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.

Two-Stage Perioperative Risk Prediction for Low Cardiac Output Syndrome in Older Adults Undergoing Cardiac Surgery: Model Development and Temporal Validation

Shengjie Ning[1]
Nanjing Medical University
Junzhe Wang[1]
Nanjing Medical University
Ruoyu Zhang[1]
Nanjing Medical University
Haoran Gan[2]
Nanjing University of Chinese Medicinen
Li Yin[1]
Nanjing Medical University
Yifei Zhou
Southeast University
Dongming Liu[2]
Nanjing University of Chinese Medicinen
Ji Wang[1]
Nanjing Medical University
Zhibing Qiu[1]
1771916816@qq.com

Nanjing Medical University

Research Article

Keywords: low cardiac output syndrome, cardiac surgery, older adults, two-stage prediction, temporal validation, calibration, decision-curve analysis

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11413566/v1

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

Additional Declarations: No competing interests reported.

Two-Stage Perioperative Risk Prediction for Low Cardiac Output Syndrome in Older Adults Undergoing Cardiac Surgery: Model Development and Temporal Validation

Shengjie Ning[1], Junzhe Wang[1], Ruoyu Zhang[1], Haoran Gan[2], Li Yin[1],Yifei Zhou[3]

Dongming Liu[2], Ji Wang[1], Zhibing Qiu[1]*

[1] Nanjing Medical University, Nanjing, Jiangsu, China; [2] Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China ;[3]Southeast University,Jiangsu, China

Corresponding author: Zhibing Qiu, e-mail：qiuzhibing2009@163.com Nanjing First Hospital (Nanjing Hospital Affiliated to Nanjing Medical University), 68 Changle Road, Qinhuai District, Nanjing 210006, Jiangsu, China

Abstract

Background: Low cardiac output syndrome (LCOS) is a major complication after cardiac surgery. Most risk assessment approaches rely primarily on preoperative information, whereas key factors such as cardiopulmonary bypass exposure, bleeding, and transfusion become available only during surgery. To align risk assessment with the progressive accumulation of perioperative information, we developed and temporally validated a two-stage strategy that generates an initial risk estimate from preoperative characteristics and then updates that estimate using routinely available intraoperative information.

Methods: This single-center retrospective cohort study included patients aged ≥65 years who underwent eligible cardiac surgery in 2024 as the development cohort. Patients treated from January through June 2025 formed the temporal validation cohort and were not used for model fitting. Model A was based on prespecified preoperative clinical variables, and Model B additionally included cardiopulmonary bypass duration and intraoperative allogeneic transfusion. Missing predictors were handled using 20 multiple imputations, and internal validation used 1,000 bootstrap resamples. Temporal validation assessed the area under the receiver operating characteristic curve (AUC), calibration-in-the-large (CITL), calibration slope, Brier score, and decision-curve net benefit. Paired between-model differences were estimated using 2,000 bootstrap resamples.

Results: The development cohort included 1,171 patients, of whom 101 developed LCOS (8.6%), and the temporal validation cohort included 606 patients, of whom 50 developed LCOS (8.3%). After 1,000 bootstrap resamples, optimism-corrected development AUCs were 0.699 (95% CI, 0.649–0.753) for Model A and 0.749 (95% CI, 0.705–0.799) for Model B. In the reserved 2025 temporal validation cohort, the AUC of Model B was 0.756 (95% CI, 0.695–0.810), compared with 0.655 (95% CI, 0.571–0.729) for Model A, yielding a paired difference of 0.101 (95% CI, 0.045–0.161; P<0.001). Model B also had a calibration slope closer to the ideal value (0.858 vs 0.684), with a CITL of −0.113. Model B showed clear incremental net benefit at risk thresholds of 5% and 10%, and its AUC remained 0.736–0.759 across sensitivity analyses. Brier scores were similar between models (0.0725 vs 0.0739).

Conclusions: Adding cardiopulmonary bypass duration and intraoperative allogeneic transfusion to the prespecified preoperative model markedly improved LCOS discrimination in temporal validation, produced a calibration slope closer to the ideal value, and yielded incremental net benefit at lower risk thresholds. This

two-stage strategy uses routinely available perioperative information to provide interpretable risk updating and may support early postoperative risk stratification, intensified monitoring, and resource allocation.

Keywords: low cardiac output syndrome; cardiac surgery; older adults; two-stage prediction; temporal validation; calibration; decision-curve analysis

Introduction

Low cardiac output syndrome (LCOS) is an important state of systemic hypoperfusion after cardiac surgery and is associated with organ dysfunction, prolonged intensive care, increased healthcare resource use, and mortality [1–3]. Identifying patients at high risk before LCOS develops is therefore important for perioperative surveillance, ICU preparedness, and optimization of early intervention strategies. Because published studies use heterogeneous hemodynamic, pharmacologic, and mechanical-support criteria [2], prediction models should not only perform well during development but also undergo rigorous validation in subsequent time periods to assess their stability in future patients.

Previous studies have identified several preoperative characteristics associated with postoperative low-output states, including advanced age, renal dysfunction, reduced left ventricular ejection fraction (LVEF), non-elective surgery, and hemodynamic instability [4], and risk scores and machine-learning models have subsequently been developed [6,7]. These studies support the feasibility of preoperative risk stratification; however, preoperative information cannot capture important process-related factors that emerge during surgery, including cardiopulmonary bypass exposure, bleeding, transfusion, and procedural complexity. Prior work suggests that adding intraoperative information to baseline risk can further improve prediction of postoperative cardiac dysfunction and other cardiac-surgical outcomes [5,8].

Accordingly, a prediction strategy that more closely reflects clinical workflow should update risk at key time points: an initial estimate before surgery, followed by a second-stage assessment at the end of surgery that incorporates newly available intraoperative information. This design preserves the prospective value of preoperative assessment while quantifying the incremental contribution of intraoperative data beyond baseline risk. Evaluation of such models should therefore extend beyond the AUC to include calibration, overall prediction error, temporal stability, and potential clinical net benefit [9–11].

We therefore developed and validated a two-stage perioperative LCOS prediction framework. A preoperative Model A was first developed and internally validated. Model B was then formed by adding cardiopulmonary bypass duration and intraoperative allogeneic transfusion to Model A. Both models were finally evaluated in a 2025 temporal validation cohort that was not used for model fitting. We focused on whether intraoperative information could provide stable incremental predictive value beyond preoperative risk, while evaluating discrimination, calibration, overall prediction error, and clinical net benefit.

Methods

Study design and setting

We conducted a single-center retrospective cohort study at Nanjing First Hospital (Nanjing Hospital Affiliated to Nanjing Medical University), Nanjing, China. Consecutive patients who underwent eligible surgery between January 1 and December 31, 2024 constituted the development cohort, whereas patients treated between January 1 and June 30, 2025 formed a non-overlapping temporal validation cohort. The 2025 data were not used for predictor selection or model fitting. The study was reported with reference to TRIPOD+AI [9], and potential risk-of-bias domains were considered using the PROBAST+AI framework [10].

