

External validation and calibration of risk models in isolated coronary artery bypass grafting: a systematic review

Lusine A. Arutyunyan, Sergei V. Shalaev, Svetlana B. Tsiryatieva, Lyudmila V. Kremneva, Vladlena V. Ostanina, Rina T. Gaibova, Maksim V. Martynenko

Corresponding author: Lusine A. Arutyunyan,
dr.lusine.arutyunyan@mail.ru

Received 22 October 2025. Revised 20 March 2026.
Accepted 10 April 2026.

How to cite: Arutyunyan L.A., Shalaev S.V., Tsiryatieva S.B., Kremneva L.V., Ostanina V.V., Gaibova R.T., Martynenko M.V. External validation and calibration of risk models in isolated coronary artery bypass grafting: a systematic review. *Patologiya krovoobrashcheniya i kardiokhirurgiya = Circulation Pathology and Cardiac Surgery*. 2026;30(2): 55-66 (In Russ.)
<https://doi.org/10.21688/1681-3472-2026-2-55-66>

Funding

The study did not have sponsorship.

Conflict of interest

The authors declare no conflict of interest.

Contribution of the authors

Conception and study design: L.A. Arutyunyan
Data collection and analysis: L.A. Arutyunyan, S.V. Shalaev, S.B. Tsiryatieva
Statistical analysis: L.A. Arutyunyan, L.V. Kremneva, V.V. Ostanina
Drafting the article: L.A. Arutyunyan, R.T. Gaibova, M.V. Martynenko
Critical revision of the article: L.A. Arutyunyan, S.V. Shalaev, S.B. Tsiryatieva
Final approval of the version to be published: all authors

ORCID

L.A. Arutyunyan, <https://orcid.org/0009-0009-5350-8426>
S.V. Shalaev, <https://orcid.org/0000-0003-2724-4016>
S.B. Tsiryatieva, <https://orcid.org/0000-0002-3881-2851>
L.V. Kremneva, <https://orcid.org/0000-0001-6331-1218>
V.V. Ostanina, <https://orcid.org/0009-0007-6635-8047>
R.T. Gaibova, <https://orcid.org/0000-0002-2399-0866>
M.V. Martynenko, <https://orcid.org/0000-0003-4003-6377>

© 2026 Arutyunyan et al.



Tyumen Regional Clinical Hospital No. 1, Tyumen, Russian Federation

Abstract

Introduction: In-hospital mortality after isolated coronary artery bypass grafting is now low, yet it differs markedly between centers and countries. Preoperative risk estimates guide patient routing, informed consent, and quality benchmarking. Because surgical technique, perioperative care, and patient profiles have shifted since 2018, risk models built on earlier cohorts lose calibration and call for re-examination in current practice.

Objective: To assess contemporary external validity and calibration of EuroSCORE II and the Society of Thoracic Surgeons risk model (STS) for short-term mortality in adults undergoing isolated coronary artery bypass grafting.

Methods: We performed a systematic search of PubMed/MEDLINE, Scopus, and Web of Science restricted to 2018–2025 and English-language publications. Eligible records were original external validations in isolated coronary artery bypass grafting or studies with a clearly reported subgroup; endpoints were in-hospital mortality or 30-day survival. We extracted discrimination (area under the ROC curve) and calibration metrics (observed-to-expected mortality ratio, Hosmer – Lemeshow p-value, calibration slope and intercept when available). Study selection and data extraction were performed independently by two reviewers; unreported metrics were recorded as “not reported”. Risk of bias was assessed using Prediction Model Risk of Bias Assessment Tool (PROBAST).

Results: Nine studies from Europe, Asia, and Latin America were included (sample sizes 438–14,118). Discrimination ranged from modest to high: for EuroSCORE II AUC 0.55–0.90 (0.82–0.90 in the largest cohorts); for the STS model AUC 0.65–0.88. Calibration was heterogeneous: overestimation in low-risk and younger patients (e.g., overall observed-to-expected ratio 0.58; 0.26 in patients <60 years) contrasted with underestimation in older adults, in procedures with cardiopulmonary bypass, and in several non-European populations. In some reports, simple centre-specific recalibration improved agreement between predicted and observed outcomes.

Conclusion: In contemporary isolated coronary artery bypass grafting, EuroSCORE II and the STS model remain useful for risk ranking but centre-level calibration checks and – when indicated – straightforward recalibration are recommended. Regular auditing of observed-to-expected mortality and stratified verification in key subgroups is also recommended.

Keywords: calibration; cardiac surgery; coronary artery bypass; external validation; hospital mortality; prognostic models; risk assessment

Внешняя валидация и калибровка моделей риска при изолированном коронарном шунтировании: систематический обзор

Л.А. Арутюнян, С.В. Шалаев, С.Б. Цирятеева, Л.В. Кремнева, В.В. Останина, Р.Т. Гаибова, М.В. Мартыненко

Для корреспонденции: Лусине Амазасповна Арутюнян, dr.lusine.arutyunyan@mail.ru

Государственное бюджетное учреждение здравоохранения Тюменской области «Областная клиническая больница № 1», Тюмень, Российская Федерация

Поступила в редакцию 22 октября 2025 г.
Исправлена 20 марта 2026 г. Принята к печати
10 апреля 2026 г.

Цитировать: Арутюнян Л.А., Шалаев С.В., Цирятьева С.Б., Кремнева Л.В., Останина В.В., Гаибова Р.Т., Мартыненко М.В. Внешняя валидация и калибровка моделей риска при изолированном коронарном шунтировании: систематический обзор. *Патология кровообращения и кардиохирургия*. 2026;30(2):55–66. <https://doi.org/10.21688/1681-3472-2026-2-55-66>

Финансирование

Исследование не имело спонсорской поддержки.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Вклад авторов

Концепция и дизайн работы: Л.А. Арутюнян

Сбор и анализ данных: Л.А. Арутюнян, С.В. Шалаев, С.Б. Цирятьева

Статистическая обработка данных: Л.А. Арутюнян, Л.В. Кремнева, В.В. Останина

Написание статьи: Л.А. Арутюнян, Р.Т. Гаибова, М.В. Мартыненко

Исправление статьи: Л.А. Арутюнян, С.В. Шалаев, С.Б. Цирятьева

Утверждение окончательного варианта статьи: все авторы

ORCID

Л.А. Арутюнян, <https://orcid.org/0009-0009-5350-8426>
С.В. Шалаев, <https://orcid.org/0000-0003-2724-4016>
С.Б. Цирятьева, <https://orcid.org/0000-0002-3881-2851>
Л.В. Кремнева, <https://orcid.org/0000-0001-6331-1218>
В.В. Останина, <https://orcid.org/0009-0007-6635-8047>
Р.Т. Гаибова, <https://orcid.org/0000-0002-2399-0866>
М.В. Мартыненко, <https://orcid.org/0000-0003-4003-6377>

© Арутюнян Л.А., Шалаев С.В., Цирятьева С.Б., Кремнева Л.В., Останина В.В., Гаибова Р.Т., Мартыненко М.В., 2026

Аннотация

Актуальность. Госпитальная летальность при изолированном коронарном шунтировании сегодня низкая, но заметно различается между центрами и странами. Предоперационная оценка риска определяет маршрутизацию пациента, информированное согласие и сопоставление качества помощи. Поскольку с 2018 г. изменились хирургическая техника, периоперационное ведение и профиль пациентов, модели риска, построенные на когортах прежних лет, теряют калибровку и нуждаются в повторной проверке в современных условиях.

