

Original Research Article

PREDICTION OF EARLY MORBIDITY IN EUROSCORE-II AND STS SCORE IN PATIENTS UNDERGOING CABG SURGERY

Muhammad Saleem¹, Muhammad Awais Saeed², Syed Anees Ur Rehman Shah³, Mohsin Ahmed⁴, Muhammad Ahsan Fareed⁵, Momina Arshad⁶

¹FCPS, FRCS (Glasgow), FRCS (CTh) (England), Associate Professor Cardiac Surgery, FIC, Faisalabad, Pakistan.

²FCPS, Ex-Resident, PIC, Lahore, Pakistan.

³MS Cardiac Surgery, Senior Registrar, PIC, Lahore, Pakistan.

⁴Ex-Postgraduate Resident Cardiac Surgery, PIC, Lahore, Pakistan.

⁵Resident Interventional Cardiology, FIC, Faisalabad, Pakistan.

⁶SilverQuill Pakistan, Lahore, Pakistan.

Received : 08/06/2026
Received in revised form : 22/09/2026
Accepted : 03/10/2026

Corresponding Author:

Dr. Muhammad Saleem,
FCPS, FRCS (Glasgow), FRCS (CTh)
(England), Associate Professor Cardiac
Surgery, FIC, Faisalabad, Pakistan.
Email: drsaleem6001@hotmail.com

DOI: 10.70034/ijmedph.2026.4.83

Source of Support: Nil,

Conflict of Interest: None declared

Int J Med Pub Health
2026; 16 (4); 557-564

ABSTRACT

Background: Accurate preoperative risk stratification is important for clinical decision-making, perioperative planning, and quality assessment in patients undergoing coronary artery bypass grafting (CABG). EuroSCORE II and the Society of Thoracic Surgeons (STS) score are widely used cardiac surgical risk models, but their predictive performance may vary across populations and clinical settings. This study compared the performance of EuroSCORE II and the STS score for predicting early postoperative morbidity in patients undergoing CABG.

Materials and Methods: A prospective observational diagnostic-accuracy study was conducted at the Punjab Institute of Cardiology, Lahore, Pakistan, from February 2024 to February 2025. A total of 400 patients aged 20-70 years undergoing CABG were enrolled using purposive sampling. EuroSCORE II and STS risk scores were calculated preoperatively, and patients were followed for seven days after surgery to assess early postoperative morbidity. Predicted morbidity was compared with observed morbidity using 2×2 contingency tables. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall classification accuracy were calculated. Associations between early morbidity and demographic or clinical risk factors were also assessed, with statistical significance set at $p < 0.05$.

Results: The mean age of the study population was 53.23 ± 11.90 years, and 244 (61.0%) patients were male. Early postoperative morbidity occurred in 23 of 400 patients (5.8%). EuroSCORE II demonstrated a sensitivity of 73.9%, specificity of 98.4%, PPV of 73.9%, NPV of 98.4%, and overall accuracy of 97.0%. The STS score showed a sensitivity of 52.2%, specificity of 98.9%, PPV of 75.0%, NPV of 97.1%, and overall accuracy of 96.3%. Both scoring systems were significantly associated with observed morbidity ($p < 0.001$). Early morbidity was significantly associated with hypertension ($p = 0.007$), obesity ($p < 0.001$), and hyperlipidemia on Fisher's exact testing ($p = 0.048$), whereas age, sex, diabetes mellitus, and smoking status were not significantly associated with the outcome.

Conclusion: Both EuroSCORE II and the STS score demonstrated high specificity and overall classification accuracy for early postoperative morbidity after CABG. EuroSCORE II showed higher sensitivity, whereas STS demonstrated marginally higher specificity. These findings support the use of structured preoperative risk assessment while also emphasizing the need for larger multicentre studies with formal calibration and discrimination analyses before either model is used as a stand-alone predictor of early morbidity in this population.

Keywords: Coronary artery bypass grafting; EuroSCORE II; Society of Thoracic Surgeons score; risk prediction; postoperative morbidity; cardiac surgery; diagnostic accuracy.

INTRODUCTION

Coronary artery bypass grafting (CABG) remains an essential revascularization strategy for patients with complex coronary artery disease, yet perioperative mortality and major postoperative complications continue to influence clinical outcomes, resource utilization, and assessment of surgical quality.^[1-4] Because crude outcome rates are strongly affected by differences in baseline patient risk and operative complexity, risk-adjusted assessment is necessary for meaningful comparison of outcomes, informed consent, perioperative planning, and institutional benchmarking.^[1,4-6] Among available models, the European System for Cardiac Operative Risk Evaluation II (EuroSCORE II) and Society of Thoracic Surgeons (STS) risk models are the most extensively evaluated tools in contemporary cardiac surgery.^[1,3,5-8] EuroSCORE II was developed from an international cohort to estimate operative mortality, whereas the procedure-specific STS models estimate mortality as well as clinically relevant outcomes including stroke, renal failure, prolonged ventilation, reoperation, deep sternal wound infection, and prolonged hospitalization.^[3-5] Current coronary-revascularization guidance consequently recommends calculation of the STS score when CABG is being considered, while EuroSCORE II remains widely used internationally for operative risk stratification.^[4,5,9]