Participants

Patients aged ≥65 years were eligible if they underwent coronary artery bypass grafting (CABG), surgical valve surgery, open aortic surgery, or a combined open cardiac procedure. For patients who underwent more than one eligible procedure during the study period, only the first procedure was retained. Transcatheter- or endovascular-only procedures, ventricular assist device implantation, and heart transplantation were excluded. The primary analysis also excluded patients with preoperative shock or missing shock status because shock may dominate both treatment decisions and LCOS risk; these patients were restored in a sensitivity analysis.

Outcome definition

The primary outcome was LCOS occurring within 48 hours after surgery after adequate preload optimization. LCOS was defined by at least one of the following: cardiac index <2.0 L/min/m²; a persistent low-output state requiring continuous inotropic support for >24 hours; or initiation of mechanical circulatory support for low cardiac output, including intra-aortic balloon pump, venoarterial extracorporeal membrane oxygenation, or another ventricular assist modality. Prophylactic mechanical circulatory support and vasoactive therapy used solely for isolated vasoplegia were not classified as LCOS. The same operational definition was applied in both calendar periods.

Predictors and two-stage modeling strategy

Predictors were prespecified before analysis on the basis of clinical knowledge, data availability, and clinical interpretability. No univariable screening or automated stepwise selection was used. Model A included age, sex, body mass index (BMI), serum creatinine, serum albumin, left ventricular ejection fraction (LVEF), New York Heart Association (NYHA) class III–IV, non-elective surgery, and previous cardiovascular surgery. Model B retained all Model A predictors and added cardiopulmonary bypass duration and intraoperative allogeneic transfusion. All coefficients were re-estimated in separate multivariable logistic-regression models; Model B was therefore not constructed by simply adding the two intraoperative terms to the linear predictor from Model A. Continuous predictors were modeled as linear terms. For estimation and reporting, age was scaled per 5 years, creatinine per 10 μmol/L, albumin per 5 g/L, LVEF per 5 percentage points, and cardiopulmonary bypass duration per 10 minutes; BMI was reported per 1 kg/m².

Data cleaning and missing data

Source records were checked for duplicate identifiers, cohort eligibility, coding consistency, impossible categories, and physiologically implausible values. Before model fitting, data-quality flags were generated from prespecified ranges and coding completeness: preoperative hemoglobin <50 or >250 g/L, creatinine <20 or >1,000 μmol/L, albumin <20 or >60 g/L, or missing surgical urgency/status. Flags prompted source-data review but did not automatically exclude a record from the primary analysis. Forty-five patients in the primary-analysis cohort had at least one flag (29 in 2024 and 16 in 2025): 18 had a hemoglobin-related flag, 20 an albumin-related flag, 7 a creatinine-related flag, and 1 missing surgical status; one patient had both hemoglobin- and albumin-related flags. In the primary analysis, out-of-range albumin, creatinine <20 μmol/L, and missing surgical status were treated as missing and included in multiple imputation; creatinine >1,000 μmol/L was retained because it could represent a genuine extreme value. Hemoglobin was not included in either prediction model; implausible hemoglobin values were treated as missing only for descriptive summaries. The sensitivity analysis excluded all 45 flagged records.

Missingness was summarized separately in the development and validation cohorts. Twenty multiply imputed datasets were generated for model predictors using iterative conditional models based on Bayesian ridge regression with posterior sampling. The imputation model was fitted only in the 2024 development cohort and applied to the 2025 cohort without refitting to preserve temporal separation. The outcome was not included in the imputation model. Binary predictors were returned to 0/1 after imputation using a threshold of 0.5. Model coefficients and variances were pooled using Rubin's rules [12,13].

Model development and internal validation

Both models were fitted using multivariable logistic regression. Continuous predictors were centered at development-cohort reference values and scaled as described above. The pooled coefficient vector was used to calculate predicted probabilities within each imputed dataset, and patient-level probabilities were then averaged across the 20 completed datasets. Apparent performance was first estimated in the development cohort. Internal validation used 1,000 bootstrap resamples to estimate optimism in the AUC, Brier score, calibration intercept, and calibration slope, with imputed datasets cycled across bootstrap repetitions. Optimism-corrected performance was calculated by subtracting mean optimism from apparent performance, consistent with resampling-based internal validation principles [9].

Temporal validation and model comparison

The frozen models were applied directly to the 2025 cohort. Discrimination was quantified using the AUC. Calibration was assessed using a plot based on deciles of predicted risk, CITL (ideal value, 0), and calibration slope (ideal value, 1), recognizing that calibration has multiple levels and cannot be fully characterized by a single statistic [10]. Overall prediction error was measured using the Brier score. Paired differences between Model B and Model A in AUC and Brier score were estimated using 2,000 patient-level bootstrap resamples, with percentile 95% confidence intervals (CIs) and two-sided bootstrap P values.

Clinical utility, sensitivity analyses, and postoperative outcomes

Decision-curve analysis (DCA) was used to quantify net benefit across threshold probabilities from 1% to 30% and to compare both models with treat-all and treat-none strategies [11]. Based on the LCOS incidence in this cohort and the intended clinical setting of early risk screening and intensified monitoring, thresholds of 5%, 10%, 15%, and 20% were prespecified for focused presentation. Differences in net benefit between Model B and Model A at these thresholds were evaluated using 2,000 bootstrap resamples.

Sensitivity analyses included complete-case analysis, restoration of patients with preoperative shock, and exclusion of all records carrying a data-quality flag. Postoperative complications were summarized descriptively by LCOS status, with absolute risk differences and 95% CIs; these concurrent associations were not interpreted causally. ICU and total hospital lengths of stay were summarized using medians and interquartile ranges.

Sample size and model complexity

All eligible records during the study period were included. We reported the number of outcome events, the conventional number of events per candidate parameter, apparent Cox–Snell and Nagelkerke R^2 values, and a heuristic shrinkage estimate. To avoid treating an events-per-variable rule as proof of sample-size adequacy, we also examined Riley target-shrinkage scenarios using a target shrinkage of 0.90 and anticipated Cox–Snell R^2 values of 0.05, 0.08, and 0.10 [14]. These post hoc scenarios were treated as model-complexity diagnostics rather than as a definitive prospective sample-size calculation.

Statistical analysis

Continuous variables were summarized as mean (standard deviation) or median [interquartile range], and categorical variables as number (percentage). Development and validation cohorts were compared descriptively using standardized mean differences (SMDs), without significance testing of baseline characteristics. All CIs were two-sided 95% intervals; $P<0.05$ was used for prespecified paired comparisons, without multiplicity adjustment for exploratory secondary analyses. Analyses were performed in Python 3.12.14 using pandas 2.2.3, NumPy 2.3.5, SciPy 1.17.0, and scikit-learn 1.8.0. Random seeds and executable analysis code were retained for reproducibility.

Electronic risk calculator

A browser-based electronic risk calculator implementing the two frozen logistic-regression equations was developed (Supplementary Tool 1). The tool centers on Model B, which showed better performance during temporal validation, while retaining Model A as the preoperative assessment option: Model A provides an estimate after entry of the shared preoperative predictors, whereas Model B provides the principal perioperative estimate after cardiopulmonary bypass duration and intraoperative allogeneic transfusion become available at the end of surgery. The calculator executes only the prespecified equations, does not refit either model or select predictors, and processes entered data locally in the browser. Because Models A and B were fitted separately and have different intercepts and coefficients, an individual patient's Model B estimate was

not constrained to exceed the Model A estimate. The two outputs represent stage-specific risk estimates and should not be interpreted as an additive score, a treatment effect, or a causal change in risk attributable to surgery.