Цель. Оценить внешнюю валидность и калибровку моделей EuroSCORE II и модели Общества торакальных хирургов (STS Score) для краткосрочной летальности у взрослых пациентов после изолированного коронарного шунтирования в современных условиях.

Методы. Проведен систематический поиск в PubMed/MEDLINE, Scopus и Web of Science с ограничениями по годам (2018–2025) и языку (английский). Включались оригинальные исследования внешней валидации у пациентов с изолированным коронарным шунтированием либо работы с четко выделяемой подгруппой; конечные точки – госпитальная летальность или 30-дневная выживаемость. Извлекались показатели различающей способности (AUC) и калибровки (отношение «наблюдаемая/ожидаемая» летальность, р-значение теста Хосмера – Лемешоу, наклон и интерсепт калибровки при наличии). Скрининг и извлечение данных выполнялись независимо двумя рецензентами; при отсутствии значений в первоисточниках они помечались как «нет данных». Риск систематической ошибки оценен с помощью Prediction Model Risk of Bias Assessment Tool (PROBAST).

Результаты. В обзор включено девять исследований из Европы, Азии и Латинской Америки (объем выборок – 438–14 118). Различающая способность варьировала от умеренной до высокой: для EuroSCORE II – AUC 0,55–0,90 (в крупнейших когортах – 0,82–0,90), для STS Score – AUC 0,65–0,88. Калибровка оказалась неоднородной: выявлена переоценка риска у молодых пациентов и в низких категориях риска (например, общее отношение «наблюдаемая/ожидаемая» летальность (О/Е) 0,58, у пациентов <60 лет О/Е 0,26) и недооценка у пожилых пациентов, при операциях на искусственном кровообращении и в ряде неевропейских популяций (до О/Е 4,2). В отдельных исследованиях локальная корректировка калибровки улучшала согласие прогноза с наблюдаемыми исходами.

Заключение. В современных сериях изолированного коронарного шунтирования EuroSCORE II и STS Score остаются полезными для ранжирования риска, однако рекомендуется проверка калибровки на уровне центра и при необходимости простая перекалибровка. Также рекомендуется регулярный аудит отношения «наблюдаемая/ожидаемая» летальность и стратифицированная верификация в ключевых подгруппах.

Ключевые слова: внешняя валидация; госпитальная летальность; калибровка; кардиохирургия; коронарное шунтирование; оценка риска; прогностические модели



Introduction

Mortality after coronary artery bypass grafting (CABG) currently remains low in absolute terms, but it is heterogeneous across different centers and populations [1]. Therefore, preoperative risk stratification is a key element of patient routing, informed consent, and quality benchmarking in cardiac surgery [2]. The most widely

used risk models – EuroSCORE II [3] and the STS (Society of Thoracic Surgeons) risk model [4] – were developed on large but diverse cohorts and, over time, inevitably suffer from calibration drift (i.e., deterioration in agreement between predicted and observed risk) as surgical techniques, indications, and patient profiles evolve. EuroSCORE II is an update of the classic EuroSCORE [5];

the model was developed on a contemporary European cohort and initially demonstrated better calibration while maintaining high discriminative ability compared with its predecessor. In parallel, the Society of Thoracic Surgeons (STS) published a new generation of risk models based on data from the large ACSD (Adult Cardiac Surgery Database), providing high predictive accuracy for cardiac operations, including CABG [6], along with detailed methodological principles for model development and validation [7].

However, even well constructed models, when transferred “to other walls,” require reassessment [8]. Classic literature on prognostic modeling emphasizes that calibration is the Achilles’ heel of predictive analytics: a poorly calibrated algorithm can mislead clinicians and cause harm in decision making. Hence, external validation requires sufficient data volumes and explicit assessment of calibration metrics [9].

Accumulated evidence on the external validity of EuroSCORE II / STS in different countries confirms heterogeneity of results. In Iranian and Tunisian cohorts, EuroSCORE II showed acceptable discrimination but required calibration tuning, which is explained by differences in risk profiles and complication patterns [10; 11]. In Greece and Serbia (studies up to 2018), similar conclusions were drawn: EuroSCORE II outperformed the old EuroSCORE version in discrimination but was not always perfectly calibrated in local settings [12; 13].

Comparative reviews and meta analyses report that the accuracy of the two aforementioned risk models is generally statistically comparable, and differences depend on the context (type of surgery, urgency, country, study period) [14]. Additional data from cardiac surgery in China (not limited to CABG) indicate good predictive ability of EuroSCORE II for short term outcomes, with in hospital mortality often being more informative than 30 day mortality for outcome assessment [15]. A particular challenge is represented by clinical subgroups, e.g., patients with atrial fibrillation: large registries have shown that in such patients EuroSCORE II may underestimate risk in low and intermediate risk ranges, and discrimination for CABG is worse than in patients without atrial fibrillation [16].

After 2017–2018, the approach to planning and performing CABG has substantially changed [17]. European guidelines on myocardial revascularization have strengthened recommendations for arterial conduit strategies (multiple arterial grafting, priority

of the left anterior descending artery, etc.), which has altered the procedure profile in contemporary patients [18]. Furthermore, the widespread adoption of enhanced recovery after surgery (ERAS) protocols for perioperative care, as well as the standardization of perioperative cardioprotection and patient management algorithms, have contributed to changes in the principles of perioperative CABG management [19]. The “architecture” of CABG surgery is also changing and becoming more differentiated: nuanced decisions regarding technique selection in the context of off pump CABG (OPCAB) and patient selection are being practiced; contemporary reviews emphasize the need for selective use of OPCAB and a re evaluation of its role in frail patients [20; 21].

