	235 (25.7)	35 (26.7)	200 (25.5)	0.770	
Body mass index, kg/m²	25.7 [23.4–28.1]	25.3 [22.6–28.0]	25.7 [23.4–28.1]	0.375	8
Obesity (BMI ≥ 30 kg/m²)	130 (14.3)	19 (14.7)	111 (14.3)	0.890	8
Comorbidities and cardiac status					
Hypertension	662 (72.3)	101 (77.1)	561 (71.6)	0.189	
Insulin-treated diabetes	90 (9.8)	20 (15.3)	70 (8.9)	0.024	
Chronic obstructive pulmonary disease	37 (4.0)	8 (6.1)	29 (3.7)	0.195	
Chronic kidney disease, not on dialysis	76 (8.3)	18 (13.7)	58 (7.4)	0.015	
Chronic dialysis	29 (3.2)	11 (8.4)	18 (2.3)	0.001	
Serum creatinine, mg/dL	0.96 [0.81–1.12]	1.03 [0.81–1.25]	0.96 [0.82–1.10]	0.055	30
Peripheral arterial disease	36 (3.9)	10 (7.6)	26 (3.3)	0.019	
Previous stroke	41 (4.5)	9 (6.9)	32 (4.1)	0.153	
Atrial fibrillation or flutter	39 (4.3)	11 (8.4)	28 (3.6)	0.011	
History of heart failure	116 (12.7)	24 (18.3)	92 (11.7)	0.036	

Values are median [interquartile range] or n (%), with percentages calculated among patients with available data in each column. Anaemia: haemoglobin < 14.3 g/dL (men) or < 13.3 g/dL (women), adjusted for altitude. Critical preoperative state: EuroSCORE II definition or preoperative intra-aortic balloon pump. Cardiopulmonary bypass and cross-clamp times are shown for on-pump patients only. EuroSCORE II is expressed as predicted mortality (%). BMI: body mass index; CABG: coronary artery bypass grafting; ICU: intensive care unit; LVEF: left ventricular ejection fraction; NYHA: New York Heart Association.

Variable	All (n = 915)	ICU stay ≥ 7 days (n = 131)	ICU stay < 7 days (n = 784)	p value	Missing, n
Demographics					
NYHA class III–IV	69 (7.6)	15 (11.5)	54 (6.9)	0.064	2
Myocardial infarction within 90 days	442 (48.3)	60 (45.8)	382 (48.7)	0.535	
Previous cardiac surgery	9 (1.0)	1 (0.8)	8 (1.0)	1.000	
Pulmonary hypertension	47 (5.1)	13 (9.9)	34 (4.3)	0.007	
Haemoglobin, g/dL	14.2 [12.8–15.5]	13.8 [11.5–15.0]	14.4 [12.9–15.7]	0.002	37
Anaemia (altitude-adjusted WHO criteria)	392 (44.6)	73 (57.5)	319 (42.5)	0.002	37
LVEF, %	52 [41–58]	48 [36–57]	52 [43–58]	0.006	16
LVEF category				< 0.001	16
≤30%	74 (8.2)	22 (17.2)	52 (6.7)		
31–50%	354 (39.4)	51 (39.8)	303 (39.3)		
≥51%	471 (52.4)	55 (43.0)	416 (54.0)		
Presentation					
Operative priority				0.442	
Elective	100 (10.9)	15 (11.5)	85 (10.8)		
Urgent	812 (88.7)	115 (87.8)	697 (88.9)		
Emergency/salvage	3 (0.3)	1 (0.8)	2 (0.3)		
Critical preoperative state	40 (4.4)	12 (9.2)	28 (3.6)	0.004	

Values are median [interquartile range] or n (%), with percentages calculated among patients with available data in each column. Anaemia: haemoglobin < 14.3 g/dL (men) or < 13.3 g/dL (women), adjusted for altitude. Critical preoperative state: EuroSCORE II definition or preoperative intra-aortic balloon pump. Cardiopulmonary bypass and cross-clamp times are shown for on-pump patients only. EuroSCORE II is expressed as predicted mortality (%). BMI: body mass index; CABG: coronary artery bypass grafting; ICU: intensive care unit; LVEF: left ventricular ejection fraction; NYHA: New York Heart Association.

Variable	All (n = 915)	ICU stay ≥ 7 days (n = 131)	ICU stay < 7 days (n = 784)	p value	Missing, n
Demographics					
Preoperative intra-aortic balloon pump	26 (2.8)	9 (6.9)	17 (2.2)	0.007	
EuroSCORE II, %	3.37 [2.48–5.22]	4.89 [3.01–8.97]	3.22 [2.40–4.77]	< 0.001	
Operative data					
CABG with concomitant procedure	161 (17.6)	47 (35.9)	114 (14.5)	< 0.001	
Off-pump CABG	109 (11.9)	11 (8.4)	98 (12.5)	0.180	
Cardiopulmonary bypass time, min (on-pump)	91 [76–115]	104 [84–153]	90 [74–112]	< 0.001	109
Aortic cross-clamp time, min (on-pump)	75 [58–94]	86 [70–124]	73 [57–92]	< 0.001	112
Operative duration, min	275 [225–334]	302 [250–390]	270 [225–330]	< 0.001	73
Intraoperative red-cell transfusion	141 (15.4)	31 (23.7)	110 (14.0)	0.005	
Early postoperative data (first 24 h)					
Cardiogenic shock at ICU admission or within 24 h before surgery	42 (4.6)	18 (13.7)	24 (3.1)	< 0.001	
PaO_2/FiO_2 on ICU admission, mmHg	242 [203–289]	233 [185–270]	247 [208–292]	0.001	93
Chest-tube drainage, first postoperative day, mL	270 [145–500]	328 [192–604]	260 [140–482]	0.005	50
Lowest haemoglobin in first 24 h, g/dL	11.3 [10.0–	10.8 [9.6–12.0]	11.4 [10.1–12.8]	< 0.001	7

Values are median [interquartile range] or n (%), with percentages calculated among patients with available data in each column. Anaemia: haemoglobin < 14.3 g/dL (men) or < 13.3 g/dL (women), adjusted for altitude. Critical preoperative state: EuroSCORE II definition or preoperative intra-aortic balloon pump. Cardiopulmonary bypass and cross-clamp times are shown for on-pump patients only. EuroSCORE II is expressed as predicted mortality (%). BMI: body mass index; CABG: coronary artery bypass grafting; ICU: intensive care unit; LVEF: left ventricular ejection fraction; NYHA: New York Heart Association.

Variable	All (n = 915)	ICU stay ≥ 7 days (n = 131)	ICU stay < 7 days (n = 784)	p value	Missing, n
Demographics					
	12.7]				

Values are median [interquartile range] or n (%), with percentages calculated among patients with available data in each column. Anaemia: haemoglobin < 14.3 g/dL (men) or < 13.3 g/dL (women), adjusted for altitude. Critical preoperative state: EuroSCORE II definition or preoperative intra-aortic balloon pump. Cardiopulmonary bypass and cross-clamp times are shown for on-pump patients only. EuroSCORE II is expressed as predicted mortality (%). BMI: body mass index; CABG: coronary artery bypass grafting; ICU: intensive care unit; LVEF: left ventricular ejection fraction; NYHA: New York Heart Association.

Table 2
Postoperative course and in-hospital outcomes by ICU length of stay

Outcome	All (n = 915)	ICU stay ≥ 7 days (n = 131)	ICU stay < 7 days (n = 784)	p value
ICU length of stay, days	3 [2–5]	9 [8–13]	3 [2–4]	< 0.001
Hospital length of stay, days	17 [13–24]	28 [20–38]	16 [12–22]	< 0.001
In-hospital death	52 (5.7)	16 (12.2)	36 (4.6)	< 0.001
Death in the ICU before day 7	36 (3.9)	0 (0.0)	36 (4.6)	0.012
Reoperation during index admission	75 (8.2)	34 (26.0)	41 (5.2)	< 0.001
Reoperation for bleeding	31 (3.4)	11 (8.4)	20 (2.6)	0.002
Postoperative renal failure	19 (2.1)	10 (7.6)	9 (1.1)	< 0.001
Postoperative dialysis	10 (1.1)	5 (3.8)	5 (0.6)	0.008
Postoperative stroke	8 (0.9)	6 (4.6)	2 (0.3)	< 0.001
Pneumonia	12 (1.3)	7 (5.3)	5 (0.6)	< 0.001
New postoperative atrial fibrillation	151 (16.5)	45 (34.4)	106 (13.5)	< 0.001
Suspected perioperative myocardial infarction	21 (2.3)	6 (4.6)	15 (1.9)	0.104
Sepsis	21 (2.3)	10 (7.6)	11 (1.4)	< 0.001
Reintubation	38 (4.2)	18 (13.7)	20 (2.6)	< 0.001
Postoperative cardiac arrest	59 (6.4)	20 (15.3)	39 (5.0)	< 0.001
Postoperative transfusion	374 (40.9)	82 (62.6)	292 (37.2)	< 0.001
Readmission	98 (10.7)	13 (9.9)	85 (10.8)	0.753

Values are median [interquartile range] or n (%). These events were not used as predictors because their timing is not recorded in the registry. All in-hospital deaths occurred in the ICU. Hospital length of stay includes the preoperative period. Suspected perioperative myocardial infarction as recorded in the registry. ICU: intensive care unit.