Ethics

The study was approved by the Ethics Committee of Nanjing First Hospital (approval No. KY20220425-05). The requirement for individual informed consent was waived. All data used for analysis were de-identified before modeling.

Results

Study population and baseline characteristics

Among 1,903 screened records, 106 were excluded because of ineligible procedure type and 20 because of preoperative shock or missing shock status. The primary analysis therefore included 1,777 patients: 1,171 in the 2024 development cohort and 606 in the 2025 temporal validation cohort (Figure 1). LCOS occurred in 101 patients in the development cohort (8.6%) and 50 in the validation cohort (8.3%).

Most baseline characteristics were broadly similar across the two calendar periods (Table 1), while retaining the case-mix differences expected in real-world cohorts over time. The largest absolute SMDs were observed for smoking history (31.6% vs 20.4%; SMD, 0.259), chronic pulmonary disease (5.8% vs 2.1%; SMD, 0.188), albumin (39.7 vs 39.2 g/L; SMD, 0.127), peripheral arterial disease (3.6% vs 1.7%; SMD, 0.121), and non-elective surgery (6.3% vs 9.2%; SMD, −0.109). Albumin was missing in 18.5% of the development cohort and 19.5% of the validation cohort; missingness for all other model predictors was ≤2.3% in each cohort.

Model estimation and internal validation

In the development cohort, the apparent AUC was 0.726 for Model A and 0.775 for Model B. After 1,000 bootstrap resamples, optimism-corrected AUCs were 0.699 (95% CI, 0.649–0.753) and 0.749 (95% CI, 0.705–0.799), respectively (Table 2). Optimism-corrected calibration slopes were similar for the two models (Model A, 0.891; 95% CI, 0.717–1.117; Model B, 0.890; 95% CI, 0.729–1.091), and corrected CITL values were close to 0 for both models.

Full estimates for Models A and B are provided in Table 3, and their adjusted ORs with 95% CIs are displayed in Figure 5 to facilitate interpretation and subsequent validation. In Model B, the odds ratio (OR) for each 10-minute increase in cardiopulmonary bypass duration was 1.099 (95% CI, 1.055–1.145), and the OR for intraoperative allogeneic transfusion was 2.121 (95% CI, 1.100–4.089), indicating additional prognostic information beyond preoperative risk. The two frozen model equations are reported below. Model B had an apparent Cox–Snell R^2 of 0.081 and a heuristic shrinkage estimate of 0.889. In the Riley target-shrinkage scenario analysis, the observed development sample was broadly compatible with the prespecified model

complexity when the assumed model fit was close to that observed; 1,000 bootstrap resamples were additionally used to quantify and correct potential optimism.

Frozen Model A equation (binary variables coded 0/1):

$$\text{logit}(pA) = -3.239180 + 0.164724[(\text{age}-72)/5] + 0.267104(\text{female}) - 0.058056(\text{BMI}-24) + 0.020655[(\text{creatinine}-74.64)/10] - 0.119693[(\text{albumin}-39.8)/5] - 0.243551[(\text{LVEF}-62)/5] + 0.297397(\text{NYHA III–IV}) + 1.451349(\text{non-elective surgery}) + 0.163118(\text{previous cardiovascular surgery}).$$

Here, $pA = 1/[1 + \exp(-\text{logit}(pA))]$.

Frozen Model B equation (binary variables coded 0/1):

$$\text{logit}(pB) = -3.815158 + 0.183861[(\text{age}-72)/5] + 0.208737(\text{female}) - 0.063808(\text{BMI}-24) + 0.028801[(\text{creatinine}-74.64)/10] + 0.016327[(\text{albumin}-39.8)/5] - 0.242465[(\text{LVEF}-62)/5] + 0.039731(\text{NYHA III–IV}) + 1.181507(\text{non-elective surgery}) + 0.094730(\text{previous cardiovascular surgery}) + 0.094491[(\text{cardiopulmonary bypass duration}-111)/10] + 0.751682(\text{intraoperative allogeneic transfusion}).$$

Here, $pB = 1/[1 + \exp(-\text{logit}(pB))]$.

Temporal validation

In the 2025 temporal validation cohort, Model A had an AUC of 0.655 (95% CI, 0.571–0.729) and Model B an AUC of 0.756 (95% CI, 0.695–0.810) (Figure 2), with a paired AUC difference of 0.101 (95% CI, 0.045–0.161; P<0.001). The calibration slope of Model B was 0.858 (95% CI, 0.566–1.151), closer to the ideal value of 1 than that of Model A (0.684; 95% CI, 0.354–1.014). CITL values were −0.114 for Model A and −0.113 for Model B, both close to 0, indicating little overall systematic deviation between mean predicted risk and the observed event rate (Figure 3).

Brier scores were 0.0739 for Model A and 0.0725 for Model B, with a numerically lower value for Model B; the paired difference was −0.0014 (95% CI, −0.0042 to 0.0018; P=0.386). Across the complementary performance domains, the principal incremental advantages of Model B were seen in risk ranking, calibration trends, and lower-threshold clinical net benefit rather than a statistically significant reduction in overall prediction error.

Decision-curve and sensitivity analyses

At threshold probabilities of 5%, 10%, 15%, and 20%, the Model B−Model A differences in net benefit were 0.0110 (95% CI, 0.0034–0.0203), 0.0141 (95% CI, 0.0039–0.0257), 0.0035 (95% CI, −0.0059 to 0.0129), and −0.0017 (95% CI, −0.0095 to 0.0066), respectively (Figure 4). Thus, Model B showed clear incremental net benefit at 5% and 10%, an uncertain small gain at 15%, and no demonstrated advantage at 20%. These findings suggest that the incremental clinical value of Model B was concentrated in the lower threshold range, which is more closely aligned with early risk screening and intensified monitoring.

Model B performed consistently across sensitivity analyses: the AUC was 0.736 in complete cases, 0.759 after inclusion of patients with preoperative shock, and 0.757 after exclusion of the 45 records carrying a data-quality flag. The corresponding Model A AUCs were 0.596, 0.670, and 0.657, respectively.

Postoperative outcomes by LCOS status

Patients with LCOS had a greater postoperative clinical burden. Compared with patients without LCOS, rates of prolonged mechanical ventilation (>24 hours), acute kidney injury, infection, and in-hospital mortality were 37.1% vs 9.7%, 15.2% vs 3.3%, 60.3% vs 12.7%, and 19.9% vs 1.2%, respectively. The median duration of the first ICU stay was 4 days in patients with LCOS and 1 day in those without LCOS. These descriptive findings further underscore the clinical relevance of early risk stratification.

