

Validation of EuroSCORE II in post-cardiac surgery patients in a tertiary institution in Hong Kong

Karen HL Ng, Kailu Wang, Takuya Fujikawa, Micky WT Kwok, Jacky YK Ho, Simon CY Chow, Joyce WY Chan, Kevin Lim, Aliss TC Chang, Ivan CH Siu, Randolph HL Wong *

ABSTRACT

Introduction: This study aimed to assess the discriminatory ability and calibration performance of the European System for Cardiac Operative Risk Evaluation (EuroSCORE) II, a widely used risk prediction tool, in predicting postoperative mortality among patients undergoing cardiac surgery at Prince of Wales Hospital (PWH) in Hong Kong.

Methods: Complete data from 4180 patients who underwent cardiac surgery at PWH between 2013 and 2023 were available for validation of EuroSCORE II and comparison of its discriminatory ability with the logistic EuroSCORE. Discriminatory performance was primarily assessed using the area under the receiver operating characteristic curve (AUROC). Calibration was evaluated using the Hosmer–Lemeshow test, coefficient of determination (R^2), and normalised root mean square error (NRMSE).

Results: EuroSCORE II demonstrated strong discrimination and good calibration for predicting 30-day mortality in the overall cohort (AUROC=0.829; Hosmer–Lemeshow $P=0.155$) and key subgroups: isolated coronary artery bypass grafting (CABG) [AUROC=0.847; $P=0.113$], isolated valve surgery (AUROC=0.810; $P=0.162$), and aortic surgery (AUROC=0.735; $P=0.549$). More than 85% of the variation in 30-day mortality (R^2) was explained across these groups. Compared with the logistic EuroSCORE, EuroSCORE II showed improved discrimination and calibration, with higher AUROC

values and lower NRMSE.

Conclusion: EuroSCORE II demonstrates strong discriminatory ability and good calibration for predicting 30-day mortality among patients undergoing cardiac surgery and within key subgroups—isolated CABG, isolated valve surgery, and aortic surgery—in this cohort.

Hong Kong Med J 2026;32:126–34

<https://doi.org/10.12809/hkmj2514125>

¹ KHL Ng, MStat

² K Wang, PhD

^{1,3} T Fujikawa, MD

^{1,3} MWT Kwok, MB, ChB, FRCS

^{1,3} JYK Ho, MB, ChB, FRCS

^{1,3} SCY Chow, MB, ChB, FRCS

^{1,3} JWY Chan, MB, BS, FRCS

^{1,3} K Lim, MB, ChB, FRCS

^{1,3} ATC Chang, MB, ChB, FRCS

^{1,3} ICH Siu, MB, ChB, MRCS

^{1,3} RHL Wong *, MB, ChB, FRCS

¹ Division of Cardiothoracic Surgery, Department of Surgery, Prince of Wales Hospital, Hong Kong SAR, China

² Centre for Health Systems and Policy Research, The Jockey Club School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China

³ Division of Cardiothoracic Surgery, Department of Surgery, The Chinese University of Hong Kong, Hong Kong SAR, China

* Corresponding author: wonh11@surgery.cuhk.edu.hk

This article was published on 13 Apr 2026 at www.hkmj.org.

New knowledge added by this study

- The European System for Cardiac Operative Risk Evaluation (EuroSCORE) II demonstrates strong discriminatory ability and good calibration for predicting 30-day mortality among patients undergoing cardiac surgery at Prince of Wales Hospital (PWH) in Hong Kong.
- EuroSCORE II demonstrates improved discrimination and calibration compared with the logistic EuroSCORE in the overall cardiac surgery cohort at PWH.
- Within the aortic subgroup, EuroSCORE II demonstrates statistically significant improvements in discrimination and calibration relative to the logistic EuroSCORE.

Implications for clinical practice or policy

- EuroSCORE II represents a reliable risk stratification tool for guiding treatment decisions, identifying high-risk patients and optimising resource allocation.
- Incorporation of additional variables into EuroSCORE II may further enhance predictive accuracy and enable tailored interventions for post-cardiac surgery patients.

Introduction

The Global Burden of Disease Results Tool of the Institute for Health Metrics and Evaluation reported that cardiovascular diseases accounted for approximately 10 383 550 deaths globally in 2017, representing 18.56% of all-cause mortality.¹ Cardiothoracic surgery plays an important role in the treatment of these conditions and in reducing associated morbidity and mortality. However, surgery carries inherent risks that vary among patients, necessitating careful evaluation of risks and benefits before proceeding. A risk stratification tool is essential for effective patient triage and the consent process.