Model 1 and Model 2

Table 3 shows the pooled odds ratios (ORs). In Model 1, the strongest predictor was a concomitant procedure (OR 3.52 (2.26–5.46)), followed by age (OR per 10 years 1.50 (1.18–1.90)), reduced LVEF (OR for 30% vs 55%, 2.62 (1.57–4.35)), higher serum creatinine (OR for 2.0 vs 1.0 mg/dL, 1.52 (1.16–2.00)) and extracardiac vascular disease (OR 1.86 (1.01–3.42)). Risk was lowest at an LVEF of about 50–55% and rose steeply below it, with a slight and imprecisely estimated rise at higher values; the association

with creatinine suggested a U-shape, with the lowest risk at about 1.0 mg/dL, although the increase at low values was imprecise (OR for 0.7 vs 1.0 mg/dL, 1.36 (0.89–2.08); Additional file 1, Fig. S1). COPD, atrial fibrillation, critical preoperative state, hemoglobin and sex had wide CIs that included 1 but were retained, as pre-specified. In Model 2, the Model 1 linear predictor remained strongly associated with the outcome (OR per unit 2.30 (1.79–2.96)), and cardiogenic shock (OR 2.38 (1.16–4.86)) and chest-tube drainage (OR per doubling 1.19 (1.02–1.39)) added information. CPB duration (OR per 30 minutes 1.10 (0.97–1.24)) and PaO_2/FiO_2 (OR per 50 mmHg decrease 1.11 (0.94–1.30)) were weakly associated, and intraoperative transfusion was not (OR 0.97 (0.58–1.63)). The final shrunken equations, with intercepts, coding and units, are given in Table 4, and a worked example in Additional file 1.

Table 3
Odds ratios for prolonged ICU stay (≥ 7 days) in Model 1 and Model 2

Predictor	Contrast	OR (95% CI)	p value
Model 1 (preoperative)			
Age	per 10 years	1.50 (1.18–1.90)	< 0.001
Female sex	yes vs no	0.86 (0.52–1.43)	0.565
LVEF	40% vs 55%	1.48 (1.16–1.90)	0.002
LVEF	30% vs 55%	2.62 (1.57–4.35)	< 0.001
Serum creatinine	0.7 vs 1.0 mg/dL	1.36 (0.89–2.08)	0.160
Serum creatinine	2.0 vs 1.0 mg/dL	1.52 (1.16–2.00)	0.003
COPD	yes vs no	1.43 (0.59–3.47)	0.428
Extracardiac vascular disease	yes vs no	1.86 (1.01–3.42)	0.047
Atrial fibrillation/flutter	yes vs no	1.56 (0.71–3.43)	0.272
Critical preoperative state	yes vs no	1.73 (0.76–3.95)	0.190
Concomitant procedure	yes vs no	3.52 (2.26–5.46)	< 0.001
Haemoglobin	per 1 g/dL decrease	1.04 (0.94–1.15)	0.458
Model 2 (24 hours after ICU admission)			
Model 1 linear predictor	per 1-unit increase	2.30 (1.79–2.96)	< 0.001
CPB duration	per 30 min	1.10 (0.97–1.24)	0.137
Intraoperative RBC transfusion	yes vs no	0.97 (0.58–1.63)	0.911
Cardiogenic shock	yes vs no	2.38 (1.16–4.86)	0.018
PaO_2/FiO_2 on ICU admission	per 50 mmHg decrease	1.11 (0.94–1.30)	0.216
Chest-tube drainage, POD 1	per doubling	1.19 (1.02–1.39)	0.024

Pooled estimates across 20 imputed datasets (Rubin's rules) from the models before shrinkage. LVEF and serum creatinine were modelled with restricted cubic splines, so their effects are shown for clinically relevant contrasts. In Model 2, the Model 1 linear predictor summarises all preoperative predictors. Extracardiac vascular disease: peripheral arterial disease or previous stroke. Concomitant procedure: any cardiac procedure in addition to CABG. CI: confidence interval; COPD: chronic obstructive pulmonary disease; CPB: cardiopulmonary bypass; LVEF: left ventricular ejection fraction; OR: odds ratio; POD: postoperative day; RBC: red blood cell.

Table 4
Final model equations (after shrinkage)

Term	Coding / unit	Model 1 coefficient	Model 2 coefficient
Intercept	—	-2.2085	-3.2312
Age	years	0.03477	0.02866
Female sex	1 = yes, 0 = no	-0.1285	-0.1059
LVEF	%	-0.05041	-0.04155
LVEF, spline term	s(LVEF); knots 33, 52, 63	0.03650	0.03008
Serum creatinine	$\log_2$(mg/dL)	-0.7554	-0.6227
Creatinine, spline term	s($\log_2$ creatinine); knots at 0.69, 0.96, 1.42 mg/dL	0.8789	0.7244
COPD	1/0	0.3095	0.2551
Extracardiac vascular disease	1/0	0.5348	0.4408
Atrial fibrillation or flutter	1/0	0.3827	0.3155
Critical preoperative state	1/0	0.4760	0.3924
Concomitant procedure	1/0	1.0872	0.8961
Preoperative haemoglobin	g/dL	-0.0328	-0.0271
CPB duration	per 30 min (off-pump = 0)	—	0.0796
Intraoperative RBC transfusion	1/0	—	-0.0253
Cardiogenic shock	1/0	—	0.7400

Predicted probability = 1 / (1 + exp(− LP)), where LP is the intercept plus the sum of each coefficient multiplied by the value of the corresponding term. Shrinkage factors were 0.864 (Model 1) and 0.855 (Model 2). Model 2 coefficients for preoperative terms equal the Model 1 coefficients before shrinkage multiplied by the shrunken Model 2 coefficient of the Model 1 linear predictor (0.7124). Spline term for a variable x with knots $k_1 < k_2 < k_3$: $s(x) = [(x - k_1)_+^3 - (x - k_2)_+^3 (k_3 - k_1)/(k_3 - k_2) + (x - k_3)_+^3 (k_2 - k_1)/(k_3 - k_2)] / (k_3 - k_1)^2$, where $(u)_+ = u$ if $u > 0$ and 0 otherwise; for creatinine the knots are on the $\log_2$ scale (− 0.535, − 0.059, 0.506). Worked example (Additional file 1): a 72-year-old man with LVEF 35%, creatinine 1.4 mg/dL, haemoglobin 12.5 g/dL, isolated CABG and no other risk factors has a Model 1 risk of 16.5%; with CPB of 130 min, no transfusion or shock, PaO_2/FiO_2 180 and first-day drainage of 800 mL, his Model 2 risk is 21.8%. A spreadsheet calculator is provided in Additional file 3. Abbreviations as in Table 3.

Term	Coding / unit	Model 1 coefficient	Model 2 coefficient
PaO_2/FiO_2 on ICU admission	per 50 mmHg	—	-0.0862
Chest-tube drainage, POD 1	$\log_2$(mL)	—	0.1492

Predicted probability = 1 / (1 + exp(− LP)), where LP is the intercept plus the sum of each coefficient multiplied by the value of the corresponding term. Shrinkage factors were 0.864 (Model 1) and 0.855 (Model 2). Model 2 coefficients for preoperative terms equal the Model 1 coefficients before shrinkage multiplied by the shrunken Model 2 coefficient of the Model 1 linear predictor (0.7124). Spline term for a variable x with knots $k_1 < k_2 < k_3$: $s(x) = [(x - k_1)_+^3 - (x - k_2)_+^3 (k_3 - k_1)/(k_3 - k_2) + (x - k_3)_+^3 (k_2 - k_1)/(k_3 - k_2)] / (k_3 - k_1)^2$, where $(u)_+ = u$ if $u > 0$ and 0 otherwise; for creatinine the knots are on the $\log_2$ scale (− 0.535, − 0.059, 0.506). Worked example (Additional file 1): a 72-year-old man with LVEF 35%, creatinine 1.4 mg/dL, haemoglobin 12.5 g/dL, isolated CABG and no other risk factors has a Model 1 risk of 16.5%; with CPB of 130 min, no transfusion or shock, PaO_2/FiO_2 180 and first-day drainage of 800 mL, his Model 2 risk is 21.8%. A spreadsheet calculator is provided in Additional file 3. Abbreviations as in Table 3.