Discussion

Principal findings

This study shows that updating preoperative baseline risk at the end of surgery with newly available intraoperative information can produce consistent and clinically meaningful performance gains. After cardiopulmonary bypass duration and intraoperative allogeneic transfusion were added to Model A, the temporal-validation AUC increased from 0.655 to 0.756, an increment of 0.101. Model B also had a calibration slope closer to 1, a CITL that remained close to 0, and clear incremental net benefit at 5% and 10% thresholds. These advantages were maintained across multiple sensitivity analyses. The principal value of the study therefore lies in representing risk updating in the true perioperative sequence and showing that routine intraoperative information can provide stable and reproducible incremental predictive information beyond preoperative risk.

Relation to previous work

Previous LCOS prediction studies have largely focused on preoperative baseline characteristics or specific surgical populations. In isolated aortic valve surgery, for example, age, renal failure, reduced LVEF, shock, and sex have been associated with low-output risk [4], and more recent work has applied clinical scores and machine-learning models to LCOS prediction [6,7]. The present study extends this work by explicitly incorporating the timing of prediction into the design: 2024 data were used for model development and internal validation, 2025 data were reserved for temporal validation, and the preoperative and intraoperatively updated models were compared directly within the same validation cohort. This moves the clinical question from “who is at higher risk before surgery?” to “can risk stratification be meaningfully updated once new intraoperative information becomes available?”, which more closely reflects the temporal sequence of perioperative decision-making.

The intraoperative update is also clinically interpretable. Cardiopulmonary bypass duration can integrate information on procedural complexity and cumulative perioperative exposures such as ischemia–reperfusion, inflammation, and hemodilution, whereas intraoperative transfusion can capture bleeding, anemia,

hemodilution, procedural complexity, and treatment requirements. Previous studies similarly suggest that adding intraoperative information to preoperative baseline risk can improve prediction of postoperative cardiac dysfunction and other cardiac-surgical outcomes [5,8]. Importantly, the second stage of our framework uses only two routinely recorded variables and requires no additional testing or complex data acquisition, supporting practical implementation.

Interpretation of discrimination, calibration, and clinical utility

AUC, Brier score, and calibration capture different dimensions of model performance. The clearest incremental gain for Model B was improved risk ranking, while its Brier score remained close to that of Model A. Considered together with the calibration and DCA findings, these results suggest that the principal value of the additional intraoperative information lies in better distinguishing patients at relatively higher risk and providing a more refined basis for early postoperative monitoring intensity and resource preparation.

Calibration in temporal validation was directionally consistent with the discrimination results. Model B had a CITL of −0.113, close to the ideal value of 0, and a calibration slope of 0.858, closer to the ideal value of 1 than the 0.684 observed for Model A. DCA further showed that the incremental net benefit of Model B was concentrated mainly within the lower 5%–10% risk-threshold range, consistent with its intended use for early risk screening and intensified monitoring. Model B may therefore complement clinical judgment as a quantitative aid for early postoperative risk stratification and monitoring intensification.

Potential clinical application

The framework can be integrated naturally into existing perioperative workflows. Model A provides an initial risk estimate before surgery; by the end of surgery, cardiopulmonary bypass duration and transfusion status are available, allowing Model B to provide the principal perioperative risk estimate without waiting for additional tests. This estimate can serve as a quantitative reference for early postoperative hemodynamic surveillance, ICU preparedness, and stratification of bedside assessment frequency. To support individualized application, we developed one dual-mode browser-based calculator centered on Model B while retaining Model A for preoperative assessment (Supplementary Tool 1), rather than two competing stand-alone tools. The stage-specific probabilities may move upward or downward and should be interpreted at their respective information time points; the difference between Model B and Model A should not be interpreted as a treatment effect or a direct measure of clinical deterioration or improvement.

Strengths and limitations

Several aspects of this study directly strengthen the quality of the prediction modeling. First, predictors were prespecified on the basis of clinical knowledge without univariable screening or automated stepwise regression, supporting clinical interpretability and transparency. Second, the development and validation cohorts were strictly separated by calendar year, and the 2025 data were not used for model fitting, more closely reflecting application to future patients. Third, the analysis incorporated 20 multiple imputations, 1,000 bootstrap resamples for internal validation, paired model comparisons, and multiple sensitivity analyses, while

evaluating complementary measures of discrimination, calibration, Brier score, and clinical utility. Adding two routinely recorded intraoperative variables increased the temporal-validation AUC by approximately 0.10, with directionally consistent improvements in calibration and lower-threshold net benefit. Full estimates and frozen equations for both models are provided together with a comparative forest plot, and one dual-mode electronic risk calculator implements Model B as the principal output while retaining Model A as the preoperative assessment option, supporting independent reconstruction and subsequent external validation.

The study has a clearly defined scope of applicability. As a single-center retrospective cohort study, model performance in other centers and different surgical case mixes requires further evaluation. Nevertheless, the directionally consistent findings from bootstrap internal validation, temporal validation, complete-case analysis, inclusion of patients with preoperative shock, and exclusion of records carrying data-quality flags support the robustness of the principal findings. Continuous predictors were modeled as linear terms to constrain model complexity. Potential nonlinear effects can be evaluated in larger external cohorts, with recalibration performed where necessary. In addition, ascertainment of LCOS partly depended on the duration of inotropic support and records of mechanical circulatory support and may therefore have been influenced by differences in clinical practice. The temporal validation originated from the same center and covered only six consecutive months; it therefore cannot substitute for geographical external validation.

Implications and future directions

The next step should prioritize geographical external validation of Model B across other centers and different surgical case mixes, with prespecified assessment of discrimination, calibration, and clinical net benefit and recalibration where necessary. The transportability of Model A as the preoperative assessment option should be evaluated in parallel to determine the stability of the overall dual-mode application. The web-based calculator remains a research prototype; multicenter applicability, usability, and clinical impact require formal evaluation before implementation. If performance is maintained, prospective implementation studies could link the 5%–10% risk range to specific hemodynamic monitoring, ICU preparedness, or bedside assessment pathways and test whether Model B-centered risk assessment improves risk identification and resource allocation.

Conclusions

In older adults undergoing cardiac surgery, adding cardiopulmonary bypass duration and intraoperative allogeneic transfusion to a preoperative risk model increased the temporal-validation AUC for LCOS prediction from 0.655 to 0.756, produced a calibration slope closer to the ideal value, and yielded clear incremental net benefit at 5%–10% risk thresholds. These advantages were maintained across multiple sensitivity analyses. The findings support Model B as the principal candidate for subsequent external validation and clinical translation, with Model A retained for initial preoperative risk estimation. This dual-

mode framework provides a transparent, interpretable, and reproducible quantitative approach to early LCOS risk stratification, intensified monitoring, and resource allocation.

Declarations

Ethics approval and consent to participate: The study was approved by the Ethics Committee of Nanjing First Hospital (approval No. KY20220425-05). The requirement for individual informed consent was waived, and all data used for analysis were de-identified before modeling.

Consent for publication: Not applicable.

Clinical trial number: not applicable.

Data and code availability: Individual-level data are not publicly shared because of patient privacy and ethics approval restrictions.

Funding: No specific funding was received for this study.

Competing interests: The authors declare no competing interests.