One widely used risk stratification tool is the European System for Cardiac Operative Risk Evaluation (EuroSCORE), a specialised scoring system that provides customised predictions of in-hospital mortality after cardiac surgery. The tool assigns scores based on various preoperative risk factors to stratify patients into different risk categories (low: EuroSCORE <4%, intermediate: 4-8%, high: >8%).² In the UK, in-hospital mortality declined from 4.0% to 2.8% between 2002 and 2016 following implementation of EuroSCORE,³ supporting its value in cardiac surgical risk assessment. The EuroSCORE comprises three versions: the additive EuroSCORE,⁴ the logistic EuroSCORE,⁵ and EuroSCORE II.⁶ In 2012, the Society for Cardiothoracic Surgery in Great Britain and Ireland recommended the use of the latest version, EuroSCORE II.⁶

Prince of Wales Hospital (PWH) in Hong Kong has adopted the logistic EuroSCORE for risk assessment since 2007. However, several publications from different countries have raised concerns regarding the accuracy of the additive and logistic EuroSCORE models, leading to the development of EuroSCORE II.⁷⁻⁹ Consequently, EuroSCORE II has been proposed as the future risk adjustment tool of the Society for Cardiothoracic Surgery in Great Britain and Ireland following successful contemporary validation.^{6,10,11}

Although EuroSCORE II has been widely used and validated, the underlying data were predominantly derived from Western populations undergoing cardiac surgery in Europe and the US.^{4,12-14} Therefore, studies evaluating the performance of EuroSCORE II in Asian populations remain limited,¹⁵⁻¹⁷ and none has been conducted specifically in Hong Kong. Furthermore, no studies have compared the performance of the logistic EuroSCORE and EuroSCORE II in the Hong Kong population. The present study aimed to address these gaps.

Moreover, Hong Kong has a higher proportion of aortic surgery than Western countries. Prince of Wales Hospital reported a surge in aortic surgeries, reaching 26% between 2021 and 2022,¹⁸ whereas the

香港一所大學醫院心臟手術患者的EuroSCORE II 驗證研究

吳愷羚、王凱路、藤川拓也、郭慧婷、何殷傑、周志英、
陳詠欣、林諾賢、鄭芷晴、蕭智軒、黃鴻亮

引言：本研究旨在評估常用的心臟手術風險評估工具EuroSCORE II在香港威爾斯親王醫院接受心臟手術的患者中是否能準確預測術後死亡風險。

方法：本研究分析2013至2023年間於威爾斯親王醫院接受心臟手術的4180名患者資料，以驗證EuroSCORE II及評估其預測表現，並與舊版本邏輯EuroSCORE比較。區辨能力以受試者工作特徵曲線下面積（AUROC）評估，以判斷模型能否有效區分高風險與低風險患者，校準度則採用Hosmer–Lemeshow檢定、決定系數（R²）及標準化均方根誤差來評估。

結果：EuroSCORE II在預測心臟手術後30天死亡風險方面表現良好，既能有效區分整體隊列中的高風險與低風險患者（AUROC=0.829；Hosmer–Lemeshow P=0.155）。在不同手術類型中，包括單純冠狀動脈旁路移植術（AUROC=0.847；P=0.113）、單純瓣膜手術（AUROC=0.810；P=0.162）及主動脈手術（AUROC=0.735；P=0.549），其預測表現同樣理想。此外，模型在各組別中均可解釋超過85%的30天死亡風險差異（R²）。與舊版本邏輯EuroSCORE相比，EuroSCORE II在區辨能力及預測準確度方面更為優勝，AUROC較高及預測誤差較低。

結論：在本研究隊列中，EuroSCORE II在預測接受心臟手術患者及主要手術類別（單純冠狀動脈旁路移植術、單純瓣膜手術及主動脈手術）的30天死亡率方面，展現良好的區辨能力及校準表現。