Discrimination, calibration and comparison with EuroSCORE II

The apparent AUC was 0.73 (95% CI 0.68–0.78) for Model 1 and 0.76 (95% CI 0.70–0.80) for Model 2; after correction for optimism, it was 0.70 and 0.73, respectively (Table 5, Fig. 3). The early postoperative variables improved Model 2 over Model 1 (pooled Wald test for the five added terms, p = 0.005: AUC difference 0.025). EuroSCORE II had an AUC of 0.68 (95% CI 0.63–0.73), lower than that of Model 1 (difference 0.052, p = 0.015) and Model 2 (difference 0.077, p = 0.001). After correction for optimism, the differences from EuroSCORE II were smaller: 0.025 for Model 1 and 0.048 for Model 2. The optimism-corrected calibration slopes before shrinkage were 0.86 and 0.86, which were applied as shrinkage factors. After shrinkage, apparent calibration was good for predicted risks below about 25%, where most patients lay (ICI 0.015 and 0.028); Model 2 underestimated risk between about 25% and 45% (E90 0.084; Fig. 4). Uncalibrated EuroSCORE II predicted a mean risk of 4.7% against an observed prevalence of 14.3% (ICI 0.110); after recalibration its ICI was 0.020. The optimism-corrected scaled Brier scores were 6.7% for Model 1, 8.0% for Model 2 and 6.9% for recalibrated EuroSCORE II, indicating modest and similar overall accuracy for the three models.

Table 5
Performance of Model 1, Model 2 and EuroSCORE II for prolonged ICU stay

Measure	Model 1 (preoperative)	Model 2 (24 hours)	EuroSCORE II
Discrimination			
AUC, apparent (95% CI)	0.73 (0.68–0.78)	0.76 (0.70–0.80)	0.68 (0.63–0.73)
AUC, optimism-corrected	0.70	0.73	0.68
Difference in AUC vs EuroSCORE II, apparent (DeLong p)	0.052 (0.015)	0.077 (0.001)	—
Difference in AUC vs EuroSCORE II, optimism-corrected	0.025	0.048	—
Calibration			
Calibration slope, optimism-corrected (before shrinkage)	0.86	0.86	1.02
Calibration-in-the-large, optimism-corrected	0.002	0.004	0.002
ICI / E90, final model (apparent)	0.015 / 0.040	0.028 / 0.084	0.020 / 0.049
Overall performance			
Brier score, apparent / optimism-corrected	0.110 / 0.115	0.107 / 0.113	0.114 / 0.114
Scaled Brier score, optimism-corrected (%)	6.7	8.0	6.9
Reclassification			
NRI vs recalibrated EuroSCORE II (95% CI)	0.31 (0.13 to 0.50)	0.42 (0.24 to 0.61)	—
IDI vs recalibrated EuroSCORE II (95% CI)	0.020 (-0.001 to 0.042)	0.039 (0.016 to 0.062)	—

Optimism was estimated with 200 bootstrap resamples in each of 20 imputed datasets, repeating model estimation (Model 1, its linear predictor and Model 2) in every resample. Calibration slope and calibration-in-the-large for EuroSCORE II refer to the logistically recalibrated score; uncalibrated EuroSCORE II had a mean predicted risk of 4.7% against an observed prevalence of 14.3% (ICI 0.110). DeLong tests used predictions averaged across imputations. Model 2 vs Model 1: AUC difference 0.025; pooled Wald test of the five added terms p = 0.005. NRI, IDI, ICI and E90 are apparent estimates. NRI: category-free net reclassification improvement; IDI: integrated discrimination improvement; ICI: integrated calibration index; E90: 90th percentile of absolute calibration error.

Measure	Model 1 (preoperative)	Model 2 (24 hours)	EuroSCORE II
Discrimination			
NRI / IDI vs Model 1 (95% CI)	—	0.28 (0.08 to 0.46) / 0.019 (0.007 to 0.032)	—
Optimism was estimated with 200 bootstrap resamples in each of 20 imputed datasets, repeating model estimation (Model 1, its linear predictor and Model 2) in every resample. Calibration slope and calibration-in-the-large for EuroSCORE II refer to the logistically recalibrated score; uncalibrated EuroSCORE II had a mean predicted risk of 4.7% against an observed prevalence of 14.3% (ICI 0.110). DeLong tests used predictions averaged across imputations. Model 2 vs Model 1: AUC difference 0.025; pooled Wald test of the five added terms p = 0.005. NRI, IDI, ICI and E90 are apparent estimates. NRI: category-free net reclassification improvement; IDI: integrated discrimination improvement; ICI: integrated calibration index; E90: 90th percentile of absolute calibration error.			

Compared with recalibrated EuroSCORE II, Model 1 had an NRI of 0.31 (95% CI 0.13 to 0.50) and an IDI of 0.020 (-0.001 to 0.042). Compared with Model 1, Model 2 had an NRI of 0.28 (0.08 to 0.46) and an IDI of 0.019 (0.007 to 0.032). The reclassification gain came mainly from patients without a prolonged stay, whose predicted risk was correctly lowered; the components for patients with a prolonged stay had confidence intervals that included zero.

Clinical utility

Both models had greater net benefit than recalibrated EuroSCORE II and then treating all patients between threshold probabilities of about 10% and 30% (Fig. 5). At a 15% threshold, the net benefit was 0.047 for Model 1, 0.057 for Model 2 and 0.035 for recalibrated EuroSCORE II; compared with recalibrated EuroSCORE II, Model 2 was equivalent to a net gain of 2.2 true positives per 100 operated patients. At a 20% threshold, Model 2 classified 20.0% of patients as high risk, with a sensitivity of 54.2%, specificity of 85.7%, positive predictive value of 38.8% and negative predictive value of 91.8% (Additional file 1, Table S4).

Point score

The preoperative score assigns 0–20 points and the 24-hour score add 0–7 points (Table 6). The AUC was 0.71 (95% CI 0.67–0.76) for the preoperative score and 0.74 (0.69–0.79) for the 24-hour score, and 0.68 and 0.70 after correction for optimism. With the 24-hour score, 61.7% of patients were in the low-risk stratum (0–4 points; observed prolonged stay 7.1%), 28.3% in the intermediate stratum (5–8 points; 18.5%) and 9.9% in the high-risk stratum (≥ 9 points; 47.3%) (Table 7, Fig. 6; observed and predicted risk for each total in Additional file 1, Table S5). The cut-off that maximized the Youden index was ≥ 7 points (sensitivity 55.0%, specificity 83.0%, positive likelihood ratio 3.24, negative likelihood ratio 0.54).

Table 6
Points score for prolonged ICU stay

Item	Points
Preoperative score	
Age 60–69 / 70–79 / ≥80 years	1 / 2 / 3
LVEF 31–50% / ≤30%	1 / 3
Serum creatinine > 2.0 mg/dL	3
Chronic obstructive pulmonary disease	1
Extracardiac vascular disease (peripheral arterial disease or previous stroke)	2
Atrial fibrillation or flutter	1
Critical preoperative state	2
Concomitant cardiac procedure	4
Anaemia (haemoglobin < 14.3 g/dL in men, < 13.3 g/dL in women)	1
Maximum preoperative score	20
Points added at 24 hours after ICU admission	
Cardiopulmonary bypass > 120 min	1
Cardiogenic shock	3
PaO_2/FiO_2 <200 on ICU admission	1
Chest-tube drainage on POD 1: 500–999 mL / ≥1,000 mL	1 / 2
Maximum 24-hour score	27
Predicted probability of ICU stay ≥ 7 days	
Preoperative score	1 / (1 + exp(−(-3.158 + 0.331 × points)))
24-hour score	1 / (1 + exp(−(-3.279 + 0.285 × points)))

Female sex, creatinine 1.21–2.0 mg/dL and intraoperative transfusion received 0 points after rounding and were omitted. The 24-hour score is the preoperative score plus the points added at 24 hours. AUC: preoperative score 0.71 (optimism-corrected 0.68); 24-hour score 0.74 (optimism-corrected 0.70). LVEF: left ventricular ejection fraction; POD: postoperative day.

Table 7
Risk strata of the points scores

Score / stratum	Points	Patients, n (%)	Events, n	Observed ICU stay ≥ 7 days, %
Preoperative score				
Low	0–3	543 (59.3)	46	8.5
Intermediate	4–6	259 (28.3)	41	15.8
High	≥ 7	113 (12.3)	44	38.9
24-hour score				
Low	0–4	565 (61.7)	40	7.1
Intermediate	5–8	259 (28.3)	48	18.5
High	≥ 9	91 (9.9)	43	47.3

Table 7
(continued). Classification performance at selected cut-offs

Cut-off	Classified high risk, %	Sensitivity, %	Specificity, %	PPV, %	NPV, %	LR+	LR−
Preoperative score							
≥ 4 points	40.7	64.9	63.4	22.8	91.5	1.77	0.55
≥ 5 points (Youden)	29.6	58.0	75.1	28.0	91.5	2.33	0.56
≥ 7 points	12.3	33.6	91.2	38.9	89.2	3.82	0.73
24-hour score							
≥ 5 points	38.3	69.5	67.0	26.0	92.9	2.10	0.46
≥ 7 points (Youden)	22.4	55.0	83.0	35.1	91.7	3.24	0.54
≥ 9 points	9.9	32.8	93.9	47.3	89.3	5.36	0.72

LR+: positive likelihood ratio; LR−: negative likelihood ratio; NPV: negative predictive value; PPV: positive predictive value.