Despite their widespread adoption, neither model demonstrates uniformly reproducible performance across populations.^[1,10-14] A meta-analysis showed good overall discrimination for both EuroSCORE II and STS, without a statistically significant difference between their discriminatory performances, but identified important geographic variation in EuroSCORE II calibration.^[1] In 4,895 Indian CABG patients, STS and EuroSCORE II had similar discrimination (AUC 0.720 versus 0.713; $p=0.58$), although EuroSCORE II overestimated and STS underestimated operative mortality.^[15] Conversely, a Chinese CABG cohort demonstrated comparable in-hospital discrimination but somewhat better longer-term performance of EuroSCORE II, while an Australian study reported numerically higher discrimination with EuroSCORE II than STS without a significant difference.^[16,17] Recent evidence further illustrates substantial heterogeneity: EuroSCORE II overestimated mortality in a Swedish nationwide CABG cohort, underestimated risk in Indonesian CABG patients, and showed poor overall discrimination alongside STS in a contemporary Brazilian CABG cohort, whereas other Asian and European validations reported satisfactory discrimination but variable calibration across surgical-risk strata and urgency.^[10-14,18-22] These inconsistencies are clinically important because models derived predominantly from European and North American populations may perform differently when transported to populations

with different case mix, comorbidity burden, surgical practice, and perioperative care.^[1,13,15,23-25]

Accordingly, direct contemporary validation of EuroSCORE II and STS in the population in which they are intended to guide CABG decision-making is necessary rather than assuming that performance observed in their derivation populations is transferable.^[1,10,13,15,23] A previous Pakistani CABG study found that the STS model discriminated mortality better than the older additive and logistic EuroSCORE algorithms, but it preceded EuroSCORE II and contemporary changes in surgical practice, leaving an important local evidence gap regarding the comparative performance of the two currently used models.^[26] Furthermore, because discrimination alone does not ensure accurate individual risk estimation, comparative evaluation should consider agreement between predicted and observed outcomes and the ability of each score to identify patients who experience important early postoperative adverse outcomes.^[1,17,24,27] Therefore, the present study will determine and compare the predictive accuracy of EuroSCORE II and the STS score in patients undergoing CABG, providing locally applicable evidence on their usefulness for perioperative risk stratification. The findings may support more reliable preoperative counselling, risk-adjusted clinical decision-making, and quality assessment in patients undergoing CABG.

MATERIALS AND METHODS

Study Design and Setting

A prospective observational diagnostic-accuracy study was conducted in the Department of Cardiology at the Punjab Institute of Cardiology (PIC), Lahore, Pakistan, from February 2024 to February 2025. Patients aged 20-70 years who were scheduled to undergo coronary artery bypass grafting (CABG) were enrolled using a purposive sampling technique. The calculated sample size was 400 patients, based on a previously reported sensitivity of 63.5% and specificity of 71.6% for prediction of early postoperative morbidity, an anticipated morbidity prevalence of 13.2%, and a 5% margin of error. Patients with pre-existing renal disease, chronic obstructive pulmonary disease, left ventricular ejection fraction (LVEF) $\leq 35\%$, acute myocardial infarction requiring urgent surgical intervention, or concomitant CABG with valvular or vascular surgery were excluded. Written informed consent was obtained from all participants before enrollment. The study protocol was approved by the institutional ethics committee of PIC, Lahore.

Data Collection and Perioperative Assessment

Baseline demographic and clinical data were recorded before surgery, including age, sex, height, weight, body mass index (BMI), smoking status, and relevant comorbidities, including diabetes mellitus, hypertension, hyperlipidemia, and obesity.

Preoperative serum creatinine, creatinine clearance, and LVEF were also documented.

Intraoperative variables included cardiopulmonary bypass (CPB) time and aortic cross-clamp time. Postoperative variables included duration of mechanical ventilation, length of intensive care unit (ICU) stay, total hospital stay, and postoperative complications. Recorded postoperative complications included renal dysfunction and neurological events, including stroke and transient ischemic attack. Other predefined postoperative clinical abnormalities, including oxygen saturation <95%, were also documented where applicable.

All patients were followed for seven days after CABG for assessment of early postoperative morbidity. Patients who developed early morbidity were compared with those who did not with respect to demographic characteristics, cardiovascular risk factors, and perioperative variables.

Preoperative Risk Assessment

The European System for Cardiac Operative Risk Evaluation II (EuroSCORE II) and Society of Thoracic Surgeons (STS) risk scores were calculated preoperatively for each patient using their respective web-based risk calculators. EuroSCORE II was calculated using the official EuroSCORE II calculator, while the STS score was determined using the STS Adult Cardiac Surgery Risk Calculator.