Author contributions: Shengjie Ning and Junzhe Wang contributed to study conception and design, methodology, data curation and analysis, interpretation of results, and drafting of the manuscript. Ruoyu Zhang, Haoran Gan, Li Yin, Yifei Zhou,Dongming Liu, and Ji Wang contributed to clinical data collection and curation, verification of results, and review and revision of the manuscript. Zhibing Qiu contributed to study conception, overall supervision, project administration, interpretation of results, and critical revision of the manuscript for important intellectual content. All authors reviewed and approved the final manuscript and agreed to be accountable for the work.

Acknowledgments: None.

References

1. Lomivorotov VV, Efremov SM, Kirov MY, Fominskiy EV, Karaskov AM. Low-cardiac-output syndrome after cardiac surgery. J Cardiothorac Vasc Anesth. 2017;31:291–308. doi:10.1053/j.jvca.2016.05.029.
2. Schoonen A, van Klei WA, van Wolfswinkel L, van Loon K. Definitions of low cardiac output syndrome after cardiac surgery and their effect on incidence of intraoperative LCOS: literature review and cohort study. Front Cardiovasc Med. 2022;9:926957. doi:10.3389/fcvm.2022.926957.
3. Duncan AE, Kartashov A, Robinson SB, Randall D, Zhang K, Luber J, et al. Risk factors, resource use, and cost of postoperative low cardiac output syndrome. J Thorac Cardiovasc Surg. 2022;163:1890–1898.e10. doi:10.1016/j.jtcvs.2020.06.125.
4. Maganti MD, Rao V, Borger MA, Ivanov J, David TE. Predictors of low cardiac output syndrome after isolated aortic valve surgery. Circulation. 2005;112(Suppl 9):I448–I452. doi:10.1161/CIRCULATIONAHA.104.526087.
5. Widyastuti Y, Stenseth R, Berg KS, Pleym H, Wahba A, Videm V. Preoperative and intraoperative prediction of risk of cardiac dysfunction following open heart surgery. Eur J Anaesthesiol. 2012;29:143–151. doi:10.1097/EJA.0b013e32834de368.

6. Mendes MA, Fabre M, Amabili P, Jaquet O, Donneau AF, Bonhomme V, et al. Development and validation of a prediction score for low cardiac output syndrome after adult cardiac surgery. J Cardiothorac Vasc Anesth. 2023;37:1967–1973. doi:10.1053/j.jvca.2023.06.025.
7. Hong L, Xu H, Ge C, Tao H, Shen X, Song X, et al. Prediction of low cardiac output syndrome in patients following cardiac surgery using machine learning. Front Med (Lausanne). 2022;9:973147. doi:10.3389/fmed.2022.973147.
8. Mori M, et al. Toward dynamic risk prediction of outcomes after coronary artery bypass grafting: improving risk prediction with intraoperative events using gradient boosting. Circ Cardiovasc Qual Outcomes. 2021;14:e007363. doi:10.1161/CIRCOUTCOMES.120.007363.
9. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KGM. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. J Clin Epidemiol. 2003;56:441–447. doi:10.1016/S0895-4356(03)00047-7.
10. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol. 2016;74:167–176. doi:10.1016/j.jclinepi.2015.12.005.
11. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26:565–574. doi:10.1177/0272989X06295361.
12. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res. 2011;20:40–49. doi:10.1002/mpr.329.
13. Austin PC, White IR, Lee DS, van Buuren S. Missing data in clinical research: a tutorial on multiple imputation. Can J Cardiol. 2021;37:1322–1331. doi:10.1016/j.cjca.2020.11.010.
14. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. BMJ. 2020;368:m441. doi:10.1136/bmj.m441.

Tables

Table 1. Baseline and perioperative characteristics by study cohort

Characteristic	2024 development cohort (n=1,171)	2025 temporal validation cohort (n=606)	SMD
Age, years	72.1 ± 4.4	72.2 ± 4.5	-0.030
Female	460/1171 (39.3%)	237/606 (39.1%)	0.004
Body mass index, kg/m²	24.1 ± 3.2	24.3 ± 3.0	-0.060
Smoking history	369/1167 (31.6%)	123/604 (20.4%)	0.259
Diabetes mellitus	358/1171 (30.6%)	187/606 (30.9%)	-0.006
Hypertension	751/1171 (64.1%)	372/606 (61.4%)	0.057
Hyperlipidemia	110/1171 (9.4%)	59/606 (9.7%)	-0.012
Preoperative chronic dialysis	12/1171 (1.0%)	5/606 (0.8%)	0.021
Peripheral arterial disease	42/1171 (3.6%)	10/606 (1.7%)	0.121
Chronic pulmonary disease	68/1171 (5.8%)	13/606 (2.1%)	0.188
History of cerebrovascular events	142/1171 (12.1%)	70/606 (11.6%)	0.018
Heart failure	1111/1171 (94.9%)	568/606 (93.7%)	0.049
Atrial fibrillation or flutter	225/1171 (19.2%)	118/606 (19.5%)	-0.007
NYHA class III–IV	672/1171 (57.4%)	332/604 (55.0%)	0.049

Table 1 (continued)

Characteristic	2024 development cohort (n=1,171)	2025 temporal validation cohort (n=606)	SMD
Previous cardiovascular surgery	143/1171 (12.2%)	70/606 (11.6%)	0.020
Hemoglobin, g/L	128.3 ± 17.2	126.8 ± 17.0	0.091
Creatinine, μmol/L	74.6 [62.2, 91.5]	73.6 [62.0, 87.2]	0.036
Albumin, g/L	39.7 ± 3.4	39.2 ± 3.6	0.127
LVEF, %	57.5 ± 9.9	57.3 ± 9.7	0.027
Non-elective surgery	74/1170 (6.3%)	56/606 (9.2%)	-0.109

Table 1 (continued)

Characteristic	2024 development cohort (n=1,171)	2025 temporal validation cohort (n=606)	SMD
Procedure: isolated CABG	440/1171 (37.6%)	225/606 (37.1%)	0.009
Procedure: valve surgery	357/1171 (30.5%)	178/606 (29.4%)	0.024
Procedure: CABG plus valve surgery	155/1171 (13.2%)	89/606 (14.7%)	-0.042
Procedure: open aortic or combined surgery	189/1171 (16.1%)	94/606 (15.5%)	0.017
Procedure: other cardiac surgery	30/1171 (2.6%)	20/606 (3.3%)	-0.044
Cardiopulmonary bypass duration, min	111.0 [89.0, 141.5]	111.0 [89.0, 143.0]	-0.001
Intraoperative allogeneic transfusion	856/1171 (73.1%)	452/604 (74.8%)	-0.040
Aortic cross-clamp duration, min	81.0 [61.0, 108.0]	83.0 [62.0, 107.0]	-0.060

Data are mean ± standard deviation, median [interquartile range], or n/N (%). Under the coding used in this study, a positive SMD indicates a higher value or prevalence of the characteristic or exposure in the 2024 cohort. BMI, body mass index; CABG, coronary artery bypass grafting; LVEF, left ventricular ejection fraction; NYHA, New York Heart Association; SMD, standardized mean difference.