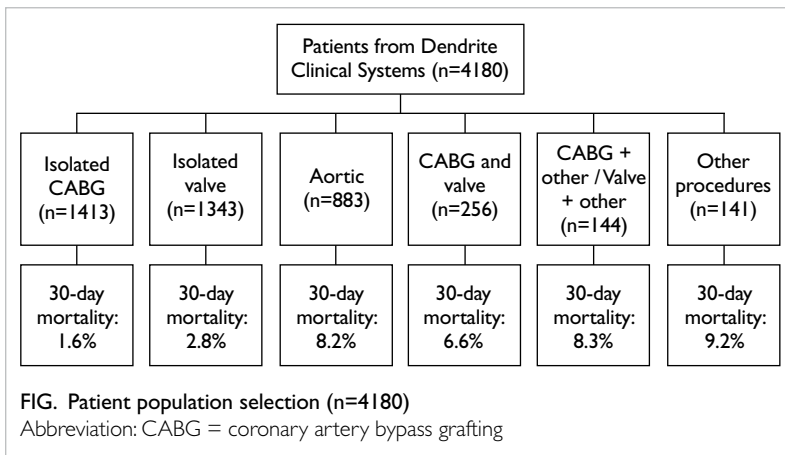
UK reported an aortic surgery prevalence of 3.47% between 2015 and 2016.³ We therefore sought to investigate whether this variation influences the validity of EuroSCORE II through subgroup analyses.

The primary objective of this study was to assess the discriminatory ability and calibration of EuroSCORE II in predicting postoperative mortality after the three main index cardiac surgeries (ie, coronary artery bypass grafting [CABG], valve surgery, and aortic surgery) at our centre.^{2,7-9,14-17,19-23} The secondary objective was to compare the discriminatory ability and calibration of EuroSCORE II with those of the logistic EuroSCORE in patients undergoing cardiac surgery.

Methods

Study design and population cohort

This retrospective validation study included patients (aged ≥18 years) who underwent all types of cardiac surgery—including CABG, valve surgery (eg, aortic valve replacement, mitral valve replacement, and tricuspid valve repair), aortic surgery, isolated or combined procedures, and other procedures (eg, left atrial appendage closure) at PWH between 1 January 2013 and 31 December 2023 (inclusive) [Fig]. Because



PWH does not perform certain cardiothoracic procedures—such as paediatric cardiac surgery, cardiac transplantation, and oesophageal surgery—records for these interventions were unavailable. For patients who underwent multiple cardiac surgeries during the same hospital admission, only the first index procedure was analysed. The minimum sample size of 225 was calculated based on estimates of area under the receiver operating characteristic curve (AUROC) from the literature⁷ and the estimated prevalence of the outcome (online supplementary Table),^{7,24} indicating that our primary cohorts for CABG, valve surgery, and aortic surgery exceeded the required sample size.

Data collection and outcomes

The Dendrite cardiac surgery database (Dendrite Clinical Systems, Oxford, United Kingdom)²⁵ was utilised for secondary data collection (Fig). This database captures clinically relevant information, including preoperative medical records and postoperative complications for patients undergoing cardiac surgery. All key variables required to calculate EuroSCORE II⁶ and the logistic EuroSCORE⁵ were extracted. Mortality, the primary outcome, was defined as death within 30 days of the index operation (regardless of place of death), consistent with previous studies.^{2,9,14,15,21}

Statistical analyses

Cases with missing or incomplete data required for calculation of EuroSCORE II were excluded from the analysis. Analyses were performed for the overall cohort and stratified by individual cardiac procedure. Complete data were available for validation of EuroSCORE II and comparison of its predictive performance with that of the logistic EuroSCORE for postoperative mortality.

Univariate and multivariate binary logistic regression analyses were conducted on all relevant variables included in the EuroSCORE II scale to

identify significant covariates associated with an increased risk of mortality.

The discriminatory performance of the predictive models was evaluated using the AUROC; values of 0.8 or above indicated strong discrimination, and 1.0 indicated perfect discrimination. Pairwise comparisons of AUROCs for individual cardiac procedures were performed using the DeLong test, with the threshold for statistical significance set at $P < 0.05$.

Calibration of the predictive model was evaluated using the Hosmer–Lemeshow goodness-of-fit test and calibration plots, through statistical and graphical assessment of agreement between observed and expected event rates within model subgroups. A P value > 0.05 and a regression line approximating the 45-degree diagonal indicated good calibration, reflecting adequate agreement between observed and predicted event rates.

Model goodness of fit was further assessed using the coefficient of determination (R^2), which quantifies the proportion of variance explained by the model, and the normalised root mean square error (NRMSE), which measures predictive accuracy by comparing predicted and observed values, normalised to the data range. Higher R^2 values and lower NRMSE values indicate better model fit.