Sensitivity analyses, temporal transportability and influential observations

Results were consistent when patients who died before day 7 were excluded (optimism-corrected AUC 0.72 and 0.75 for Models 1 and 2), for the composite outcome of prolonged stay or death (0.73 and 0.77), and in complete cases (0.70 and 0.73); re-including records with implausible ICU stay in the complete-case data reduced them slightly (0.68 and 0.71) (Additional file 1, Table S6). Discrimination

was higher for a threshold of ≥ 10 days (0.72 and 0.77) and lower for ≥ 5 days (0.67 and 0.68), in isolated CABG (0.64 and 0.67) and after excluding 2020–2021 (0.67 and 0.67). The LASSO model retained almost all candidate predictors and had an optimism-corrected AUC of 0.72 (Table S7).

The prevalence of prolonged stay varied by year, from 7.4% in 2025 to 26.3% in 2021. In internal–external cross-validation, the AUC of Model 2 was 0.77 in 2019–2020, 0.74 in 2021–2022 and 0.65 in 2023–2026 (Model 1: 0.73, 0.71 and 0.66); in the most recent period Model 2 did not discriminate better than Model 1. Calibration-in-the-large varied substantially, with observed-to-expected ratios for Model 2 of 0.58, 1.60 and 1.28, and calibration slopes of 0.90, 0.79 and 0.52 (Table S8). The maximum Cook's distance was 0.024 (Model 2); the most influential records corresponded to plausible clinical courses, such as cardiogenic shock followed by a long ICU stay and death, and none contained data errors (Table S9).

Discussion

In this cohort of 915 patients undergoing CABG in a Latin American center where most operations are urgent, one in seven patients stayed in the ICU for 7 days or longer. A model based on ten preoperative variables predicted this outcome with moderate discrimination (optimism-corrected AUC 0.70) and adding five variables available 24 hours after ICU admission improved discrimination modestly (0.73). Model 2 discriminated better than EuroSCORE II and had greater net benefit; the advantage of Model 1 over EuroSCORE II was modest after correction for optimism, and overall accuracy was similar for the three. After shrinkage, apparent calibration was good at the low predicted risks where most patients lay. However, baseline risk varied considerably over time, so the models will need local recalibration before they are used.

Definitions of prolonged ICU stay vary widely [5, 12, 13]. We chose 7 days a priori because this threshold captures patients whose stay is driven by complications rather than by discharge logistics, and because it has been used in large derivation and validation cohorts [9, 14]. The frequency of complications in our patients with a prolonged stay (Table 2) supports this choice. Many studies in cardiac surgery have addressed ICU readmission rather than length of stay, but the risk factors are largely shared [40, 41]. In a cohort of 4,869 patients, Thomson et al. found that cerebrovascular disease and renal failure increased the risk of readmission and death [40], consistent with the roles of extracardiac vascular disease and creatinine in our model.

The discrimination of our models lies within the range reported for other models of prolonged ICU stay. In the systematic review and validation study by Ettema et al., the best general cardiac surgery models reached AUCs of 0.71–0.75 [8]. Rotar et al. reported an AUC of 0.745 for a stay of at least 72 hours [10], and a subsequent multi-institutional model reached a bootstrap-corrected C-statistic of 0.71 [22]. The CardiOttawa model, which used the same 7-day threshold, had a C-statistic of 0.85 at derivation and 0.78 at external validation [9]. These figures, together with the 0.70–0.73 range of our models, indicate that ICU stay is only partly determined by patient characteristics. Organisational factors such as the availability of ward beds, discharge practices and staffing influence it heavily [4, 13].

A concomitant procedure was the strongest preoperative predictor, in line with the inclusion of procedure type in the CardiOttawa model [9]. Age and LVEF are core components of preoperative risk scores such as EuroSCORE II [20, 42], and EuroSCORE has been associated with ICU stays longer than 48 hours [43]. The nonlinear relation with LVEF suggests that risk increases mainly when LVEF falls below about 50%, the threshold used by EuroSCORE II. Higher creatinine reflects chronic kidney disease, which is associated with worse outcomes and higher costs after cardiac surgery [16, 44]. The suggestion of increased risk at low creatinine values was imprecise; if confirmed, it might reflect low muscle mass or frailty. Preoperative anemia has been associated with transfusion and longer ICU stay [45]. In our model its independent contribution was small, perhaps because hemoglobin is higher at altitude and because other predictors capture part of its effect; for the score we used altitude-adjusted thresholds [36].

EuroSCORE II did not discriminate well for prolonged stay (AUC 0.68), similar to the AUC of 0.71 reported for the original EuroSCORE by Ettema et al. [8]. As a mortality score, it underestimated the probability of prolonged stay about threefold, and its calibration required re-estimation. In a previous study in our institution, EuroSCORE II also discriminated and calibrated poorly for in-hospital mortality after CABG [18], probably because of our case mix, which is dominated by urgent surgery in patients admitted with acute coronary syndromes. Our preoperative model does not replace EuroSCORE II for mortality, but at the same point in time it discriminates modestly better for anticipating ICU resource use.

The 24-hour model added CPB duration, intraoperative transfusion, shock, oxygenation and bleeding. Longer CPB increases the risk of vasoplegia and acute kidney injury [15, 16], and Rosenfeld et al. identified longer CPB time as a predictor of stays of 7 days or longer after isolated CABG [14]. In our model CPB duration added little once the preoperative linear predictor was known, probably because concomitant procedures, which account for most long CPB runs, were already captured. Reoperation for bleeding is associated with longer ICU stay, higher costs and worse long-term outcomes [17, 46, 47]. Increased chest-tube drainage is independently associated with adverse outcomes [48]; in our data, first-day drainage predicted prolonged stay and is known early, often before a decision to re-explore is made. Cardiogenic shock at ICU admission reflects low cardiac output, one of the main causes of a complicated course [3]. Early PaO_2/FiO_2 has predicted prolonged ICU stay in other cohorts [49]. Its weaker association in our data may partly reflect the lower arterial oxygen tension expected at 2,600 m, which suggests that thresholds derived at sea level should not be transposed directly. Intraoperative transfusion, a marker of morbidity in other studies [50], did not add information once bleeding and preoperative risk were known.

Postoperative complications such as renal failure, reoperation for bleeding, stroke and perioperative myocardial infarction were deliberately treated as outcomes rather than predictors. These complications were far more frequent in patients with a prolonged stay (Table 2), but they occur on the causal pathway to a prolonged stay, their timing is not recorded, and including them produces a model that explains the outcome rather than predicts it at a defined time [51]. Perioperative myocardial injury and stroke remain important determinants of prognosis and length of stay after cardiac surgery [52, 53]. Preventing these

complications, through meticulous haemostasias, renal protection and neurological monitoring, is the most direct way to shorten ICU stay.

A prediction model does not change the recovery trajectory by itself; its value depends on the decisions it informs. Before surgery, Model 1 can help to schedule high-risk operations, for example combined procedures in patients with reduced LVEF, when ICU capacity is available; to inform patients and families; and to prompt optimization of modifiable factors such as anemia and renal function. At 24 hours, Model 2 can identify patients who are likely to remain in the ICU for a week, in whom early mobilization, delirium prevention and nutritional assessment may be prioritized. Prediction of prolonged ICU stay has, for example, been proposed as a way to identify nutritional risk after cardiac surgery [54]. The high negative predictive value of the low-risk stratum (about 93% of patients in it did not have a prolonged stay) may be the most useful operational property, because it identifies patients suitable for early step-down. Whether using the models improves outcomes or efficiency must be tested prospectively.

The temporal analysis showed that the prevalence of prolonged stay more than doubled, from 11.3% in 2019–2020 to 23.2% in 2021–2022. This period partly coincided with the COVID-19 pandemic, which disrupted cardiac surgery practice and ICU capacity [38], although other changes in case mix or organization may have contributed. Discrimination was reasonably stable in the first two periods but lower in 2023–2026, when the early postoperative variables added nothing, and calibration-in-the-large varied substantially. This pattern is expected for an outcome that depends on organizational factors not captured by patient data [4, 13]. It implies that the model intercept should be updated locally before use and that performance should be monitored periodically.

Strengths and limitations

The strengths of this study include pre-specified predictors with explicit timing, multiple imputation, bootstrap validation of model estimation with shrinkage, a formal comparison with EuroSCORE II including decision-curve analysis, temporal validation, several sensitivity analyses, a point score, complete reporting of the model equations and a publicly available dataset.