Table 2. Model performance in the development and temporal validation cohorts

Evaluation stage	Model	n	LCOS events	AUC (95% CI)	Brier score	CITL (95% CI)	Calibration slope (95% CI)
2024 development cohort, apparent	A	1171	101	0.726	0.0728	0.001（−0.212～0.214）	1.008（0.766～1.250）
2024 development cohort, optimism-corrected	A	1171	101	0.699（0.649～0.753）	0.0754	0.009（−0.208～0.268）	0.891（0.717～1.117）
2025 temporal validation	A	606	50	0.655（0.571～0.729）	0.0739	−0.114（−0.415～0.187）	0.684（0.354～1.014）
2024 development cohort, apparent	B	1171	101	0.775	0.0706	0.001（−0.216～0.218）	1.008（0.796～1.219）
2024 development cohort, optimism-corrected	B	1171	101	0.749（0.705～0.799）	0.0736	0.009（−0.225～0.269）	0.890（0.729～1.091）
2025 temporal validation	B	606	50	0.756（0.695～0.810）	0.0725	−0.113（−0.418～0.192）	0.858（0.566～1.151）

Confidence intervals for optimism-corrected development estimates were derived from 1,000 bootstrap resamples. Temporal-validation intervals were estimated in the 2025 cohort. The ideal CITL is 0 and the ideal calibration slope is 1. AUC, area under the receiver operating characteristic curve; CI, confidence interval; CITL, calibration-in-the-large.

Paired temporal-validation comparison: ΔAUC (Model B − Model A) = 0.101 (95% CI, 0.045–0.161; P<0.001); ΔBrier score = −0.0014 (95% CI, −0.0042 to 0.0018; P=0.386).

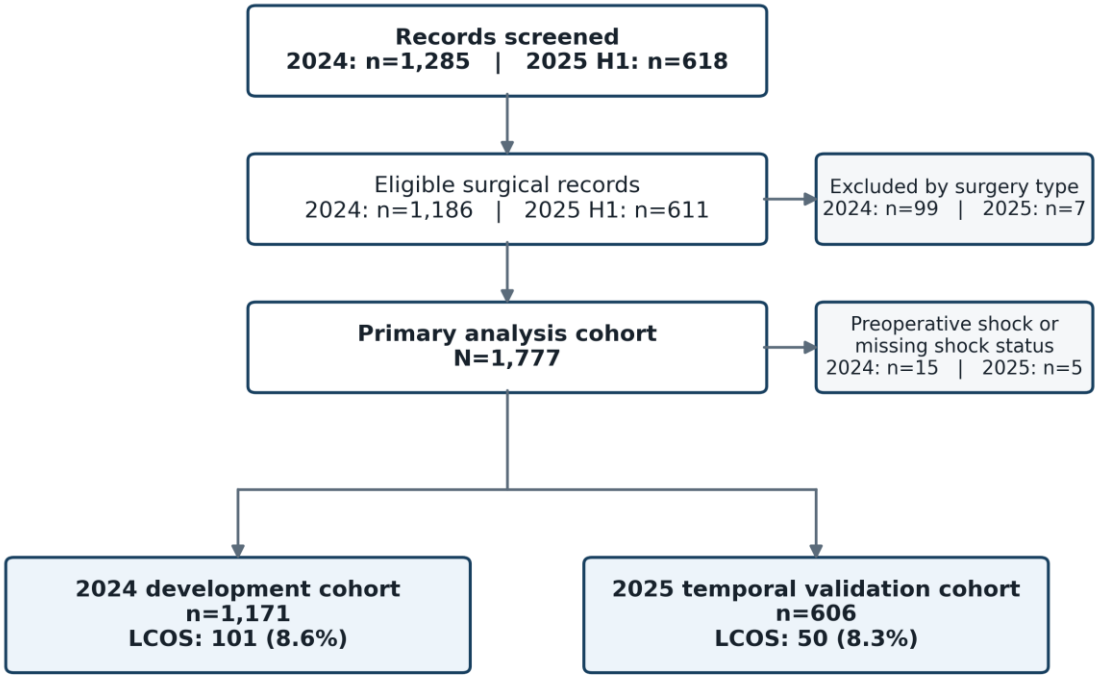
Table 3. Pooled multivariable logistic-regression estimates

Predictor	Model A OR (95% CI)	P value	Model B OR (95% CI)	P value
Age, per 5-year increase	1.179 (0.928–1.499)	0.178	1.202 (0.939–1.538)	0.144
Female sex (reference: male)	1.306 (0.828–2.060)	0.250	1.232 (0.773–1.964)	0.380
BMI, per 1 kg/m² increase	0.944 (0.880–1.012)	0.102	0.938 (0.874–1.007)	0.078
Creatinine, per 10 µmol/L increase	1.021 (1.000–1.042)	0.051	1.029 (1.007–1.051)	0.008
Albumin, per 5 g/L increase	0.887 (0.626–1.258)	0.501	1.016 (0.706–1.464)	0.930
LVEF, per 5-percentage-point increase	0.784 (0.713–0.862)	<0.001	0.785 (0.711–0.866)	<0.001
NYHA class III–IV	1.346 (0.838–2.163)	0.219	1.041 (0.637–1.699)	0.874
Non-elective surgery	4.269 (2.325–7.839)	<0.001	3.259 (1.746–6.086)	<0.001
Previous cardiovascular surgery	1.177 (0.647–2.143)	0.594	1.099 (0.594–2.035)	0.763
Cardiopulmonary bypass duration, per 10-min increase	—	—	1.099 (1.055–1.145)	<0.001
Intraoperative allogeneic transfusion	—	—	2.121 (1.100–4.089)	0.025

Estimates were pooled across 20 multiply imputed datasets using Rubin's rules. All predictors were retained according to the prespecified analysis plan irrespective of P value; P values are descriptive and were not used for variable selection. BMI, body mass index; CI, confidence interval; LVEF, left ventricular ejection fraction; NYHA, New York Heart Association; OR, odds ratio.