In parallel, standards for reporting and critical appraisal of prognostic models have evolved. Whereas in the past authors often relied solely on the area under the ROC curve (AUC), recent years have seen the “gold standard” become reports that must include calibration assessment: the observed to expected (O/E) ratio, the Hosmer–Lemeshow (HL) test, calibration slope, and calibration in the large (CITL). This is explicitly stated in the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) statement, its extension for machine learning models (TRIPOD+AI), and methodological reviews on calibration [22]. The picture is further complemented by the updated Prediction Model Risk of Bias Assessment Tool (PROBAST) and its AI extension (PROBAST+AI), which provide a unified framework for external validation and assessment of model applicability in local contexts [23; 24].

Thus, validations from the previous era of CABG (roughly prior to 2018) reflect a different clinical and organizational landscape (different anesthesia and rehabilitation protocols, different patient mix and surgical technique), and they naturally suffer from calibration drift, yielding unstable risk estimates when extrapolated to current practice. Consequently, direct comparison with contemporary series is inappropriate; for a clinically relevant answer, the search should be limited to more recent studies that align with current standards of care and reporting [19].

The aim of this systematic review is to summarize evidence from recent studies (later than 2018) on the external validation of EuroSCORE II / STS in adult patients undergoing isolated CABG, with a focus on AUC ranges as a measure of discrimination, calibration (O/E, Hosmer–

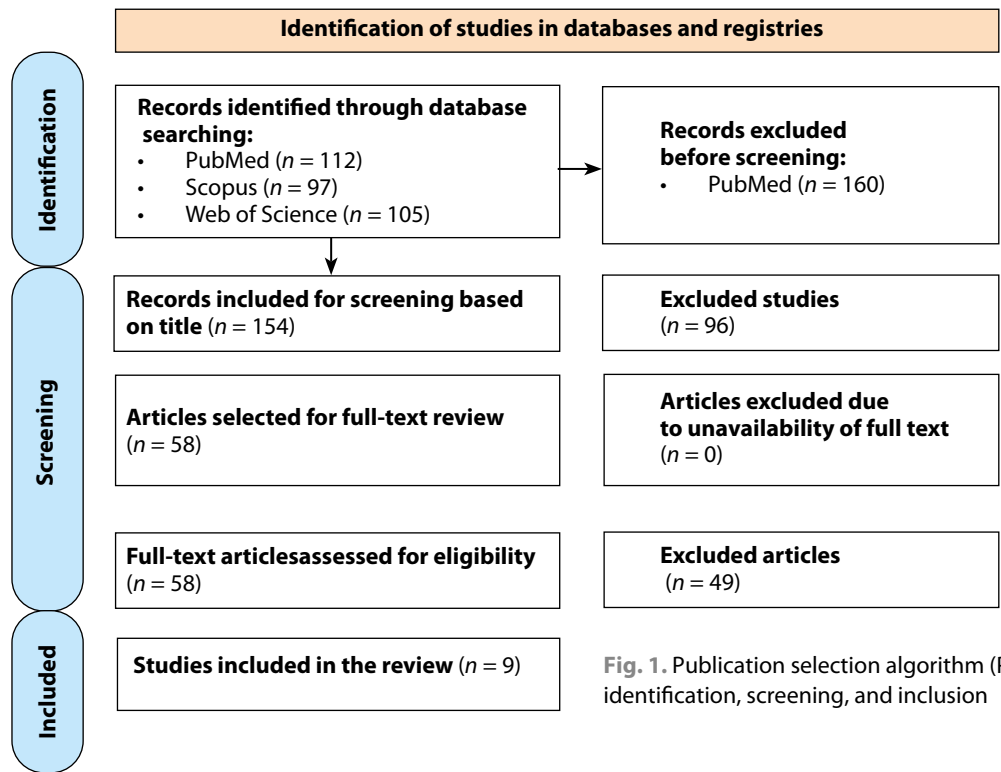


Fig. 1. Publication selection algorithm (PRISMA-2020): identification, screening, and inclusion

Lemeshow test, slope/CITL), model performance in key subgroups (age, urgency, OPCAB), and implications for everyday clinical practice (when local recalibration or selection of an alternative model is needed).

Methods

Study Design

Systematic review of external validations of the EuroSCORE II and STS risk models in adult patients undergoing isolated coronary artery bypass grafting. Endpoints were in hospital mortality or 30 day survival. Reporting followed the PRISMA 2020 statement [25]. The review protocol was not prospectively registered.

Information Sources and Search

The search was performed in PubMed/MEDLINE, Scopus, and Web of Science (last update: October 6, 2025) using combinations of terms describing the models (EuroSCORE II, STS score), the intervention (coronary artery bypass, CABG), and model assessment (validation, calibration, discrimination, performance). Additionally, reference lists of relevant articles were screened. The

full search strategy for each database is provided in the supplementary material (Appendix).

Inclusion Criteria

Original studies reporting external validation of the aforementioned models in adult patients undergoing isolated CABG; publications from 2018 to 2025; English language; availability of the full text and at least one model performance metric (AUC and/or calibration measures: O/E, HL, slope/CITL).

Selection and Data Extraction

Selection was performed in two stages (title/abstract; full text) with documentation of reasons for exclusion. For the included studies, we extracted the design, population and sample size, outcome definition, AUC values [95% CI], O/E, Hosmer–Lemeshow test p value, calibration slope, and calibration in the large (CITL); where values were not reported, they were marked as “not available” (n/a).

Quality Assessment and Synthesis

Quality and applicability of the studies were assessed using the Prediction Model Risk of Bias Assessment Tool (PROBAST) [24] across four domains: participants,

Table 1. Studies included in the review

Authors, year, country	Sample size (n)	Outcome	AUC [95% CI] (ESII)	AUC [95% CI] (STS)	O/E (ESII)	O/E (STS)	HL p (ESII)	HL p (STS)	Slope/CITL (ESII)
Silverborn et al., 2023, Sweden [26]	14,118	30 day survival	0.82 [0.79–0.85]	n/a	0.58	n/a	n/a	n/a	n/a
Zahara et al., 2023, Indonesia [27]	2,628	In hospital mortality	0.72 [0.67–0.77]	n/a	n/a	n/a	0.02	n/a	0.814 / 0.009
Mokhtar et al., 2023, Malaysia [28]	1,059	30 day survival	0.844 [0.705–0.983]	n/a	n/a	n/a	0.768	n/a	n/a
Wolf et al., 2024, Brazil [29]	438	In hospital mortality	0.697 [0.593–0.802]	0.646 [0.532–0.760]	1.59	3.60	n/a	n/a	n/a
Hirnle et al., 2024, Poland [30]	5,771	In hospital mortality	0.72 [0.68–0.77]	n/a	1.62 (on pump) / 0.98 (off pump)	n/a	<0.001 (on pump) / 0.34 (off pump)	n/a	n/a
Musa et al., 2019, Malaysia [31]	1,718	In hospital mortality	0.70 [0.640–0.759]	n/a	n/a	n/a	0.089 (overall cohort) / 0.032 (isolated CABG)	n/a	n/a
Hogervorst et al., 2018, Netherlands [32]	2,296 (overall) / 1,059 (isolated CABG)	In hospital mortality	0.884 (isolated CABG)	n/a	0.86 (isolated CABG)	n/a	n/a	n/a	n/a
Gao et al., 2021, China [33]	1,628	In hospital mortality	0.900	0.879	n/a	n/a	0.071	0.104	n/a
Rodríguez Lima et al., 2025, Colombia [34]	1,009	In hospital mortality	0.55 [0.48–0.63]	n/a	4.2	n/a	n/a	n/a	n/a