Statistical analyses were performed using SPSS (Windows version 29.0; IBM Corp, Armonk [NY], United States), Microsoft Excel 2019, and R software (RStudio, version 2024.04.2).

Results

Patient characteristics

The study cohort comprised 4180 patients (Fig). Table 1 summarises the characteristics of the overall cohort and relevant subgroups. The median age was 63 years (interquartile range, 56–69), and 29.2% ($n=1222$) were women. Aortic operations were performed in 21.1% ($n=883$) of patients and the majority underwent a single non-coronary procedure (36.9%, $n=1541$). For the overall cohort, the median logistic EuroSCORE value was 5.8 (interquartile range, 2.6–13.7), whereas median EuroSCORE II value was 2.4 (interquartile range, 1.2–5.3). The institutional 30-day mortality rate for all cardiac procedures was 4.2%.

Primary outcome

Discriminatory and calibration performance

The AUROC for EuroSCORE II was 0.829, indicating strong discriminatory ability. The Hosmer–Lemeshow P value for EuroSCORE II was 0.155, indicating no statistically significant difference between predicted and observed values (online supplementary Fig a). Accordingly, EuroSCORE II demonstrated acceptable calibration.

TABLE I. Patient demographics and characteristics*

	All patients (n=4180)	Isolated CABG (n=1413)	Isolated valve (n=1343)	Aortic (n=883)	CABG and valve (n=256)	CABG + other / Valve + other (n=144)	Others (n=141)
Preoperative factors							
Patient-related factors							
Age, y	63 (56-69)	63 (57-69)	63 (56-69)	61 (53-68)	66 (60-71)	63.5 (55-69)	61 (52-69)
Female sex	1222 (29.2%)	238 (16.8%)	590 (43.9%)	195 (22.1%)	61 (23.8%)	71 (49.3%)	67 (47.5%)
COPD	413 (9.9%)	131 (9.3%)	127 (9.5%)	104 (11.8%)	29 (11.3%)	10 (6.9%)	12 (8.5%)
Extracardiac arteriopathy	508 (12.2%)	93 (6.6%)	27 (2.0%)	357 (40.4%)	18 (7.0%)	5 (3.5%)	8 (5.7%)
History of poor mobility	50 (1.2%)	16 (1.1%)	10 (0.7%)	15 (1.7%)	2 (0.8%)	4 (2.8%)	3 (2.1%)
Previous cardiac surgery	312 (7.5%)	5 (0.4%)	110 (8.2%)	164 (18.6%)	6 (2.3%)	11 (7.6%)	16 (11.3%)
Active endocarditis	131 (3.1%)	0	104 (7.7%)	9 (1.0%)	6 (2.3%)	12 (8.3%)	0
Critical preoperative condition	445 (10.6%)	170 (12.0%)	123 (9.2%)	66 (7.5%)	31 (12.1%)	26 (18.1%)	29 (20.6%)
Renal impairment requiring dialysis	104 (2.5%)	41 (2.9%)	33 (2.5%)	8 (0.9%)	10 (3.9%)	6 (4.2%)	6 (4.3%)
Insulin-dependent diabetes mellitus	227 (5.4%)	182 (12.9%)	22 (1.6%)	2 (0.2%)	11 (4.3%)	5 (3.5%)	5 (3.5%)
BMI, kg/m ²	24.3 (22.0-27.1)	24.9 (22.8-27.7)	23.3 (21.1-26.0)	24.8 (22.3-27.8)	23.8 (21.8-26.4)	23.5 (21.5-26.6)	24.1 (21.0-26.8)
Current smoker	665 (15.9%)	251 (17.8%)	126 (9.4%)	207 (23.4%)	44 (17.2%)	15 (10.4%)	22 (15.6%)
Hypercholesterolaemia	1782 (42.6%)	963 (68.2%)	362 (27.0%)	240 (27.2%)	129 (50.4%)	50 (34.7%)	38 (27.0%)
Hypertension	2704 (64.7%)	1087 (76.9%)	661 (49.2%)	631 (71.5%)	186 (72.7%)	71 (49.3%)	68 (48.2%)
Neurological dysfunction	95 (2.3%)	23 (1.6%)	21 (1.6%)	30 (3.4%)	10 (3.9%)	4 (2.8%)	7 (5.0%)
Cardiac-related factors							
CCS class IV	294 (7.0%)	127 (9.0%)	14 (1.0%)	119 (13.5%)	16 (6.3%)	7 (4.9%)	11 (7.8%)
LVEF ≤30%	173 (4.1%)	79 (5.6%)	40 (3.0%)	22 (2.5%)	18 (7.0%)	10 (6.9%)	4 (2.8%)
Left main stem disease	608 (14.5%)	549 (38.9%)	6 (0.4%)	12 (1.4%)	34 (13.3%)	5 (3.5%)	2 (1.4%)
Recent MI within 90 days	616 (14.7%)	488 (34.5%)	13 (1.0%)	23 (2.6%)	60 (23.4%)	12 (8.3%)	20 (14.2%)
Pulmonary hypertension	1278 (30.6%)	141 (10.0%)	792 (59.0%)	95 (10.8%)	129 (50.4%)	84 (58.3%)	37 (26.2%)
NYHA class ≥III	1024 (24.5%)	206 (14.6%)	433 (32.2%)	190 (21.5%)	96 (37.5%)	54 (37.5%)	45 (31.9%)
Operation-related factors							
Surgery on the thoracic aorta	883 (21.1%)	0	0	883 (100%)	0	0	0
Emergency surgery	530 (12.7%)	29 (2.1%)	52 (3.9%)	381 (43.1%)	12 (4.7%)	17 (11.8%)	39 (27.7%)
Type of surgery							
1 coronary procedure	1413 (33.8%)	1413 (100%)	0	0	0	0	0
1 non-coronary procedure	1541 (36.9%)	0	845 (62.9%)	560 (63.4%)	0	8 (5.6%)	128 (90.8%)
2 procedures	851 (20.4%)	0	344 (25.6%)	276 (31.3%)	164 (64.1%)	54 (37.5%)	13 (9.2%)
≥3 procedures	375 (9.0%)	0	154 (11.5%)	47 (5.3%)	92 (35.9%)	82 (56.9%)	0
Postoperative outcomes							
30-day mortality	174 (4.2%)	22 (1.6%)	38 (2.8%)	72 (8.2%)	17 (6.6%)	12 (8.3%)	13 (9.2%)
Logistic EuroSCORE	5.8 (2.6-13.7)	2.3 (1.4-4.8)	6.0 (3.4-11.1)	18.4 (10.1-30.1)	6.9 (3.7-13.8)	7.5 (4.4-16.5)	5.8 (2.7-19.9)
EuroSCORE II	2.4 (1.2-5.3)	1.3 (0.8-2.4)	2.3 (1.3-4.8)	5.2 (2.9-9.3)	3.6 (2.2-7.5)	3.6 (2.5-9.8)	2.0 (1.0-6.0)