For EuroSCORE II, predicted risk was categorized as low (<4%), intermediate (4%-9%), or high (>9%). For the primary diagnostic-accuracy analysis, a EuroSCORE II value $\geq 4\%$ was classified as a positive prediction for early morbidity, whereas a value <4% was classified as negative.

For the STS score, predicted risk was categorized as low (<4%), intermediate (4%-8%), or high (>8%). An STS score $\geq 4\%$ was classified as a positive prediction for early morbidity, while a score <4% was classified as negative. Predicted morbidity according to each scoring system was subsequently compared with the observed postoperative morbidity status.

Outcome Assessment

The primary outcome was early postoperative morbidity occurring within seven days after CABG. The predictive performance of EuroSCORE II and the STS score was assessed by comparing the preoperatively predicted morbidity classification with observed postoperative morbidity. Patients were therefore classified as true positive, false positive, true negative, or false negative for each risk model.

Diagnostic performance was summarized using sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall classification accuracy.

Statistical Analysis

Data were entered and analyzed using IBM SPSS Statistics for Windows, version 26.0 (IBM Corp., Armonk, NY, USA). Continuous variables, including age, BMI, serum creatinine, CPB time, aortic cross-clamp time, ICU stay, and total hospital stay, were summarized as mean $\pm$ standard deviation (SD). Categorical variables, including sex, cardiovascular risk factors, postoperative complications, and predicted morbidity according to EuroSCORE II and the STS score, were reported as frequencies and percentages.

Two-by-two contingency tables were constructed for each scoring system using observed early morbidity as the reference outcome. Sensitivity, specificity, PPV, NPV, and overall accuracy were calculated from these tables. Associations between predicted and observed morbidity were evaluated using the appropriate categorical statistical test. The data were further stratified according to age, sex, diabetes mellitus, hypertension, hyperlipidemia, smoking status, and obesity, and postoperative morbidity was compared across these strata. Pearson's chi-square test was used when its assumptions were satisfied, while Fisher's exact test was used when expected cell frequencies were small. A two-sided p-value <0.05 was considered statistically significant.

RESULTS

A total of 400 patients undergoing coronary artery bypass grafting were included. The mean age was 53.23 ± 11.90 years, with an age range of 24-70 years. Of the study population, 244 (61.0%) were male and 156 (39.0%) were female. The mean height was 166.09 ± 9.06 cm, mean weight was 75.10 ± 12.10 kg, and mean body mass index was 27.60 ± 8.30 kg/m². The mean serum creatinine concentration was 1.05 ± 0.40 mg/dL. Hypertension was present in 203 (50.8%) patients, diabetes mellitus in 180 (45.0%), smoking in 143 (35.8%), hyperlipidemia in 385 (96.3%), and obesity in 159 (39.8%) patients. [Table 1]

Table 1: Baseline demographic and clinical characteristics of the study population

Characteristic	Value
Age, years	53.23 ± 11.90
Age range, years	24-70
Male sex	244 (61.0)
Female sex	156 (39.0)
Height, cm	166.09 ± 9.06
Weight, kg	75.10 ± 12.10
BMI, kg/m ²	27.60 ± 8.30
Serum creatinine, mg/dL	1.05 ± 0.40
Hypertension	203 (50.8)
Diabetes mellitus	180 (45.0)

Current smoking	143 (35.8)
Hyperlipidemia	385 (96.3)
Obesity	159 (39.8)

Data are presented as mean $\pm$ standard deviation or n (%), unless otherwise stated. BMI: body mass index. The mean cardiopulmonary bypass time was 92.90 ± 23.30 minutes, while the mean aortic cross-clamp time was 43.70 ± 9.90 minutes. Prolonged postoperative ventilation of 24 hours or longer was

required in 21 (5.3%) patients. The mean intensive care unit stay was 4.18 ± 1.20 days, and the mean total hospital stay was 9.10 ± 1.69 days. Postoperative renal failure occurred in seven (1.8%) patients, while neurological complications occurred in five (1.3%) patients. [Table 2]

Table 2: Intraoperative and postoperative characteristics

Parameter	Value
Cardiopulmonary bypass time, min	92.90 ± 23.30
Aortic cross-clamp time, min	43.70 ± 9.90
Prolonged ventilation ≥ 24 h	21 (5.3)
ICU stay, days	4.18 ± 1.20
Total hospital stay, days	9.10 ± 1.69
Renal failure	7 (1.8)
Neurological complication	5 (1.3)

Data are presented as mean $\pm$ standard deviation or n (%). ICU: intensive care unit.