Note: CI – confidence interval; isolated CABG – isolated coronary artery bypass grafting; n/a – not available; AUC – area under the ROC curve; ESII – EuroSCORE II; HL p – Hosmer–Lemeshow test p value; O/E – observed to expected mortality ratio; off pump – off pump CABG; on pump – on pump CABG; slope/CITL – calibration slope / calibration in the large; STS – Society of Thoracic Surgeons.

predictors, outcome, analysis. Summary domain level ratings are reported below. Synthesis was narrative, summarizing discrimination ranges and calibration direction; quantitative meta analysis was not planned due to clinical and methodological heterogeneity.

Results

Selection of Publications

A total of 314 records were identified (PubMed – 112; Scopus – 97; Web of Science – 105). After duplicate removal, 154 records were screened for eligibility. Based on title and abstract, irrelevant publications were excluded; 58 articles underwent full text assessment. The main reasons for exclusion at this stage were: absence

of a separate isolated CABG subgroup, lack of external validation, and non English texts. Finally, 9 studies meeting the criteria were included in the qualitative synthesis. The search and study inclusion algorithm is shown in Fig. 1.

Description of Included Studies

The studies included in the review are presented in Table 1. Nine studies from Europe, Asia, and Latin America include sample sizes ranging from 438 to 14,118 patients. A retrospective design (registries / single center series) predominates; populations consist of isolated CABG. The outcome for validation is in hospital or 30 day mortality. Several studies present subgroups by age, urgency, and technique (on pump / off pump).

Table 2. Risk of bias and applicability assessment (PROBAST)

Authors, year, country	Participants (risk of bias / applicability)	Predictors (risk of bias / applicability)	Outcome (risk of bias / applicability)	Analysis (risk of bias / applicability)	Overall risk of bias	Overall applicability	Brief rationale
Silverborn et al., 2023, Sweden [26]	Low / High	Low / High	Low / High	Low / High	Low	High	National registry of isolated CABG; correct ESII calculation; 30 day outcome; sufficient number of events; AUC and O/E reported
Zahara et al., 2023, Indonesia [27]	Low / High	Low / High	Low / High	Moderate / High	Moderate	High	Single center study; isolated CABG; AUC, HL, slope/CITL reported; limited details on statistical procedures (no O/E by strata)
Mokhtar et al., 2023, Malaysia [28]	Low / High	Low / High	Low / High	Low / High	Low	High	Isolated CABG; AUC with CI and HL p reported; sufficient number of observations
Wolf et al., 2024, Brazil [29]	Low / High	Low / High	Low / High	Moderate / High	Moderate	High	Two centers; isolated CABG; AUC reported; O/E and p provided; relatively few events, thus wide CIs
Hirnle et al., 2024, Poland [30]	Low / Limited	Low / High	Low / High	Low / High	Low	Limited	National registry; subgroup aged ≥80 years (limitation of applicability to the general population); on /off pump reported separately
Musa et al., 2019, Malaysia [31]	Low / High	Low / High	Low / High	Moderate / High	Moderate	High	CABG with an identifiable isolated CABG subgroup; HL significant in isolated CABG; AUC and methods partially limited
Hogervorst et al., 2018, Netherlands [32]	Low / High	Low / High	Low / High	Low / High	Low	High	Single center study; isolated CABG subgroup with AUC and O/E reported; appropriate methodology
Gao et al., 2021, China [33]	Low / High	Low / High	Low / High	Moderate / High	Moderate	High	Isolated CABG; AUC and HL p for ESII/STS reported; adequate sample size and reporting
Rodríguez Lima et al., 2025, Colombia [34]	Low / High	Low / High	Low / High	Low / High	Low	High	Isolated CABG; low AUC and marked O/E shift; acceptable statistics, but CIs wide

Note: CI – confidence interval; isolated CABG – isolated coronary artery bypass grafting; CABG – coronary artery bypass grafting; n/a – not available; AUC – area under the ROC curve; ESII – EuroSCORE II; HL p – Hosmer–Lemeshow test p value; O/E – observed to expected mortality ratio; off pump – off pump CABG; on pump – on pump CABG; slope/CITL – calibration slope / calibration in the large; STS – Society of Thoracic Surgeons.

Risk of Bias Assessment

The results of the PROBAST assessment are presented in Table 2. Of the nine included studies, five had a low overall risk of bias, and four had a moderate risk (mainly due to the “Analysis” domain: small number of events or incomplete reporting of calibration metrics). Applicability was high in all studies, except for the study by Hirnle et al. (2024), in which the sample was restricted to patients aged ≥ 80 years [30].

Model Performance Metrics

Discriminative ability (AUC). For EuroSCORE II, the range was 0.55–0.90: high AUC was observed in the Swedish national registry (0.82) [26] and the Chinese series (0.90) [33]; moderate values were seen in Southeast Asia (0.72–0.84) [27; 31]; a low value was found in the Colombian cohort (0.55) [34]. For STS, reported values ranged from 0.646 to 0.879 (two studies reporting AUC for STS), with both models exceeding 0.80 in stable patients [29; 33].

Calibration. The direction of miscalibration was heterogeneous. In the Swedish cohort, overestimation of risk was observed in low risk and younger patients (overall O/E 0.58; O/E 0.26 in patients < 60 years), with improved calibration in older ages and higher risk categories [26]. The Indonesian study showed a significant Hosmer–Lemeshow test ($p = 0.02$) and a slope of 0.814 with a small intercept shift (CITL 0.009) – indicating underestimation of absolute risk [27]. In the Brazilian prospective series, EuroSCORE II showed calibration that did not formally reach statistical significance at the 0.05 threshold (O/E 1.59; $p = 0.056$); however, the direction of miscalibration indicated a trend toward risk underestimation, whereas STS showed marked miscalibration (O/E 3.60; $p < 0.001$) [29]. In the Polish registry of patients ≥ 80 years, calibration differed by technique: on pump – O/E 1.62, HL $p < 0.001$; off pump – O/E 0.98, HL $p = 0.34$ [30]. The Colombian cohort demonstrated marked underestimation (O/E 4.2) alongside low AUC [34].