This study also has limitations. First, it is a retrospective single-center study, and validation was internal and internal–external only. External validation in other centers is required before clinical use, ideally in samples with at least 100 events [55]. Second, ICU stay was recorded in whole days, and ten records with implausible values were excluded; re-including them in a complete-case analysis reduced discrimination slightly without changing the conclusions. Third, the registry definition of cardiogenic shock combines preoperative shock and shock on ICU admission, and variables that are likely to be important at 24 hours, such as vasoactive-inotropic support, lactate, duration of mechanical ventilation and frailty, were not available with reliable timing. Fourth, death competes with prolonged stay; we addressed this with two alternative outcome definitions. Fifth, the number of events was below that required by Riley's criteria for the number of parameters used, which we handled with bootstrap-based shrinkage, although shrinkage cannot fully compensate for a limited sample size [56]. Sixth, our case

mix, with nearly nine in ten operations urgent, limits generalizability to elective-dominant settings. Performance was also lower in isolated CABG, in which the strongest predictor, a concomitant procedure, is absent. Finally, we could not compare our models with the Society of Thoracic Surgeons length-of-stay models, which were not calculated in the registry.

Conclusions

In patients undergoing CABG in a Colombian High-altitude university hospital, preoperative data predicted ICU stay of 7 days or longer with moderate discrimination, and data available 24 hours after ICU admission improved prediction modestly. The 24-hour model discriminated better than EuroSCORE II, and the preoperative model modestly so. A simple point score identifies low-, intermediate- and high-risk groups. Because baseline risk varied over time and validation was internal, the models should be externally validated and recalibrated locally before they are used to support ICU planning.

Abbreviations

AUC
area under the receiver operating characteristic (ROC) curve
CABG
coronary artery bypass grafting
CI
confidence interval
CITL
calibration-in-the-large
COPD
chronic obstructive pulmonary disease
CPB
cardiopulmonary bypass
E90
90th percentile of absolute calibration error
ICI
integrated calibration index
ICU
intensive care unit
IDI
integrated discrimination improvement
IQR
interquartile range
LASSO
least absolute shrinkage and selection operator
LVEF

left ventricular ejection fraction
NRI
net reclassification improvement
OR
odds ratio
POD
postoperative day
RBC
red blood cell
TRIPOD
Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis.

Declarations

Ethics approval and consent to participate

The study was approved by the Research Ethics Committee of Hospital Universitario Mayor–Méderi (approval DVO005 2333-CV1731), which waived the requirement for informed consent because of the retrospective design and the use of de-identified registry data. The study was conducted in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This study received no external funding.

Author Contribution

D.R.R.L. conceived the study, contributed to the study design, and participated in data analysis and interpretation of the results. E.E.R.A. drafted the original manuscript, participated in data analysis and interpretation, and contributed to the critical review, editing, and revision of subsequent versions of the manuscript. D.M.A.A., N.Y.L.G., M.J.N.Q., and B.Y.G.R. contributed to data collection, data curation, and

data analysis. All authors critically reviewed the manuscript, read and approved the final version, and agreed to be accountable for the work.

Acknowledgement

N/A

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available due to institutional and ethical considerations regarding the confidentiality of patient-level data but are available from the corresponding author upon reasonable request.

References

1. World Health Organization. Cardiovascular diseases (CVDs) [Internet], Geneva WHO. 2025 [cited 2026 Sep 10]. Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).
2. Lawton JS, Tamis-Holland JE, Bangalore S, Bates ER, Beckie TM, Bischoff JM, et al. 2021 ACC/AHA/SCAI guideline for coronary artery revascularization: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation. 2022;145(3):e18–114.
3. Stephens RS, Whitman GJ. Postoperative critical care of the adult cardiac surgical patient. Part I: routine postoperative care. Crit Care Med. 2015;43(7):1477–97.
4. Widyastuti Y, Stenseth R, Wahba A, Pleym H, Videm V. Length of intensive care unit stay following cardiac surgery: is it impossible to find a universal prediction model? Interact Cardiovasc Thorac Surg. 2012;15(5):825–32.
5. Almashrafi A, Elmontsri M, Aylin P. Systematic review of factors influencing length of stay in ICU after adult cardiac surgery. BMC Health Serv Res. 2016;16:318.
6. Jóskowiak D, Kappert U, Matschke K, Tugtekin SM. Prolonged intensive care unit stay of patients after cardiac surgery: initial clinical results and follow-up. Thorac Cardiovasc Surg. 2013;61(8):701–7.
7. Hein OV, Birnbaum J, Wernecke K, England M, Konertz W, Spies C. Prolonged intensive care unit stay in cardiac surgery: risk factors and long-term-survival. Ann Thorac Surg. 2006;81(3):880–5.
8. Ettema RG, Peelen LM, Schuurmans MJ, Nierich AP, Kalkman CJ, Moons KG. Prediction models for prolonged intensive care unit stay after cardiac surgery: systematic review and validation study. Circulation. 2010;122(7):682–9.

9. Sun LY, Bader Eddeen A, Ruel M, MacPhee E, Mesana TG. Derivation and validation of a clinical model to predict intensive care unit length of stay after cardiac surgery. J Am Heart Assoc. 2020;9(21):e017847.
10. Rotar EP, Beller JP, Smolkin ME, Chancellor WZ, Ailawadi G, Yarboro LT, et al. Prediction of prolonged intensive care unit length of stay following cardiac surgery. Semin Thorac Cardiovasc Surg. 2022;34(1):172–9.
11. De Cocker J, Messaoudi N, Stockman BA, Bossaert LL, Rodrigus IE. Preoperative prediction of intensive care unit stay following cardiac surgery. Eur J Cardiothorac Surg. 2011;39(1):60–7.
12. Messaoudi N, De Cocker J, Stockman B, Bossaert LL, Rodrigus IE. Prediction of prolonged length of stay in the intensive care unit after cardiac surgery: the need for a multi-institutional risk scoring system. J Card Surg. 2009;24(2):127–33.
13. Bettex D, Rudiger A. Length of ICU stay after cardiac surgery: too long or too short? J Cardiothorac Vasc Anesth. 2018;32(6):2692–3.
14. Rosenfeld R, Smith JM, Woods SE, Engel AM. Predictors and outcomes of extended intensive care unit length of stay in patients undergoing coronary artery bypass graft surgery. J Card Surg. 2006;21(2):146–50.
15. Busse LW, Barker N, Petersen C. Vasoplegic syndrome following cardiothoracic surgery—review of pathophysiology and update of treatment options. Crit Care. 2020;24(1):36.
16. Wang Y, Bellomo R. Cardiac surgery-associated acute kidney injury: risk factors, pathophysiology and treatment. Nat Rev Nephrol. 2017;13(11):697–711.
17. Ruel M, Chan V, Boodhwani M, McDonald B, Ni X, Gill G, et al. How detrimental is reexploration for bleeding after cardiac surgery? J Thorac Cardiovasc Surg. 2017;154(3):927–35.
18. Rodríguez Lima DR, Rodríguez Aparicio EE, Otálora González L, Hernández D, González-Muñoz A. Performance of the EuroSCORE in coronary artery bypass graft in Colombia, a middle-income country: a retrospective cohort. World J Cardiol. 2025;17(3):100506.
19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inf. 2009;42(2):377–81.
20. Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, et al. EuroSCORE II. Eur J Cardiothorac Surg. 2012;41(4):734–44.
21. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378.
22. Wisniewski AM, Wang XQ, Sutherland G, Rotar EP, Strobel RJ, Young A, et al. Multi-institutional model to predict intensive care unit length of stay after cardiac surgery. J Thorac Cardiovasc Surg. 2025;170(3):784–92.
23. Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ. 2020;368:m441.

24. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Softw. 2011;45(3):1–67.
25. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. 2nd ed. Cham: Springer; 2019.
26. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016;69:245–7.
27. Harrell FE Jr. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. 2nd ed. Cham: Springer; 2015.
28. Wahl S, Boulesteix AL, Zierer A, Thorand B, van de Wiel MA. Assessment of predictive performance in incomplete data by combining internal validation and multiple imputation. BMC Med Res Methodol. 2016;16(1):144.
29. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230.
30. Austin PC, Steyerberg EW. The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. Stat Med. 2019;38(21):4051–65.
31. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837–45.
32. Pencina MJ, D'Agostino RB, Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med. 2011;30(1):11–21.
33. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak. 2006;26(6):565–74.
34. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res. 2019;3:18.
35. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med. 2006;25(1):127–41.
36. World Health Organization. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. Vitamin and Mineral Nutrition Information System. WHO/NMH/: Geneva: WHO; 2011. NHD/MNM/11.1.
37. Sullivan LM, Massaro JM, D'Agostino RB, Sr. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. Stat Med. 2004;23(10):1631–60.
38. Nguyen TC, Thourani VH, Nissen AP, Habib RH, Dearani JA, Ropski A, et al. The effect of COVID-19 on adult cardiac surgery in the United States in 717 103 patients. Ann Thorac Surg. 2022;113(3):738–46.
39. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1–22.
40. Thomson R, Fletcher N, Valencia O, Sharma V. Readmission to the intensive care unit following cardiac surgery: a derived and validated risk prediction model in 4,869 patients. J Cardiothorac Vasc

Anesth. 2018;32(6):2685–91.