Abbreviations: BMI = body mass index; CABG = coronary artery bypass grafting; CCS = Canadian Cardiovascular Society; COPD = chronic obstructive pulmonary disease; EuroSCORE = European System for Cardiac Operative Risk Evaluation; LVEF = left ventricular ejection fraction; MI = myocardial infarction; NYHA = New York Heart Association

* Data are shown as median (interquartile range) or No. (%)

Comparison between logistic EuroSCORE and EuroSCORE II

EuroSCORE II demonstrated a statistically significant improvement in discriminatory performance compared with the logistic EuroSCORE (DeLong $P=0.006$). Additionally, EuroSCORE II showed superior calibration, supported by a significant Hosmer–Lemeshow test result for the logistic EuroSCORE ($P<0.001$). Calibration curves comparing observed and predicted 30-day mortality were consistent with these findings, further indicating better calibration with EuroSCORE II than with the logistic EuroSCORE. More than 90% of the variation in 30-day mortality was explained by both models (R^2 for EuroSCORE II=98.7%; R^2 for logistic EuroSCORE=99.1%). Notably, EuroSCORE II demonstrated a substantially lower NRMSE (5.7%) than the logistic EuroSCORE (56.4%), indicating reduced dispersion and relative variability in predictions (online supplementary Fig a).

Subgroup analysis

Isolated coronary artery bypass surgery

In this subgroup, EuroSCORE II demonstrated strong discriminatory performance (AUROC=0.847) and acceptable calibration (Hosmer–Lemeshow $P=0.113$). There was no statistically significant difference in discriminatory performance between EuroSCORE II and the logistic EuroSCORE (DeLong $P=0.529$). However, EuroSCORE II showed better calibration, supported by a significant Hosmer–Lemeshow test result for the logistic EuroSCORE ($P<0.001$) and calibration curves favouring EuroSCORE. More than 85% of the variation in 30-day mortality was explained by both models (R^2 for EuroSCORE II=87.7%; R^2 for logistic EuroSCORE=91.4%). Compared with the logistic EuroSCORE (40.6%), EuroSCORE II demonstrated a lower NRMSE (13.0%), indicating reduced dispersion and relative variability (online supplementary Fig b).