Discussion

In contemporary cohorts of patients undergoing isolated coronary artery bypass grafting, the EuroSCORE II and STS models maintain satisfactory discriminative ability (generally AUC ~ 0.70 – 0.85), but they demonstrate unstable calibration: the direction and magnitude of miscalibration depend on age, risk profile, surgical technique, and geographic region, indicating the need for local validation and recalibration.

The data presented in Table 1 show a characteristic pattern: in large national registries, discrimination remains high (AUC 0.82–0.90), whereas values are more modest in medium sized cohorts. Calibration is even more heterogeneous: overestimation of risk in younger and low risk patients contrasts with underestimation in the elderly during on pump surgery and in non European populations.

These findings are consistent with recent publications. Retrospective and prospective series from Europe and Asia confirm that the discrimination of EuroSCORE II and STS is generally comparable, whereas calibration differs and is sensitive to the population, outcomes, and study period. For example, a large single center study from 2023 demonstrated reduced reliability of EuroSCORE II in high risk patients. The authors concluded that this model does not reflect the actual mortality risk in a high risk cohort [35]. A similar trend is observed in valve surgery: modest AUC of EuroSCORE II and calibration drift in high risk groups, directly indicating the need for updating/recalibration before application in the context of new technologies and patient profiles [36]. Reports from the STS ACS in 2020 [7] and 2022 [37] demonstrate continued reduction in baseline CABG mortality in the United States due to systemic improvements, which inevitably affects the calibration of historical models.

Comparative studies of STS and EuroSCORE II in non US centers show heterogeneous results: some studies find a slight advantage for STS in AUC, others for EuroSCORE II; nevertheless, a general trend is observed – STS is more often miscalibrated outside the United States [38].

The possible mechanisms underlying differences in model accuracy essentially boil down to several interrelated shifts in the clinical context. Since 2018, the “architecture” of the operation and perioperative management have changed: widespread implementation of ERACS/fast track programs, more frequent use of off pump surgery in frail and elderly patients, a focus on arterial conduit strategies – all these factors reduce observed mortality compared with the era in which the models were trained, thereby shifting their calibration [19]. These trends are enshrined in current European revascularization guidelines and are supported by registry reports documenting further improvements in outcomes for standard procedures [18].

Age and urgency of intervention notably influence the calibration profile. The aggregated data reveal a clear age gradient: risk is often overestimated in younger patients,

whereas underestimation is observed in older patients, especially during on pump surgery; switching to off pump in some elderly patients mitigates this bias. This is consistent with a meta analysis comparing on pump and off pump CABG in elderly patients [39].

Another systemic factor is frailty. This phenotype is not captured by either EuroSCORE II or STS, yet it is strongly associated with complication rates, including 30 day mortality after CABG. Recent reviews and meta analyses report increased relative risks of 2–5 fold [40; 41]. Even with acceptable AUC, failure to account for frailty formally worsens calibration in geriatric subgroups.

From a methodological perspective, the key point remains unchanged: calibration is the Achilles' heel of prognostic models [9; 42]. External validation should include calibration curves and at least four metrics – O/E, Hosmer–Lemeshow test p value, slope, and calibration in the large (CITL) [42]. When miscalibration is present, local recalibration is necessary: the intercept is adjusted, and, in the case of probability “shrinkage,” the slope is also adjusted; these are simple and reproducible procedures that restore agreement with the data from a specific center without loss of discriminative ability [42].

Limitations

This review is based on a small number of studies, which is itself due to the deliberately limited time frame (2018–2025) and strict inclusion criteria. Another substantial limitation is the incompleteness of reporting in the selected studies:

1. Calibration metrics are not presented in all articles: several studies lack one or more key calibration measures (O/E, HL, slope/CITL), which limits comparability and precludes a pooled assessment of calibration across studies.
2. Incomplete confidence intervals for AUC: some publications report only the point estimate of AUC without 95% CIs, reducing the precision of comparisons of discrimination.
3. Inconsistent specification of the outcome: some studies use in hospital mortality, others use 30 day survival; in the absence of unambiguous convertibility, this hinders unified aggregation of metrics.
4. Limited subgroup reporting: not all studies provide stratification (age, urgency, on/off pump), making

it impossible to perform a comparable subgroup synthesis of calibration/discrimination.

5. Incomplete presentation of STS scores: in several studies, the STS model was either not validated or its metrics (AUC, calibration) were not reported, so direct comparison between EuroSCORE II and STS is not possible for all cohorts.
6. Inconsistencies in reporting statistical tests: some articles report p values for goodness of fit tests that are not the Hosmer–Lemeshow test, without explicitly specifying the method.
7. Infeasibility of meta analysis: the aggregation of missing calibration data and heterogeneity of outcomes preclude a valid quantitative synthesis; a narrative synthesis was therefore used.

These limitations are typical of risk validation studies in cardiac surgery and largely reflect the general lack of transparent reporting [9; 42].

Implications for clinical practice. EuroSCORE II and STS should be regarded primarily as tools for quality auditing and benchmarking between centers, rather than as precise calculators of individual risk. Every cardiac surgery center is advised to perform regular audits of O/E and stratified metrics. It is recommended that frailty be included as an additional predictor in future versions of the models.

Priority directions for future research: 1) prospective multicenter validations with a full set of calibration metrics and TRIPOD compliant reporting; 2) development of simple algorithms for local recalibration suitable for routine use; 3) evaluation of the prognostic value of frailty indicators as supplements to existing scoring systems.

Conclusion

Against the background of gradually declining mortality in patients after isolated coronary artery bypass grafting, the issue of validating and calibrating prognostic models remains highly relevant. In the contemporary era of CABG, the EuroSCORE II and STS models still retain utility for risk stratification, provided that calibration is checked. Miscalibration is most pronounced in patient subgroups (demographics, comorbidities, surgical technique) – here, stratified approaches are particularly important. Regular validation (O/E, HL, slope/CITL), transparent reporting, and simple adjustments allow these scores to be used safely as one of the tools supporting clinical decision making.