Early morbidity was observed in 23 of 400 patients (5.8%). EuroSCORE II classified 23 patients as having predicted morbidity, of whom 17 were true-positive and six were false-positive cases. Among the 377 patients classified as not having predicted morbidity, six were false-negative and 371 were true-

negative cases. This corresponded to a sensitivity of 73.9%, specificity of 98.4%, positive predictive value of 73.9%, negative predictive value of 98.4%, and overall classification accuracy of 97.0%. The association between EuroSCORE II classification and observed early morbidity was statistically significant ($p < 0.001$). [Table 3]

Table 3: Predictive performance of EuroSCORE II and STS score for early morbidity

Performance measure	EuroSCORE II	STS score
True positive, n	17	12
False positive, n	6	4
False negative, n	6	11
True negative, n	371	373
Sensitivity, %	73.9	52.2
Specificity, %	98.4	98.9
Positive predictive value, %	73.9	75.0
Negative predictive value, %	98.4	97.1
Overall accuracy, %	97.0	96.3
Association with observed morbidity, p-value	< 0.001	< 0.001

EuroSCORE II: European System for Cardiac Operative Risk Evaluation II; STS: Society of Thoracic Surgeons.

The STS score classified 16 patients as having predicted morbidity, including 12 true-positive and four false-positive cases. Among the 384 patients classified as not having predicted morbidity, 11 were false-negative and 373 were true-negative cases. The corresponding sensitivity was 52.2%, specificity was 98.9%, positive predictive value was 75.0%, negative predictive value was 97.1%, and overall classification accuracy was 96.3%. The association between STS classification and observed early morbidity was statistically significant ($p < 0.001$) (Table 3). The comparative sensitivity, specificity, and overall accuracy of both scoring systems are presented graphically in Figure 1.

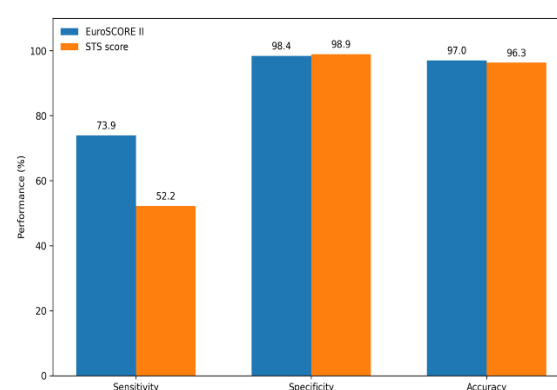


Figure 1: Diagnostic performance of EuroSCORE II and STS score for prediction of early morbidity

EuroSCORE II: European System for Cardiac Operative Risk Evaluation II; STS: Society of Thoracic Surgeons.

Following stratification, early morbidity occurred in three of 31 patients aged 20-30 years (9.7%), one of 25 patients aged 31-40 years (4.0%), seven of 69

patients aged 41-50 years (10.1%), five of 143 patients aged 51-60 years (3.5%), and seven of 132 patients aged 61-70 years (5.3%); the distribution did not differ significantly across age categories ($p=0.301$). Early morbidity occurred in 16 of 244 male patients (6.6%) and seven of 156 female patients (4.5%; $p=0.386$). The corresponding frequencies were 11 of 180 patients with diabetes (6.1%) and 12 of 220 without diabetes (5.5%; $p=0.779$).

Early morbidity occurred in 18 of 203 hypertensive patients (8.9%) compared with five of 197 non-

hypertensive patients (2.5%; $p=0.007$). Among patients with hyperlipidemia, 20 of 385 (5.2%) developed early morbidity compared with three of 15 patients without hyperlipidemia (20.0%; Fisher's exact $p=0.048$). Morbidity occurred in four of 143 smokers (2.8%) compared with 19 of 257 non-smokers (7.4%; $p=0.058$). Among patients classified as obese, 17 of 159 (10.7%) experienced early morbidity compared with six of 241 non-obese patients (2.5%; $p<0.001$). [Table 4]

Table 4: Early morbidity according to demographic and clinical characteristics

Variable	Early morbidity, n (%)	No early morbidity, n (%)	p-value
Age, years			0.301
20-30	3 (9.7)	28 (90.3)	
31-40	1 (4.0)	24 (96.0)	
41-50	7 (10.1)	62 (89.9)	
51-60	5 (3.5)	138 (96.5)	
61-70	7 (5.3)	125 (94.7)	
Sex			0.386
Male	16 (6.6)	228 (93.4)	
Female	7 (4.5)	149 (95.5)	
Diabetes mellitus			0.779
Yes	11 (6.1)	169 (93.9)	
No	12 (5.5)	208 (94.5)	
Hypertension			0.007
Yes	18 (8.9)	185 (91.1)	
No	5 (2.5)	192 (97.5)	
Hyperlipidemia			0.048*
Yes	20 (5.2)	365 (94.8)	
No	3 (20.0)	12 (80.0)	
Smoking			0.058
Yes	4 (2.8)	139 (97.2)	
No	19 (7.4)	238 (92.6)	
Obesity			<0.001
Yes	17 (10.7)	142 (89.3)	
No	6 (2.5)	235 (97.5)	

Pearson's chi-square test was used except where noted. *Fisher's exact test was used for hyperlipidemia because of the small expected cell frequency.