Figures and legends

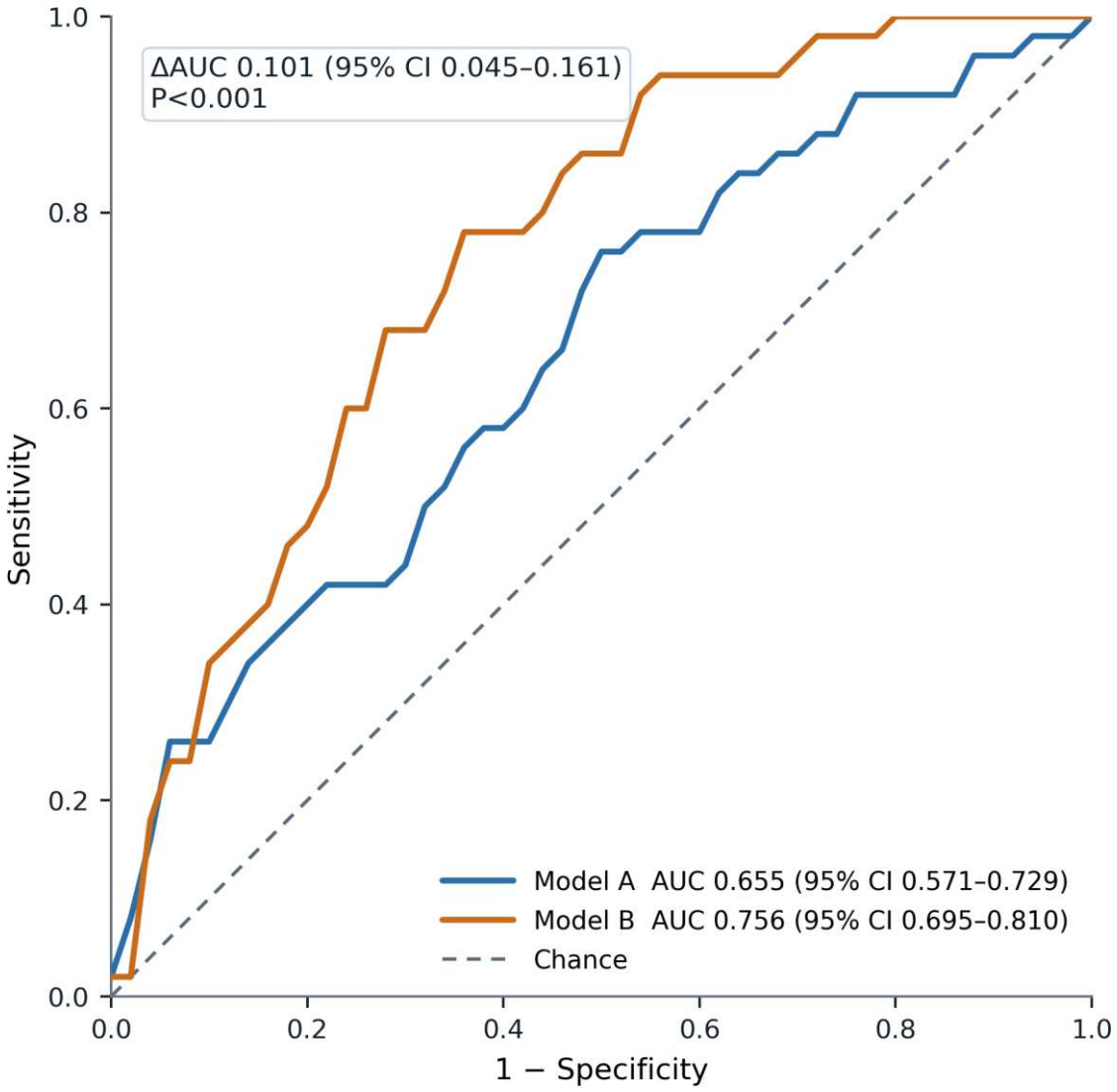
Figure 1. Study flow



LCOS, low cardiac output syndrome; H1, first half of the year.

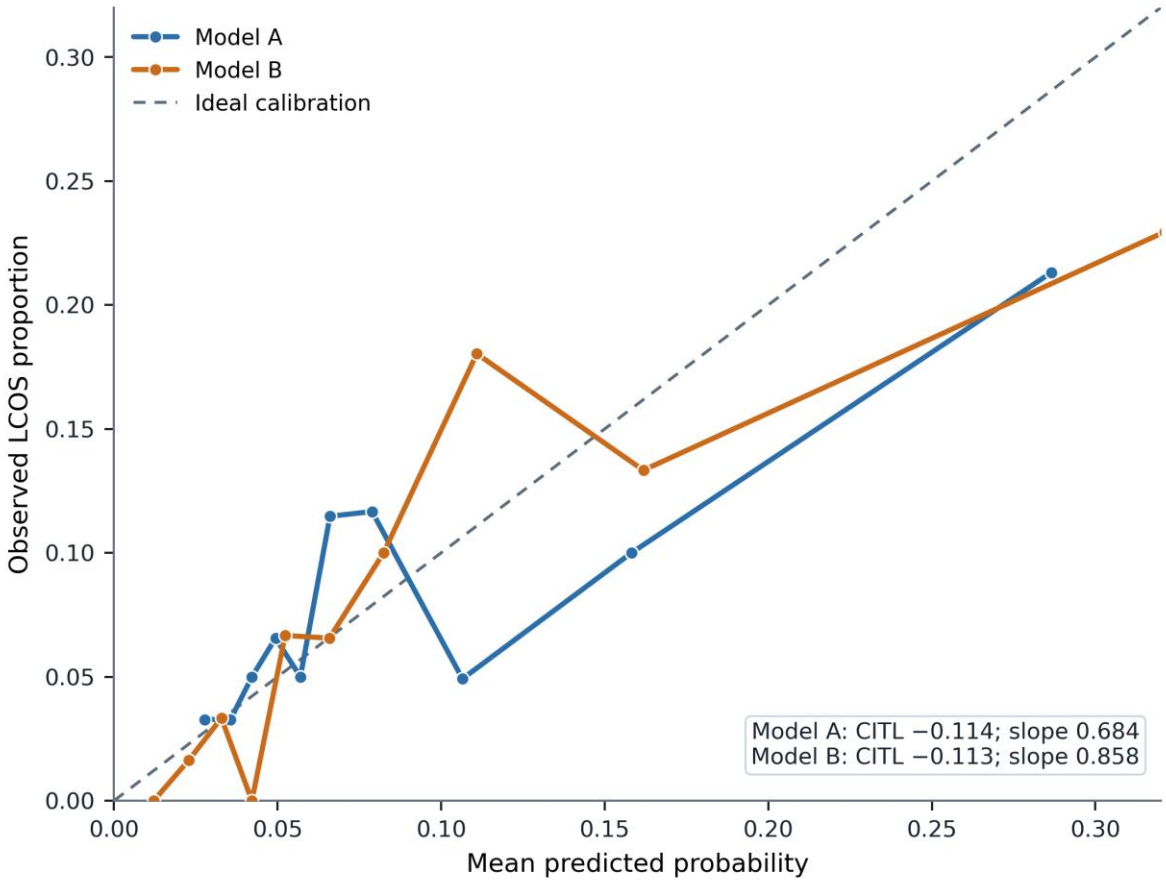
The 2024 cohort was used for model development and bootstrap internal validation; the non-overlapping 2025 cohort was reserved for temporal validation. H1, first half of the year; LCOS, low cardiac output syndrome.