41. Kolat P, Guttenberger P, Ried M, Kapahnke J, Haneya A, Schmid C, et al. ICU readmission after cardiac surgery—still a matter of concern? Thorac Cardiovasc Surg. 2020;68(5):384–8.
42. Ad N, Holmes SD, Patel J, Pritchard G, Shuman DJ, Halpin L. Comparison of EuroSCORE II, original EuroSCORE, and The Society of Thoracic Surgeons risk score in cardiac surgery patients. Ann Thorac Surg. 2016;102(2):573–9.
43. Meadows K, Gibbens R, Gerrard C, Vuylsteke A. Prediction of patient length of stay on the intensive care unit following cardiac surgery: a logistic regression analysis based on the cardiac operative mortality risk calculator, EuroSCORE. J Cardiothorac Vasc Anesth. 2018;32(6):2676–82.
44. Lysak N, Bihorac A, Hobson C. Mortality and cost of acute and chronic kidney disease after cardiac surgery. Curr Opin Anaesthesiol. 2017;30(1):113–7.
45. Hung M, Besser M, Sharples LD, Nair SK, Klein AA. The prevalence and association with transfusion, intensive care unit stay and mortality of pre-operative anaemia in a cohort of cardiac surgery patients. Anaesthesia. 2011;66(9):812–8.
46. Brown JA, Kilic A, Aranda-Michel E, Navid F, Serna-Gallegos D, Bianco V, et al. Long-term outcomes of reoperation for bleeding after cardiac surgery. Semin Thorac Cardiovasc Surg. 2021;33(3):764–73.
47. Christensen MC, Krapf S, Kempel A, von Heymann C. Costs of excessive postoperative hemorrhage in cardiac surgery. J Thorac Cardiovasc Surg. 2009;138(3):687–93.
48. Christensen MC, Dziewior F, Kempel A, von Heymann C. Increased chest tube drainage is independently associated with adverse outcome after cardiac surgery. J Cardiothorac Vasc Anesth. 2012;26(1):46–51.
49. Kapadohos T, Angelopoulos E, Vasileiadis I, Nanas S, Kotanidou A, Karabinis A, et al. Determinants of prolonged intensive care unit stay in patients after cardiac surgery: a prospective observational study. J Thorac Dis. 2017;9(1):70–9.
50. Koch CG, Li L, Duncan AI, Mihaljevic T, Cosgrove DM, Loop FD, et al. Morbidity and mortality risk associated with red blood cell and blood-component transfusion in isolated coronary artery bypass grafting. Crit Care Med. 2006;34(6):1608–16.
51. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51–8.
52. Devereaux PJ, Lamy A, Chan MTV, Allard RV, Lomivorotov VV, Landoni G, et al. High-sensitivity troponin I after cardiac surgery and 30-day mortality. N Engl J Med. 2022;386(9):827–36.
53. McKhann GM, Grega MA, Borowicz LM Jr, Baumgartner WA, Selnes OA. Stroke and encephalopathy after cardiac surgery: an update. Stroke. 2006;37(2):562–71.
54. Stoppe C, Ney J, Lomivorotov VV, Efremov SM, Benstoem C, Hill A, et al. Prediction of prolonged ICU stay in cardiac surgery patients as a useful method to identify nutrition risk in cardiac surgery

patients: a post hoc analysis of a prospective observational study. JPEN J Parenter Enter Nutr. 2019;43(6):768–79.

55. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Stat Med. 2016;35(2):214–26.
56. Riley RD, Snell KIE, Martin GP, Whittle R, Archer L, Sperrin M, et al. Penalization and shrinkage methods produced unreliable clinical prediction models especially when sample size was small. J Clin Epidemiol. 2021;132:88–96.

Figures

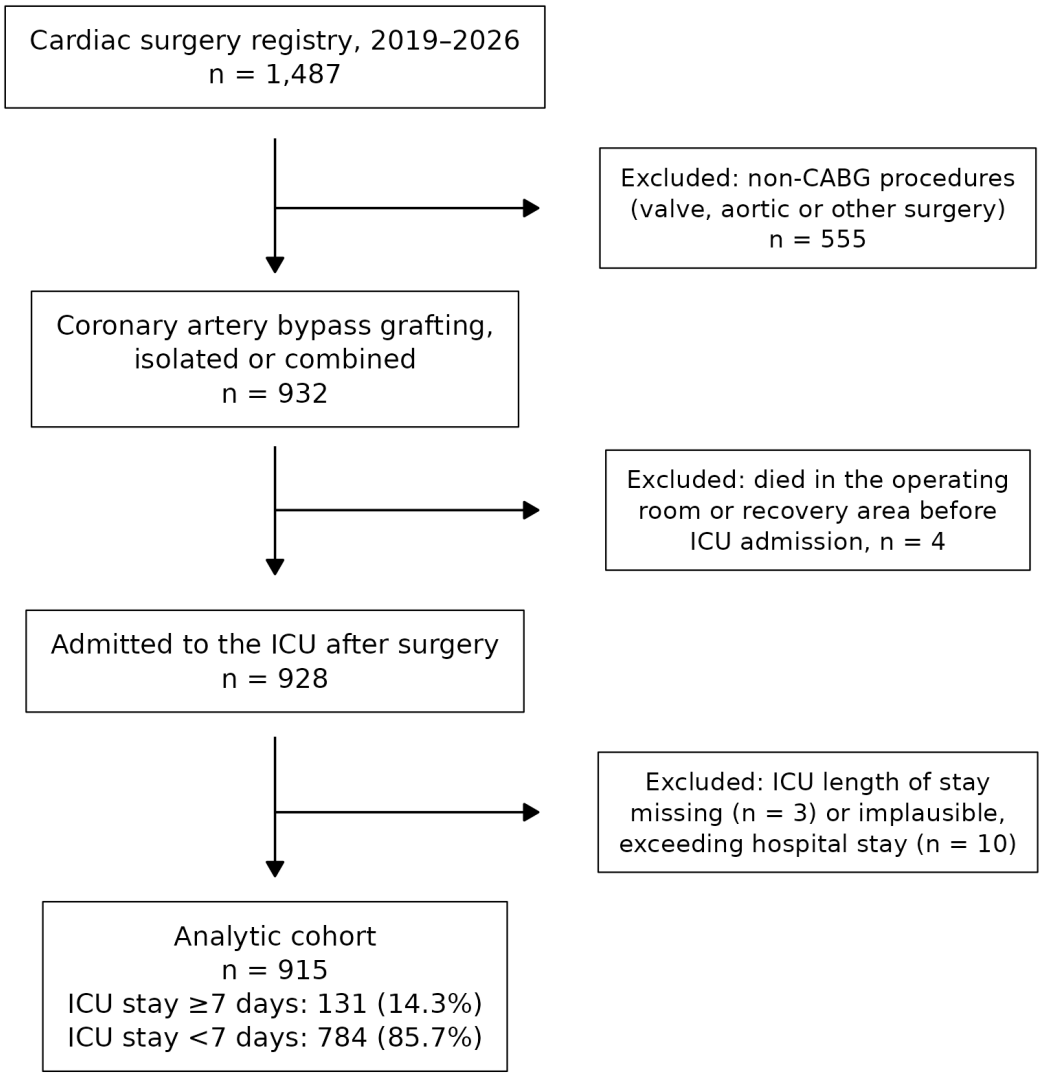


Figure 1

Flow of participants. Registry procedures between January 2019 and February 2026. ICU length of stay was considered implausible when it exceeded total hospital length of stay.