DISCUSSION

The present study evaluated the predictive performance of EuroSCORE II and the Society of Thoracic Surgeons (STS) score for early postoperative morbidity after coronary artery bypass grafting (CABG) in a contemporary Pakistani cohort. Both models showed high specificity and overall classification accuracy, but EuroSCORE II identified a greater proportion of patients who subsequently developed early morbidity, with a sensitivity of 73.9% compared with 52.2% for the STS score. Specificity was similar for EuroSCORE II and STS (98.4% and 98.9%, respectively), and their overall classification accuracies were likewise close (97.0% and 96.3%). These findings address an important local evidence gap because comparative performance of contemporary cardiac surgical risk models cannot be assumed to remain constant across populations, institutions, or outcome definitions. The results therefore suggest that, using the prespecified thresholds employed in this cohort, both tools were effective at identifying patients without subsequent

early morbidity, whereas EuroSCORE II captured a larger proportion of those who experienced an event. This distinction is clinically relevant because high specificity alone may yield apparently high overall accuracy when the outcome is uncommon; early morbidity occurred in only 5.8% of the cohort, making sensitivity and predictive values particularly important when judging practical performance.

The observed performance should nevertheless be interpreted in the context of how prediction models behave when transported beyond their derivation populations. Risk models such as EuroSCORE II and STS combine multiple baseline and procedural characteristics into estimates derived from populations with specific case mixes, treatment pathways, and underlying event rates. A large systematic review and meta-analysis by Sinha and colleagues found broadly good discrimination for both EuroSCORE II and STS in adult cardiac surgery, with pooled areas under the receiver-operating-characteristic curve of 0.783 and 0.757, respectively, and no statistically significant difference between them; importantly, EuroSCORE

II calibration varied significantly across geographical regions.^[1] This geographical heterogeneity provides a plausible framework for the present findings and reinforces the need for local validation rather than simple adoption of performance estimates derived elsewhere. Moreover, the current study evaluated dichotomised predictions against early morbidity rather than comparing continuous predicted probabilities with operative mortality. Sensitivity, specificity, and classification accuracy in the present analysis therefore describe performance at the selected $\geq 4\%$ thresholds and should not be interpreted as equivalent to discrimination measured by the area under a receiver-operating-characteristic curve or calibration assessed by observed-to-expected event ratios.

Several previous studies support the broader finding that both EuroSCORE II and STS can provide clinically useful risk stratification after CABG, while often demonstrating relatively small differences in overall performance. Shales and colleagues evaluated 4,895 Indian patients undergoing CABG and reported satisfactory and nearly identical discrimination for operative mortality, with an area under the curve of 0.713 for EuroSCORE II and 0.720 for STS; the difference was not significant ($p=0.58$).^[15] Their observation that neither model clearly dominated the other is consistent with the present study's very similar overall classification accuracies and specificities. Gao and colleagues likewise compared EuroSCORE II and STS in patients undergoing CABG and found that both models had useful predictive ability for in-hospital and medium-term mortality, although performance varied according to the outcome and follow-up period.^[16] The meta-analysis by Sinha et al further supports this pattern, demonstrating good overall discrimination for both scores without compelling evidence of a clinically dominant model across pooled cardiac-surgical populations.^[1] Similarly, Koo and colleagues identified EuroSCORE II and STS among the better-performing contemporary cardiac-surgical risk models in an Australian population, while emphasizing that optimal thresholds are population dependent.^[7] Taken together, these studies are compatible with the present observation that the two systems achieved comparable overall classification performance, while individual components of performance—most notably sensitivity—could differ. The current findings extend this evidence to early postoperative morbidity in a Pakistani CABG population rather than operative mortality alone, although this endpoint difference limits direct numerical comparison with previous validation studies.

Other reports have shown materially different relative performance, emphasizing the instability of risk-model transportability across clinical settings. In an earlier Pakistani study of 380 patients undergoing isolated CABG, Qadir and colleagues found that the STS algorithm performed better than the additive and logistic versions of the original EuroSCORE for