Figure 2. Receiver operating characteristic curves in the 2025 temporal validation cohort



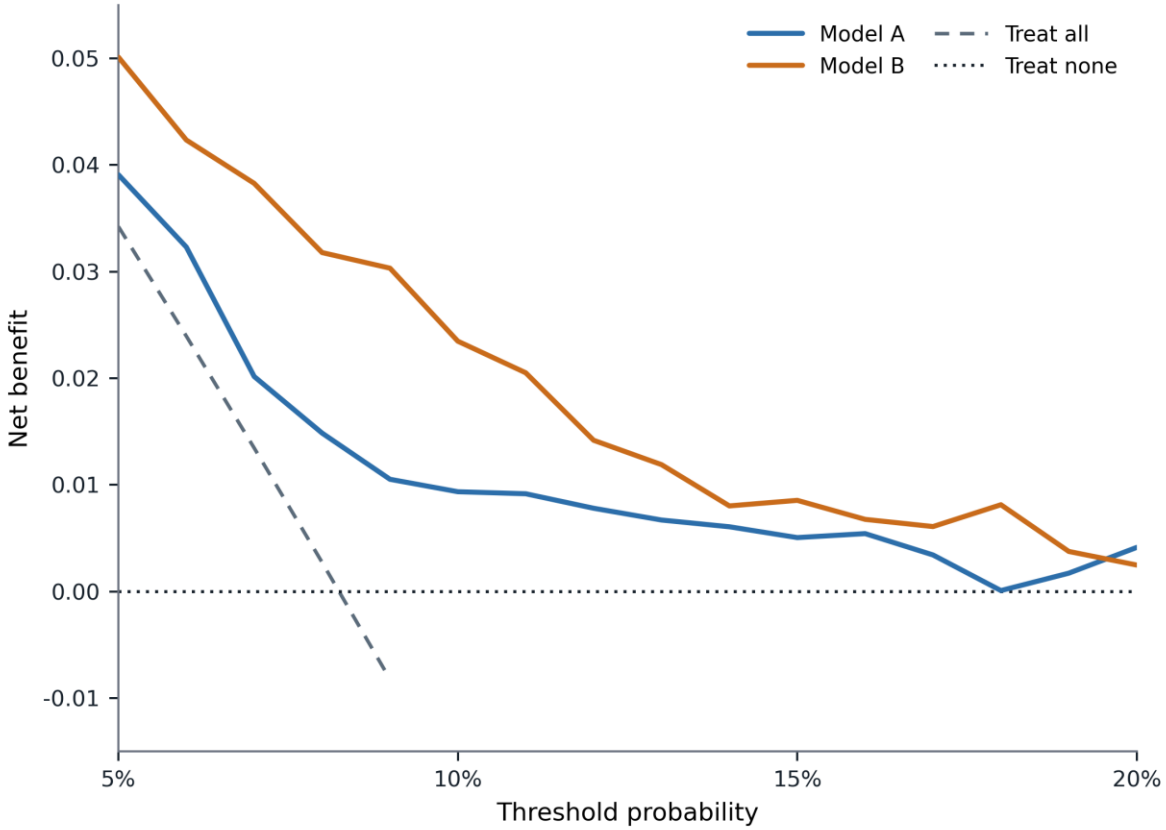
Model A was based on preoperative predictors; Model B additionally included cardiopulmonary bypass duration and intraoperative allogeneic transfusion. AUC, area under the receiver operating characteristic curve; CI, confidence interval.

Figure 3. Calibration by deciles of predicted risk in the 2025 temporal validation cohort



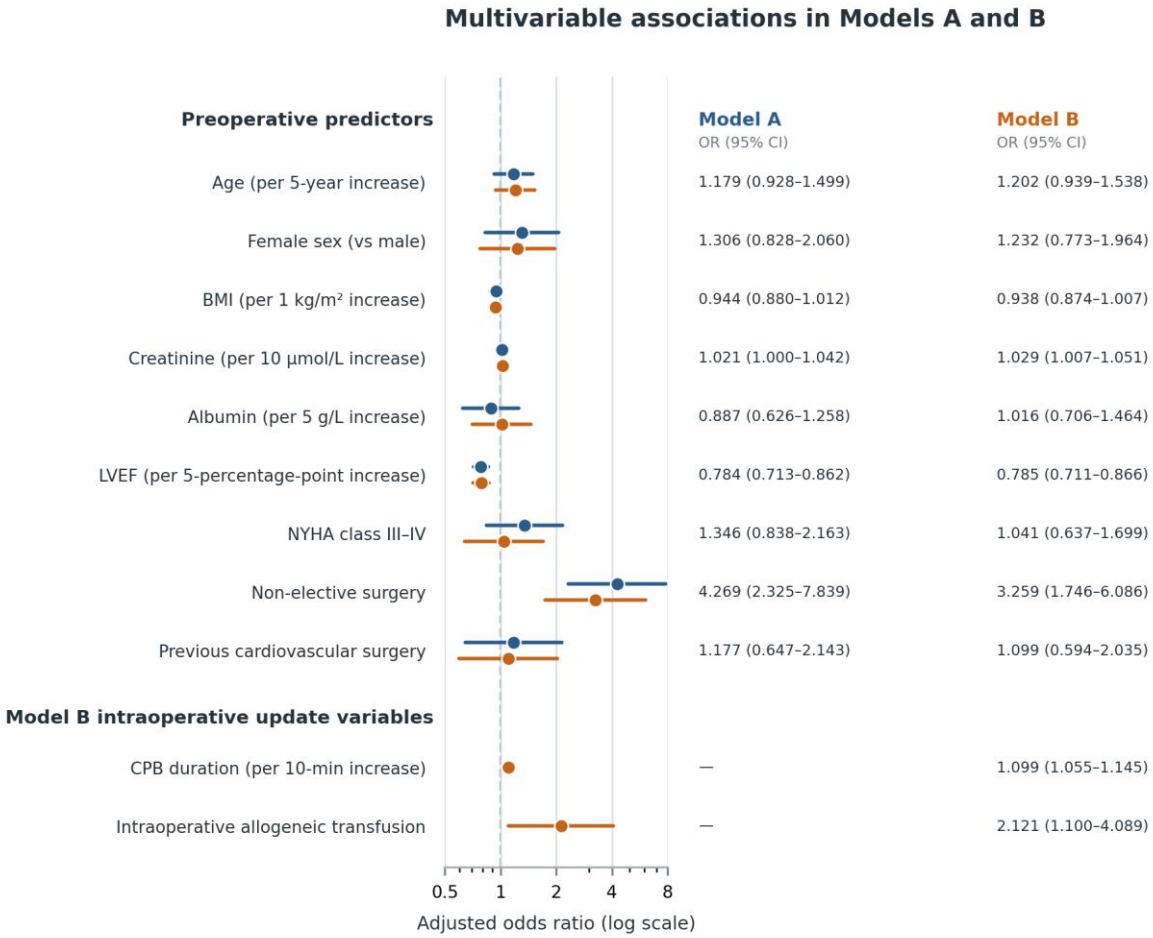
Points show the mean predicted risk and observed LCOS proportion within each risk decile. CITL, calibration-in-the-large; LCOS, low cardiac output syndrome.

Figure 4. Decision-curve analysis in the 2025 temporal validation cohort



Net benefit is shown across the clinically focused 5%–20% threshold range. Model B had greater net benefit than Model A at 5%, 10%, and 15%, but not at 20%; the 95% CI for the between-model difference excluded 0 only at 5% and 10%.

Figure 5. Adjusted odds ratios from Models A and B



Points denote adjusted odds ratios (ORs), horizontal lines denote 95% CIs, and the vertical dashed line indicates the null value of OR=1. Blue denotes Model A and orange denotes Model B; dashes in the Model A column identify the two intraoperative variables added only to Model B. Estimates were obtained from multivariable logistic-regression models pooled across 20 multiply imputed datasets using Rubin's rules. All predictors were prespecified and retained; the forest plot was not used for variable selection. BMI, body mass index; CI, confidence interval; CPB, cardiopulmonary bypass; LVEF, left ventricular ejection fraction; NYHA, New York Heart Association; OR, odds ratio.