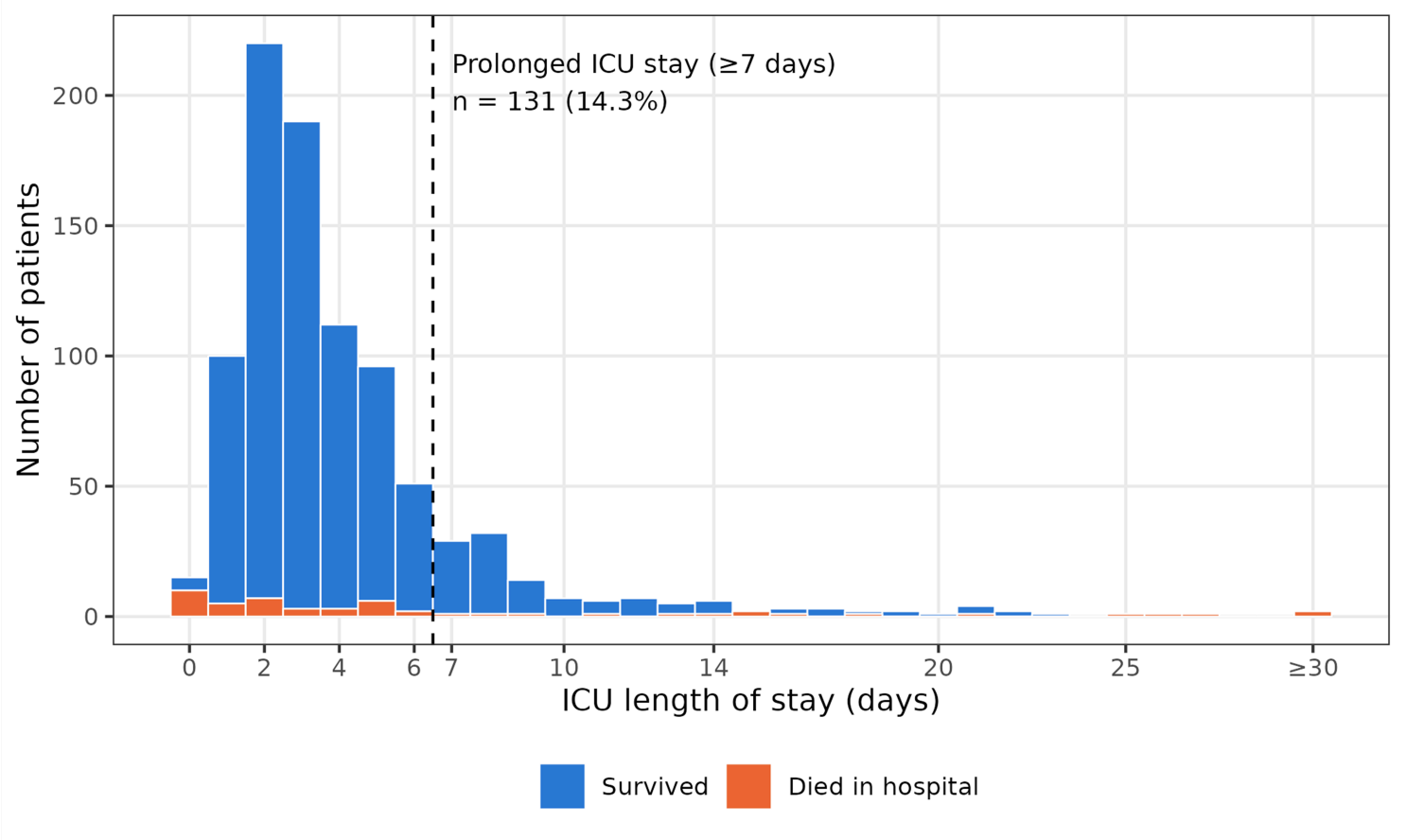


Figure 2

Distribution of ICU length of stay. Stays of 30 days or more are grouped in the last bar. The dashed line marks the threshold for prolonged ICU stay (≥7 days). Colours indicate in-hospital vital status; all in-hospital deaths occurred in the ICU.

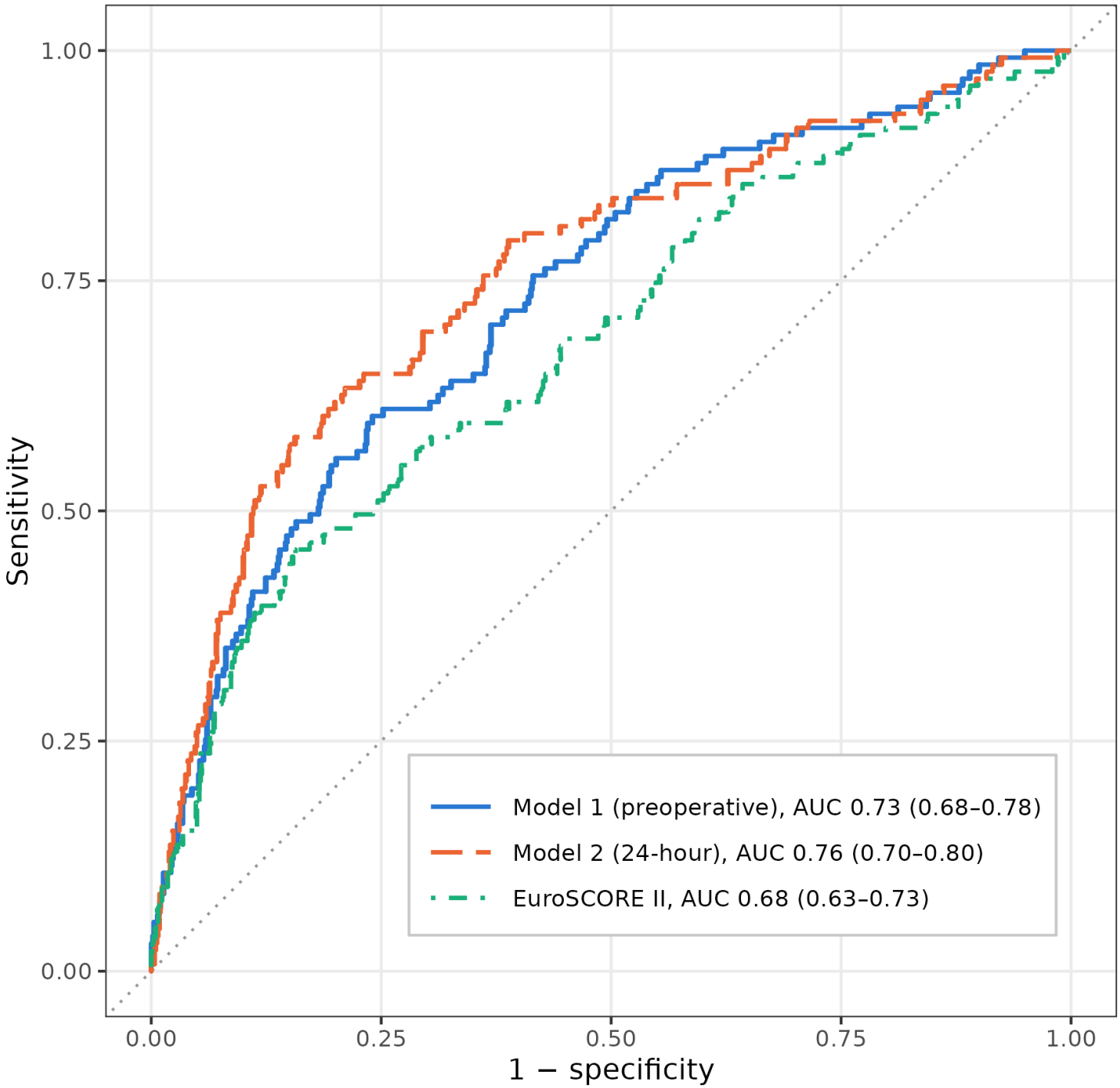


Figure 3

Receiver operating characteristic curves for Model 1, Model 2 and EuroSCORE II. Apparent AUCs with 95% confidence intervals (DeLong method, pooled across 20 imputed datasets). Optimism-corrected AUCs were 0.70 (Model 1) and 0.73 (Model 2).

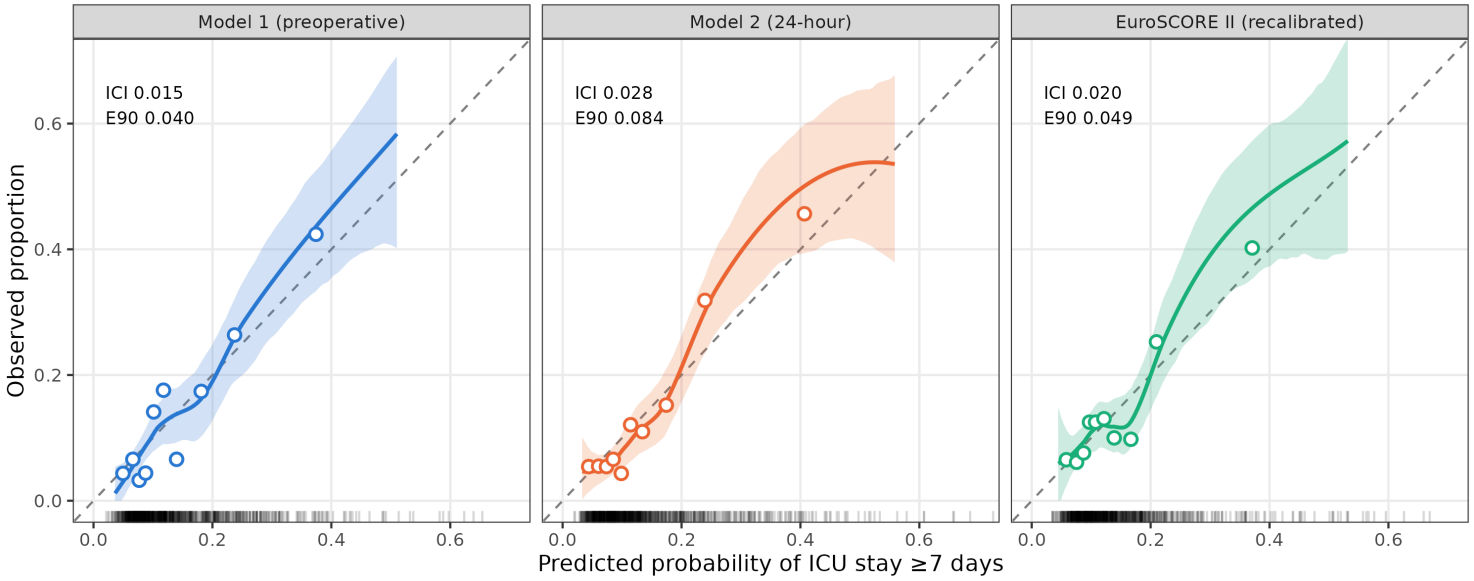


Figure 4

Calibration of Model 1, Model 2 and recalibrated EuroSCORE II. Lines are loess-smoothed calibration curves with bootstrap 95% confidence bands; circles are deciles of predicted risk; the rug shows the distribution of predictions; the dashed line indicates perfect calibration. Models 1 and 2 are shown after bootstrap-based shrinkage. EuroSCORE II was logistically recalibrated, because as a mortality score its uncalibrated predictions underestimate the probability of prolonged ICU stay. ICI: integrated calibration index; E90: 90th percentile of absolute error.