mortality prediction; the reported areas under the curve were 0.899 for STS, 0.866 for additive EuroSCORE, and 0.842 for logistic EuroSCORE, with STS also demonstrating the strongest calibration.^[25] Although that study evaluated earlier EuroSCORE formulations rather than EuroSCORE II, its relative advantage for STS contrasts with the higher sensitivity observed for EuroSCORE II in the present cohort. Differences in score generation, temporal changes in CABG practice, case mix, and the use of mortality rather than morbidity as the outcome may account for part of this discrepancy. More recent studies have also documented substantial variation. Silverborn and colleagues, in a nationwide Swedish cohort of 14,118 CABG patients, reported good overall discrimination for EuroSCORE II (AUC 0.82) but considerable overestimation of mortality, with an overall observed-to-expected ratio of 0.58; performance also varied by age and baseline surgical risk.^[18] By contrast, Wolf and Amato reported relatively poor overall discrimination in 438 Brazilian patients undergoing isolated CABG, with AUCs of 0.697 for EuroSCORE II and 0.646 for STS, although performance improved substantially among clinically stable patients.^[20] The Indian study by Shales et al also found systematic differences in calibration: EuroSCORE II tended to overestimate mortality whereas STS underestimated it.^[15] These contrasting observations demonstrate that a model can show reasonable discrimination yet remain poorly calibrated, or can perform differently across subgroups despite acceptable overall metrics. The high classification accuracy in the present study should therefore not be interpreted as evidence that either model is universally well calibrated, particularly because calibration was not formally assessed.

Beyond model performance, the study identified variation in observed morbidity across several clinical risk strata. Hypertension and obesity were associated with higher rates of early morbidity, whereas age, sex, diabetes, and smoking status were not significantly associated with the outcome in the stratified analyses. The association with hypertension is clinically plausible given its relationship with vascular dysfunction, left ventricular remodeling, renal vulnerability, and perioperative haemodynamic instability, while obesity may capture a cluster of metabolic, respiratory, and technical factors relevant to postoperative recovery. However, these associations should be interpreted cautiously because the study was not designed primarily to establish causal risk factors, and no multivariable adjustment was undertaken to determine whether hypertension or obesity independently predicted morbidity. The hyperlipidemia finding warrants even greater caution: morbidity was numerically more frequent among the small group without documented hyperlipidemia than among those with hyperlipidemia, and only 15 patients were classified as not having hyperlipidemia. The small cell size

makes this estimate unstable and susceptible to sparse-data effects. Similarly, the absence of statistically significant associations for established cardiovascular risk factors such as diabetes or age should not be regarded as evidence of no biological effect; the low number of morbidity events substantially limited statistical power for subgroup comparisons.

An important feature of the present results is the difference between sensitivity and overall accuracy. Although both scoring systems exceeded 96% overall classification accuracy, this figure was strongly influenced by the predominance of patients without early morbidity. With only 23 morbidity events among 400 participants, a model classifying most patients as event-free would already achieve high apparent accuracy. Sensitivity therefore provides additional information: EuroSCORE II identified 17 of the 23 morbidity cases, whereas STS identified.^[12] Conversely, STS generated slightly fewer false-positive predictions and consequently showed marginally higher specificity. Positive predictive values were similar, while negative predictive values were high for both models. These data indicate that the apparent similarity in overall accuracy masks meaningful differences in how the scores classified the relatively small group who experienced postoperative morbidity. A future analysis incorporating receiver-operating-characteristic curves, confidence intervals around sensitivity and specificity, and direct statistical comparison of paired model predictions would permit a more complete assessment of whether these differences exceed what might be expected by chance.

The study has several strengths. It included 400 consecutively assessed CABG candidates over a defined contemporary period at a major tertiary cardiac centre and applied both risk scores preoperatively to the same patients, allowing within-cohort comparison against prospectively observed postoperative outcomes. Uniform postoperative follow-up during the first week reduced differential outcome ascertainment, and the study provides contemporary data from Pakistan, a setting in which external validation of EuroSCORE II and STS remains limited. The direct calculation of true-positive, false-positive, false-negative, and true-negative classifications also provides clinically interpretable measures of model performance rather than relying solely on an aggregate association statistic. In addition, evaluation within a local population is particularly relevant because international studies have consistently shown that the calibration and discrimination of cardiac surgical risk models vary across geographical regions and risk strata.^[1,15,18,20]

Several limitations should also be acknowledged. First, this was a single-centre study using purposive sampling, which may limit external validity and introduce selection effects. Second, only 23 early morbidity events occurred; consequently, estimates of sensitivity and subgroup associations are based on

relatively few events and may have wide confidence intervals. Third, the primary endpoint requires particular caution. EuroSCORE II was developed principally to estimate operative mortality, whereas the STS models generate procedure-specific estimates for mortality and multiple morbidity outcomes. In the present study, EuroSCORE II and STS values were dichotomised using a threshold of 4% and evaluated against early postoperative morbidity. This approach is pragmatic but does not constitute conventional external validation of either model's original endpoint, and the selected cut-off requires independent validation. Fourth, the analysis did not assess calibration using calibration plots, calibration slope/intercept, Brier score, or observed-to-expected ratios, nor did it evaluate discrimination using receiver-operating-characteristic curves. Consequently, the reported sensitivity, specificity, and accuracy characterize performance only at the prespecified thresholds. Fifth, the study evaluated morbidity within seven days; events occurring later during the conventional 30-day postoperative period were not captured. Finally, the stratified analyses were unadjusted and involved multiple comparisons, so individual associations should be regarded as exploratory.