References

1. Semchenko A.N., Zaitsev I.V., Shevchenko A.M., Semchenko A.V. Influence of incomplete revascularization on the outcomes of coronary artery bypass grafting in patients with ischemic heart disease. *Patologiya krovoobrashcheniya i kardiokhirurgiya = Circulation Pathology and Cardiac Surgery*. 2021;25(1):74-84. (In Russ.) <http://dx.doi.org/10.21688/1681-3472-2021-1-74-84>
2. Kamenshchikov N.O., Podoksenov Yu.K., Diakova M.L., Boyko A.M., Kozlov B.N. Acute kidney injury in cardiac surgery: preoperative predictive diagnostics. *Patologiya krovoobrashcheniya i kardiokhirurgiya = Circulation Pathology and Cardiac Surgery*. 2021;25(1):40-51. (In Russ.) <http://dx.doi.org/10.21688/1681-3472-2021-1-40-51>
3. Nashef S.A., Roques F., Sharples L.D., Nilsson J., Smith C, Goldstone A.R., Lockowandt U. EuroSCORE II. *Eur J Cardiothorac Surg*. 2012;41(4):734-44. PMID: 22378855. <https://doi.org/10.1093/ejcts/ezs043>
4. O'Brien S.M., Feng L., He X., Xian Y., Jacobs JP, Badhwar V., Kurlansky P.A., Furnary A.P., Cleveland J.C. Jr, Lobdell K.W., Vassileva C., Wyler Von Ballmoos M.C., Thourani V.H., Rankin J.S., Edgerton J.R., D'Agostino R.S., Desai N.D., Edwards F.H., Shahian D.M. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 2-Statistical Methods and Results. *Ann Thorac Surg*. 2018;105(5):1419-1428. PMID: 29577924. <https://doi.org/10.1016/j.athoracsur.2018.03.003>
5. Nashef S.A., Roques F., Michel P., Gauducheau E., Lemeshow S., Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg*. 1999;16(1):9-13. PMID: 10456395. [https://doi.org/10.1016/S1010-7940\(99\)00134-7](https://doi.org/10.1016/S1010-7940(99)00134-7)
6. Shahian D.M., O'Brien S.M., Filardo G., Ferraris V.A., Haan C.K., Rich J.B., Normand S.L., DeLong E.R., Shewan C.M., Dokholyan R.S., Peterson E.D., Edwards F.H., Anderson R.P.; Society of Thoracic Surgeons Quality Measurement Task Force. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 1 – coronary artery bypass grafting surgery. *Ann Thorac Surg*. 2009;88(1 Suppl):S2-22. PMID: 19559822. <https://doi.org/10.1016/j.athoracsur.2009.05.053>
7. Bowdish M.E., D'Agostino R.S., Thourani V.H., Desai N., Shahian D.M., Fernandez F.G., Badhwar V. The Society of Thoracic Surgeons Adult Cardiac Surgery Database: 2020 Update on Outcomes and Research. *Ann Thorac Surg*. 2020;109(6):1646-1655. PMID: 32247780. <https://doi.org/10.1016/j.athoracsur.2020.03.003>
8. Efremov S.M., Shmidt R.E., Zyryanova A.V., Shmidt T.A., Kamenskikh M.S., Shmatov D.V. The relationship between diastolic function characteristics and severe myocardial injury in coronary artery bypass graft surgery: a retrospective cohort study. *Patologiya krovoobrashcheniya i kardiokhirurgiya = Circulation Pathology and Cardiac Surgery*. 2024;28(2):51-61. (In Russ.) <https://doi.org/10.21688/1681-3472-2024-2-51-61>
9. Van Calster B., McLernon D.J., van Smeden M., Wynants L., Steyerberg E.W.; Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17(1):230. PMID: 31842878; PMCID: PMC6912996. <https://doi.org/10.1186/s12916-019-1466-7>
10. El Hadj Sidi C., Mgarrech I., Tarmiz A., Jerbi S. External validation of the European System for Cardiac Operative Risk Evaluation II in a Tunisian population. *J Card Surg*. 2019;34(5):266-273. PMID: 30873659. <https://doi.org/10.1111/jocs.14015>
11. Atashi A., Amini S., Tashnizi M.A., Moeinipour A.A., Aazami M.H., Tohidnezhad F., Ghasemi E., Eslami S. External Validation of European System for Cardiac Operative Risk Evaluation II (EuroSCORE II) for Risk Prioritization in an Iranian Population. *Braz J Cardiovasc Surg*. 2018;33(1):40-46. PMID: 29617500; PMCID: PMC5873780. <https://doi.org/10.21470/1678-9741-2017-0030>
12. Stavridis G., Panaretos D., Kadda O., Panagiotakos D.B. Validation of the EuroSCORE II in a Greek Cardiac Surgical Population: A Prospective Study. *Open Cardiovasc Med J*. 2017;11:94-101. PMID: 29204219; PMCID: PMC5688389. <https://doi.org/10.2174/1874192401711010094>
13. Nežić D., Spasić T., Micović S., Kosević D., Petrović I., Lausević Vuk L., Unić-Stojanović D., Borzanović M. Consecutive Observational Study to Validate EuroSCORE II Performances on a Single-Center, Contemporary Cardiac Surgical Cohort. *J Cardiothorac Vasc Anesth*. 2016;30(2):345-51. PMID: 26873386. <https://doi.org/10.1053/j.jvca.2015.11.011>
14. Sullivan P.G., Wallach J.D., Ioannidis J.P. Meta-Analysis Comparing Established Risk Prediction Models (EuroSCORE II, STS Score, and ACEF Score) for Perioperative Mortality During Cardiac Surgery. *Am J Cardiol*. 2016;118(10):1574-1582. PMID: 27687052. <https://doi.org/10.1016/j.amjcard.2016.08.024>
15. Liu P.H., Shih H.H., Kang P.L., Pan J.Y., Wu T.H., Wu C.J. Performance of the EuroSCORE II Model in Predicting Short-Term Mortality of General Cardiac Surgery: A Single-Center Study in Taiwan. *Acta Cardiol Sin*. 2022;38(4):495-503. PMID: 35873132; PMCID: PMC9295035. [https://doi.org/10.6515/ACS.202207_38\(4\).20220128B](https://doi.org/10.6515/ACS.202207_38(4).20220128B)
16. Kuźma Ł., Kowalewski M., Wańha W., Dąbrowski E.J., Jasiński M., Widenka K., Deja M., Bartuś K., Hirnle T., Wojakowski W., Lorusso R., Tobota Z., Maruszewski B.J., Suwalski P., KROK Investigators. Validation of EuroSCORE II in atrial fibrillation heart surgery patients from the KROK Registry. *Sci Rep*. 2023;13(1):13024. PMID: 37563207; PMCID: PMC10415263. <https://doi.org/10.1038/s41598-023-39983-w>
17. Kravchenko I.V., Podoksenov Yu.K., Tyo M.A., Kalashnikova T.P., Churilina E.A., Kozulin M.S., Gusakova A.M., Kozlov B.N., Kamenshchikov N.O. Prevention of respiratory complications in patients at risk of adverse respiratory events by perioperative nitric oxide delivery during cardiac surgery under cardiopulmonary bypass: a single-center prospective randomized study. *Patologiya krovoobrashcheniya i kardiokhirurgiya = Circulation Pathology and Cardiac Surgery*. 2024;28(3):78-93. (In Russ.) <https://doi.org/10.21688/1681-3472-2024-3-78-93>
18. Neumann F.J., Sousa-Uva M., Ahlsson A., Alfonso F., Banning A.P., Benedetto U., Byrne R.A., Collet J.P., Falk V., Head S.J., Jüni P., Kastrati A., Koller A., Kristensen S.D., Niebauer J., Richter D.J., Seferovic P.M., Sibbing D., Stefanini G.G., Windecker S., Yadav R., Zembala M.O., ESC Scientific Document Group. 2018 ESC/EACTS Guidelines on myocardial revascularization. *Eur Heart J*. 2019;40(2):87-165. PMID: 30165437. <https://doi.org/10.1093/eurheartj/ehy507>