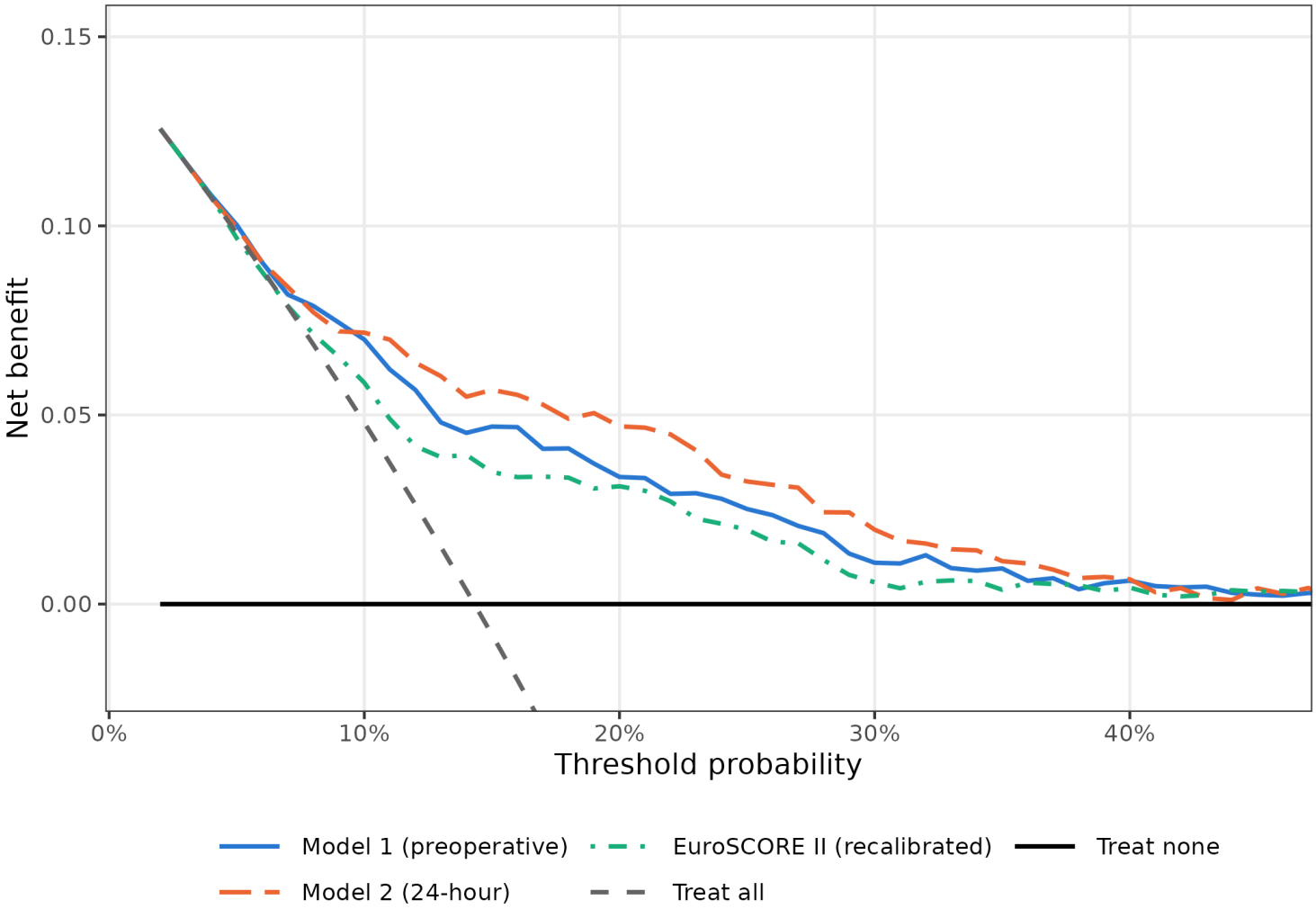


Figure 5

Decision-curve analysis. Net benefit of Model 1, Model 2 and recalibrated EuroSCORE II compared with classifying all or no patients as high risk, across threshold probabilities.

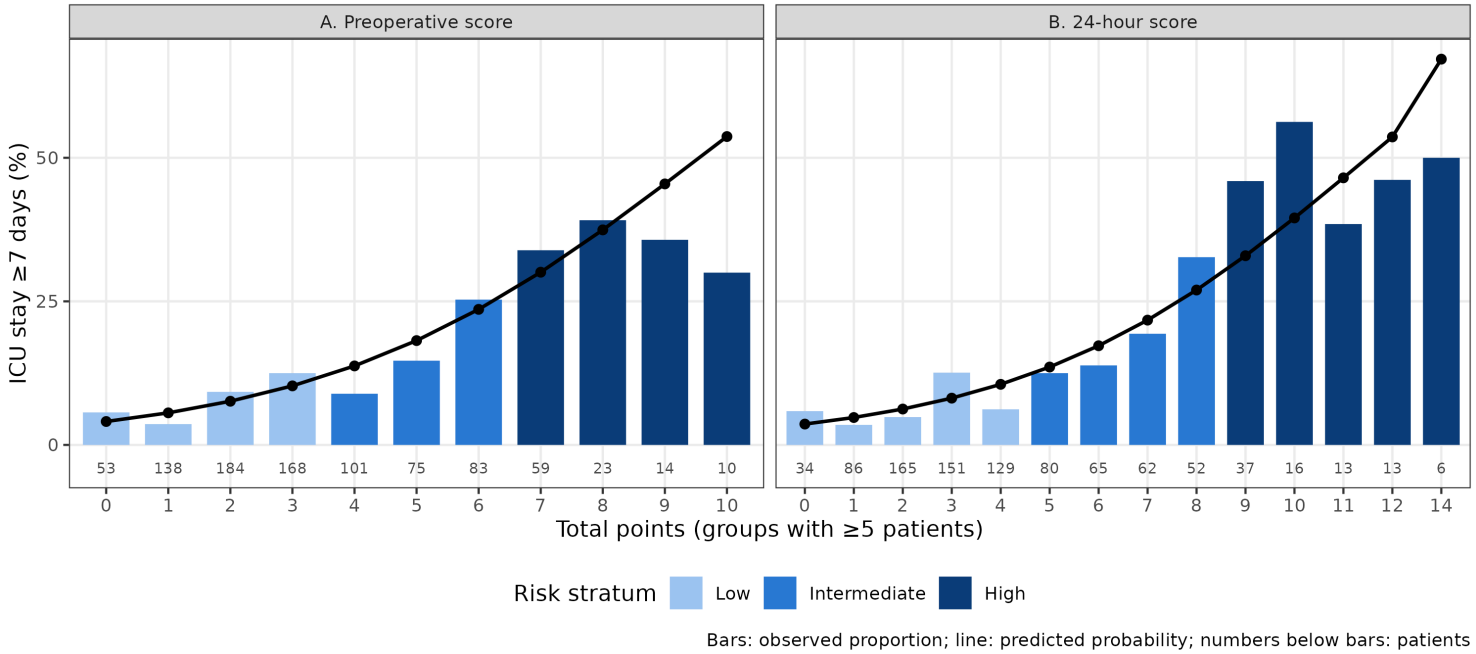


Figure 6

Observed and predicted probability of prolonged ICU stay by points score. A, preoperative score; B, 24-hour score. Bars show the observed proportion with ICU stay ≥7 days, colour-coded by risk stratum; the line shows the predicted probability; numbers below the bars are the numbers of patients. Only totals with at least five patients are shown.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1Supplementary.docx
- Additionalfile2TRIPODAIchecklist.docx
- Additionalfile3Riskcalculator.xlsx