These findings have several implications for further research. Larger multicentre studies from Pakistan and other South Asian populations should externally validate EuroSCORE II and the contemporary STS models using their intended outcome definitions and should report discrimination, calibration, overall prediction error, and clinically relevant decision thresholds. Separate analyses for mortality and major morbidity would avoid conflating distinct endpoints, while model recalibration may be required if systematic differences between predicted and observed risks are demonstrated. Future studies should also assess whether locally prevalent clinical characteristics improve prediction when added to established models and should use sufficient event numbers to permit robust multivariable analysis and subgroup validation.

CONCLUSION

Both EuroSCORE II and the Society of Thoracic Surgeons (STS) score demonstrated high specificity and overall classification accuracy for early postoperative morbidity in patients undergoing coronary artery bypass grafting. At the prespecified threshold, EuroSCORE II showed greater sensitivity than the STS score, while both models produced similarly high negative predictive values. These findings indicate that structured preoperative risk assessment can assist in identifying patients at increased risk of adverse early postoperative outcomes; however, the performance of these models is dependent on the population, outcome definition, and threshold applied.

Because EuroSCORE II was originally developed for operative mortality and the STS system provides outcome-specific risk estimates, the present findings should be interpreted as local comparative evidence rather than definitive validation of either model for early morbidity. Larger multicentre studies incorporating receiver-operating-characteristic analysis, formal calibration assessment, confidence intervals for diagnostic-performance measures, and externally validated morbidity definitions are required before either score can be recommended as a stand-alone tool for predicting early postoperative morbidity in the Pakistani CABG population.