19. Grant M.C., Crisafi C., Alvarez A., Arora R.C., Brindle M.E., Chatterjee S., Ender J., Fletcher N., Gregory A.J., Gunaydin S., Jahangiri M., Ljungqvist O., Lobdell K.W., Morton V., Reddy V.S., Salenger R., Sander M., Zarbock A., Engelman D.T. Perioperative Care in Cardiac Surgery: A Joint Consensus Statement by the Enhanced Recovery After Surgery (ERAS) Cardiac Society, ERAS International Society, and The Society of Thoracic Surgeons (STS). *Ann Thorac Surg.* 2024;117(4):669-689. PMID: 38284956. <https://doi.org/10.1016/j.athoracsur.2023.12.006>
20. Chan J., Dimagli A., Dong T., Fudulu D.P., Sinha S., Angelini G.D. Trend and early clinical outcomes of off-pump coronary artery bypass grafting in the UK. *Eur J Cardiothorac Surg.* 2023;64(2):ezad272. PMID: 37522886; PMCID: PMC10876163. <https://doi.org/10.1093/ejcts/ezad272>
21. Akchurin R.S., Shiryayev A.A., Galyautdinov D.M., Vasiliev V.P., Kurbanov S.K., Andreev A.V., Zaikovskiy V.Yu., Mayorov G.B. 1-year outcomes of coronary artery bypass grafting in patients with target artery distal calcinosis. *Patologiya krovoobrashcheniya i kardiokhirurgiya = Circulation Pathology and Cardiac Surgery.* 2022;26(1):55-65. (In Russ.) <http://dx.doi.org/10.21688/1681-3472-2022-1-55-65>
22. Collins G.S., Reitsma J.B., Altman D.G., Moons K.G. Transparent reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. *J Clin Epidemiol.* 2015;68(2):134-43. PMID: 25579640. <https://doi.org/10.1016/j.jclinepi.2014.11.010>
23. Moons K.G.M., Damen J.A.A., Kaul T., Hooft L., Andaur Navarro C., Dhiman P., Beam A.L., Van Calster B., Celi L.A., Denaxas S., Denniston A.K., Ghassemi M., Heinze G., Kengne A.P., Maier-Hein L., Liu X., Logullo P., McCradden M.D., Liu N., Oakden-Rayner L., Singh K., Ting D.S., Wynants L., Yang B., Reitsma J.B., Riley R.D., Collins G.S., Van Smeden M. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ.* 2025;388:e082505. PMID: 40127903; PMCID: PMC11931409. <https://doi.org/10.1136/bmj-2024-082505>
24. Wolff R.F., Moons K.G.M., Riley R.D., Whiting P.F., Westwood M., Collins G.S., Reitsma J.B., Kleijnen J., Mallett S.; PROBAST Group†. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med.* 2019;170(1):51-58. PMID: 30596875. <https://doi.org/10.7326/M18-1376>
25. Page M.J., McKenzie J.E., Bossuyt P.M., Boutron I., Hoffmann T.C., Mulrow C.D., Shamseer L., Tetzlaff J.M., Akl E.A., Brennan S.E., Chou R., Glanville J., Grimshaw J.M., Hróbjartsson A., Lalu M.M., Li T., Loder EW, Mayo-Wilson E., McDonald S., McGuinness L.A., Stewart L.A., Thomas J., Tricco A.C., Welch V.A., Whiting P., Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71. PMID: 33782057; PMCID: PMC8005924. <https://doi.org/10.1136/bmj.n71>
26. 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. PMID: 36658617; PMCID: PMC9850511. <https://doi.org/10.1186/s13019-023-02141-4>
27. Zahara R., Soeharto D.F., Widyantoro B., Sugisman, Herlambang B. Validation of EuroSCORE II Scoring System on Isolated CABG Patient in Indonesia. *Egypt Heart J.* 2023;75(1):86. PMID: 37828408; PMCID: PMC10570231. <https://doi.org/10.1186/s43044-023-00410-0>
28. Mokhtar M.N., Abdullah F.H., Abdul Rahman M.R., Ooi J.S.M. Evaluation of cardiac operative risk and outcome using EuroSCORE II for coronary artery bypass graft surgery. *Anaesth Pain Intensive Care.* 2023;27(3):307-14. <https://doi.org/10.35975/apic.v27i3.2228>
29. Wolf P.J.W., Amato V.L. 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. PMID: 39241182; PMCID: PMC11378809. <https://doi.org/10.21470/1678-9741-2023-0282>
30. Hirnle G., Aboul-Hassan S.S., Knapik P., Tobota Z., Maruszewski B.J., Rogowski J., Pawliszak W., Bugajski P., Deja M., Hrapkowicz T. Performance of EuroSCORE II in Octogenarians Undergoing Coronary Artery Surgery (from the KROK Registry). *J Clin Med.* 2024;13(22):6863. PMID: 39598006; PMCID: PMC11594418. <https://doi.org/10.3390/jcm13226863>
31. Musa A.F., Cheong X.P., Dillon J., Nordin R.B. 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;7:534. PMID: 32913630; PMCID: PMC7459866. <https://doi.org/10.12688/f1000research.14760.2>
32. Hogervorst E.K., Rosseel P.M.J., van de Watering L.M.G., Brand A., Bentala M., van der Meer B.J.M., van der Bom J.G. Prospective validation of the EuroSCORE II risk model in a single Dutch cardiac surgery centre. *Neth Heart J.* 2018;26(11):540-551. PMID: 30232783; PMCID: PMC6220016. <https://doi.org/10.1007/s12471-018-1161-x>
33. Gao F., Shan L., Wang C., Meng X., Chen J., Han L., Zhang Y., Li Z. Predictive Ability of European Heart Surgery Risk Assessment System II (EuroSCORE II) and the Society of Thoracic Surgeons (STS) Score for in-Hospital Mortality of Patients Undergoing Coronary Artery Bypass Grafting. *Int J Gen Med.* 2021;14:8509-8519. PMID: 34824547; PMCID: PMC8610380. <https://doi.org/10.2147/IJGM.S338819>
34. Rodríguez Lima D.R., Rodríguez Aparicio E.E., Otálora González L., Hernández D.C., 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. PMID: 40161567; PMCID: PMC11947952. <https://doi.org/10.4330/wjc.v17.i3.100506>
35. Mastroiacovo G., Bonomi A., Ludergrani M., Franchi M., Maragna R., Pirola S., Baggiano A., Caglio A., Pontone G., Polvani G., Merlino L. Is EuroSCORE II still a reliable predictor for cardiac surgery mortality in 2022? A retrospective study. *Eur J Cardiothorac Surg.* 2022;64(3):ezad294. PMID: 37669150; PMCID: PMC10722878. <https://doi.org/10.1093/ejcts/ezad294>
36. Di Bacco L., D'Alonzo M., Baudo M., Montisci A., Di Eusanio M., Folliguet T., Solinas M., Miceli A., Fischlein T., Rosati F., Muneretto C. Reliability of EuroSCORE II on Prediction of Thirty-Day Mortality and Long-Term Results in Patients Treated with Sutureless Valves. *J Clin Med.* 2024;13(13):3986. PMID: 38999550; PMCID: PMC11242606. <https://doi.org/10.3390/jcm13133986>