REFERENCES

1. Sinha S, Dimagli A, Dixon L, Gaudino M, Caputo M, Vohra HA, et al. Systematic review and meta-analysis of mortality risk prediction models in adult cardiac surgery. *Interact Cardiovasc Thorac Surg.* 2021;33(5):673-686. doi: 10.1093/icvts/ivab151.
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. *JACC.* 2022;79(2):e21-e129. <https://doi.org/10.1016/j.jacc.2021.09.006>.
3. Nashef SAM, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, et al. EuroSCORE II. *Eur J Cardiothorac Surg.* 2012;41(4):734-744. doi:10.1093/ejcts/ezs043.
4. Shahian DM, Jacobs JP, Badhwar V, Kurlansky PA, Furnary AP, Cleveland JC Jr, et al. The Society of Thoracic Surgeons 2018 adult cardiac surgery risk models: Part 1—background, design considerations, and model development. *Ann Thorac Surg.* 2018;105(5):1411-1418. doi: 10.1016/j.athoracsur.2018.03.002.
5. STS.org [Internet]. 2026 [cited 2026 Sep 30]. ACS/Operative Risk Calculator | STS. Available from: <https://www.sts.org/resources/acs-operative-risk-calculator>.
6. 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):573-579. doi: 10.1016/j.athoracsur.2016.01.105.
7. Koo SK, Dignan R, Lo EYW, Williams C, Xuan W. Evidence-based determination of cut-off points for increased cardiac-surgery mortality risk with EuroSCORE II and STS. *Heart Lung Circ.* 2022;31(4):590-601. doi: 10.1016/j.hlc.2021.08.026.
8. Saka E, Öztürk E, Yüksel AE, Kocabaş NS. Comparison of EuroSCORE II and STS risk scoring systems in patients who underwent open-heart surgery. *Türk J Anaesthesiol Reanim.* 2025;53(4):163-169. doi:10.4274/TJAR.2025.241778.
9. Shahian DM, Jacobs JP, Badhwar V, Kurlansky PA, Furnary AP, Cleveland JC, et al. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 1—Background, Design Considerations, and Model Development. *Ann Thorac Surg.* 2018;105(5):1411-8. doi: 10.1016/j.athoracsur.2018.03.002.
10. Kuplay H, Bayer Erdoğan S, Baştopçu M, Karpuzoğlu E, Er H. Performance of the EuroSCORE II and the STS score for cardiac surgery in octogenarians. *Türk Gogus Kalp Damar Cerrahisi Derg.* 2021;29(2):174-182. doi: 10.5606/tgkdc.dergisi.2021.21403.
11. Liu PH, Shih HH, Kang PL, Pan JY, Wu TH, Wu CJ. Performance of the EuroSCORE II model in predicting short-term mortality of general cardiac surgery. *Acta Cardiol Sin.* 2022;38(4):495-503. doi:10.6515/ACS.202207_38(4).20220128B.
12. Sembiring YE, Ginting A, Puruhito, Budiono. Validation of EuroSCORE II to predict mortality in post-cardiac surgery patients in East Java tertiary hospital. *Med J Indones.* 2021;30(1):54-59. doi:10.13181/mji.0a.204536.
13. Zahara R, Soeharto DF, Widyantoro B, Sugisman, Herlambang B. Validation of EuroSCORE II scoring system on isolated CABG patient in Indonesia. *Egypt Heart J.* 2023; 75:86. doi:10.1186/s43044-023-00410-0.
14. Kumiawaty J, Setianto BY, Widyastuti Y, Supomo S, Boom CE, Ancilla C. Validation for EuroSCORE II in the Indonesian cardiac surgical population: a retrospective, multicenter study. *Expert Rev Cardiovasc Ther.* 2022;20(6):491-496. doi: 10.1080/14779072.2022.2078703.
15. Shales S, Rao SUM, Khapli S, Ghorai PA, Behera SK, Ghosh AK, et al. Comparison of EuroSCORE II and STS score for risk prediction in Indian patients undergoing coronary artery bypass grafting. *Indian J Thorac Cardiovasc Surg.* 2021;37(6):623-630. doi:10.1007/s12055-021-01186-1.
16. Gao F, Shan L, Wang C, Meng X, Chen J, Han L, et al. Predictive Ability of European Heart Surgery Risk Assessment System II (EuroSCORE II) and the Society of Thoracic Surgeons (STS) Score for in-Hospital and Medium-Term Mortality of Patients Undergoing Coronary Artery Bypass Grafting. *Int J Gen Med.* 2021; 14:8509-8519. doi: 10.2147/IJGM.S338819.
17. Koo SK, Dignan R, Lo EYW, Williams C, Xuan W. Evidence-Based Determination of Cut-Off Points for Increased Cardiac-Surgery Mortality Risk with EuroSCORE II and STS: The Best-Performing Risk Scoring Models in a Single-Centre Australian Population. *Heart Lung Circ.* 2022;31(4):590-601. doi: 10.1016/j.hlc.2021.08.026.
18. Silverborn M, Nielsen S, Karlsson M. The performance of EuroSCORE II in CABG patients in relation to sex, age, and surgical risk: a nationwide study in 14,118 patients. *J Cardiothorac Surg.* 2023;18(1):40. doi: 10.1186/s13019-023-02141-4.
19. Fang SY, Chen JW, Chou HW, Chan CY, Wu IH, Chou NK, et al. Validation of the European system for cardiac operative risk evaluation II in a large Taiwan cardiac surgical centre. *J Formos Med Assoc.* 2023;122(12):1265-1273. doi: 10.1016/j.jfma.2023.05.028.
20. Wolf PJW, Amato VL. Assessment of EuroSCORE II and STS Score Performance and the Impact of Surgical Urgency in Isolated Coronary Artery Bypass Graft Surgery at a Referral Center in São Paulo, Brazil. *Braz J Cardiovasc Surg.* 2024;39(5):e20230282. doi: 10.21470/1678-9741-2023-0282.
21. Mastroiaco G, Bonomi A, Ludergrani M, Franchi M, Maragna R, Pirola S, et al. Is EuroSCORE II still a reliable predictor for cardiac surgery mortality in 2022? A retrospective study. *Eur J Cardiothorac Surg.* 2022;64(3): ezad294. doi: 10.1093/ejcts/ezad294.
22. Ng KHL, Wang K, Fujikawa T, Kwok MWT, Ho JYK, Chow SCY, et al. Validation of EuroSCORE II in post-cardiac surgery patients in a tertiary institution in Hong Kong. *Hong Kong Med J.* 2026;32(2):126-134. doi: 10.12809/hkmj2514125.
23. Boukhmis A, Himeur H, Haddad A, Guerchani MK, Djafer B. External validation of the EuroSCORE II and the STS-PROM for patients undergoing isolated coronary artery bypass grafting in a developing country. *Eur Heart J.* 2025;46(Suppl 1): ehaf784.1672. doi: 10.1093/eurheartj/ehaf784.1672.
24. Musa AF, Cheong XP, Dillon J, Nordin RB. Validation of EuroSCORE II in patients undergoing coronary artery bypass grafting (CABG) surgery at the National Heart Institute, Kuala Lumpur: a retrospective review. *F1000Res.* 2018 May 2; 7:534. doi: 10.12688/f1000research.14760.2.
25. Qadir I, Salick MM, Perveen S, Sharif H. Mortality from isolated coronary bypass surgery: a comparison of the Society of Thoracic Surgeons and the EuroSCORE risk prediction algorithms. *Interact Cardiovasc Thorac Surg.* 2012;14(3):258-262. doi:10.1093/icvts/ivr072.
26. Biancari F, Vasques F, Mikkola R, Martin M, Lahtinen J, Heikkinen J. Validation of EuroSCORE II in patients undergoing coronary artery bypass surgery. *Ann Thorac Surg.* 2012 Jun;93(6):1930-5. doi: 10.1016/j.athoracsur.2012.02.064.
27. Grant SW, Hickey GL, Dimarakis I, et al. How does EuroSCORE II perform in UK cardiac surgery: an analysis of 23,740 patients from the Society for Cardiothoracic Surgery in Great Britain and Ireland National Database. *Heart.* 2012. doi:10.1136/heartjnl-2012-302483.