37. Kim K.M., Arghami A., Habib R., Daneshmand M.A., Parsons N., Elhalabi Z., Krohn C., Thourani V., Bowdish M.E. The Society of Thoracic Surgeons Adult Cardiac Surgery Database: 2022 Update on Outcomes and Research. *Ann Thorac Surg.* 2023;115(3):566-574. PMID: 36623634.
<https://doi.org/10.1016/j.athoracsur.2022.12.033>
38. Saka E., Öztürk E., Yüksel A.E., Kocabaş N.S. Comparison of EuroSCORE II and STS Risk Scoring Systems in Patients who Underwent Open-heart Surgery. *Turk J Anaesthesiol Reanim.* 2025;53(4):163-169. PMID: 40631722; PMCID: PMC12288423.
<https://doi.org/10.4274/TJAR.2025.241778>
39. Machado R.J., Saraiva F.A., Mancio J., Sousa P., Cerqueira R.J., Barros A.S., Lourenço A.P., Leite-Moreira A.F. A systematic review and meta-analysis of randomized controlled studies comparing off-pump versus on-pump coronary artery bypass grafting in the elderly. *J Cardiovasc Surg (Torino).* 2022;63(1):60-68. PMID: 34792312. <https://doi.org/10.23736/S0021-9509.21.12012-9>
40. Chen S., Zhang S., Cai S., Wang H. Impact of frailty on outcomes following coronary artery bypass grafting: a systematic review and meta-analysis. *BMC Surg.* 2024;24(1):419. PMID: 39725933; PMCID: PMC11673924.
<https://doi.org/10.1186/s12893-024-02728-1>
41. Chandiramani A., Ali J.M. Frailty in Cardiac Surgery-Assessment Tools, Impact on Outcomes, and Optimisation Strategies: A Narrative Review. *J Cardiovasc Dev Dis.* 2025;12(4):127. PMID: 40278186; PMCID: PMC12028191.
<https://doi.org/10.3390/jcdd12040127>
42. Binuya M.A.E., Engelhardt E.G., Schats W., Schmidt M.K., Steyerberg E.W. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. *BMC Med Res Methodol.* 2022;22(1):316. PMID: 36510134; PMCID: PMC9742671. <https://doi.org/10.1186/s12874-022-01801-8>

Supplement. Search strategy.

Last search date: 06.10.2025

PubMed/MEDLINE

Nº	Request	Result
#1	«EuroSCORE II»[Title/Abstract] OR «EuroSCORE 2»[Title/Abstract] OR «STS score»[Title/Abstract] OR «STS risk»[Title/Abstract] OR «Society of Thoracic Surgeons»[Title/Abstract]	5 857
#2	«coronary artery bypass»[MeSH Terms] OR «coronary artery bypass»[Title/Abstract] OR «CABG»[Title/Abstract] OR «coronary bypass»[Title/Abstract]	82 543
#3	validat*[Title/Abstract] OR calibrat*[Title/Abstract] OR discriminat*[Title/Abstract] OR performance[Title/Abstract]	2 925 498
#4	#1 AND #2 AND #3 Filters: English, 2018–2025	112

Scopus

Nº	Request	Result
#1	TITLE-ABS-KEY(«EuroSCORE II» OR «EuroSCORE 2» OR «STS score» OR «STS risk» OR «Society of Thoracic Surgeons»)	6 412
#2	TITLE-ABS-KEY(«coronary artery bypass» OR «CABG» OR «coronary bypass»)	57 836
#3	TITLE-ABS-KEY(validat* OR calibrat* OR discriminat* OR performance)	4 127 614
#4	#1 AND #2 AND #3 AND PUBYEAR > 2017 AND PUBYEAR < 2026 AND LANGUAGE(english)	97

Web of Science (Core Collection)

Nº	Request	Result
#1	TS=(«EuroSCORE II» OR «EuroSCORE 2» OR «STS score» OR «STS risk» OR «Society of Thoracic Surgeons»)	5 943
#2	TS=(«coronary artery bypass» OR «CABG» OR «coronary bypass»)	51 724
#3	TS=(validat* OR calibrat* OR discriminat* OR performance)	3 486 251
#4	#1 AND #2 AND #3 Indexes: SCI-EXPANDED, ESCI Timespan: 2018–2025 Language: English	105

Outcomes

Source	Record count
PubMed/MEDLINE	112
Scopus	97
Web of Science	105
Total identified	314

Notes

* – truncation to retrieve all word forms (e.g., validat* = validation, validated, validating).

The number of records for queries #1–#3 is presented without date or language restrictions. Restrictions were applied at the combination stage (query #4).

Additionally, a manual review of the reference lists of included articles and relevant reviews was performed (no additional records were identified).

After removal of 160 duplicates, 154 records were eligible for screening.