Isolated valve surgery

In this subgroup, EuroSCORE II demonstrated strong discriminatory performance (AUROC=0.810) and acceptable calibration (Hosmer–Lemeshow $P=0.162$). There was no statistically significant difference in discriminatory performance between EuroSCORE II and the logistic EuroSCORE (DeLong $P=0.160$). Nevertheless, EuroSCORE II demonstrated superior calibration, supported by a significant Hosmer–Lemeshow test result for the logistic EuroSCORE ($P<0.001$) and calibration curves favouring EuroSCORE II. More than 90% of the variation in 30-day mortality was explained by both models (R^2 for EuroSCORE II=94.7%; R^2 for logistic EuroSCORE=94.4%). Compared with the logistic EuroSCORE (80.4%), EuroSCORE II

demonstrated a lower NRMSE (21.8%), indicating reduced dispersion and relative variability (online supplementary Fig c).

Aortic surgery

In this subgroup, EuroSCORE II demonstrated satisfactory discriminatory performance (AUROC=0.735) and good calibration (Hosmer–Lemeshow $P=0.549$). It also showed a statistically significant improvement in discrimination compared with the logistic EuroSCORE (DeLong $P<0.001$). Calibration was also superior, supported by a significant Hosmer–Lemeshow test result for the logistic EuroSCORE ($P<0.001$) and calibration curves favouring EuroSCORE II. More than 90% of the variation in 30-day mortality was explained by EuroSCORE II (R^2 for EuroSCORE II=96.1%; R^2 for logistic EuroSCORE=76.6%). EuroSCORE II also demonstrated a lower NRMSE (6.6%) than the logistic EuroSCORE (98.8%), indicating reduced dispersion and relative variability (online supplementary Fig d).

Combined valve and coronary artery bypass surgery

In this subgroup, EuroSCORE II demonstrated fair discriminatory performance (AUROC=0.694) and good calibration (Hosmer–Lemeshow $P=0.606$). There was no statistically significant difference in discriminatory performance between EuroSCORE II and the logistic EuroSCORE (DeLong $P=0.913$). Both models exhibited adequate calibration (EuroSCORE II $P=0.606$; logistic EuroSCORE $P=0.280$) [online supplementary Fig e].

Combined valve or coronary artery bypass surgery and other procedures

In this subgroup, EuroSCORE II demonstrated strong discriminatory performance (AUROC=0.862) and acceptable calibration (Hosmer–Lemeshow $P=0.159$). There was no statistically significant difference in discriminatory performance between EuroSCORE II and the logistic EuroSCORE (DeLong $P=0.248$). However, EuroSCORE II exhibited superior calibration compared with the logistic EuroSCORE (Hosmer–Lemeshow $P=0.062$) [online supplementary Fig f].

Other procedures

In this subgroup, EuroSCORE II demonstrated strong discriminatory performance (AUROC=0.872) but poor calibration (Hosmer–Lemeshow $P<0.001$). There was no statistically significant difference in discriminatory performance between EuroSCORE II and the logistic EuroSCORE (DeLong $P=0.626$). Notably, calibration curves favoured EuroSCORE II over the logistic EuroSCORE (online supplementary Fig g).

Multivariate binary logistic regression analysis

Furthermore, comparison of EuroSCORE II variables with multivariable analyses from PWH database identified dialysis as an additional significant predictor of increased 30-day mortality (adjusted odds ratio=3.401) among patients undergoing cardiac surgery (Table 2).

Discussion

In the present study, EuroSCORE II demonstrated strong discriminatory performance and good calibration in the overall cohort and three key subgroups (isolated CABG, isolated valve surgery and aortic surgery). Moreover, EuroSCORE II outperformed the logistic EuroSCORE in both discrimination and calibration across the overall cohort and these principal subgroups.

Our results are consistent with validation studies conducted in several European countries (Italy,²⁶ Greece,²⁷ Serbia,²⁸ Spain,²⁹ and Hungary³⁰), which demonstrated strong discriminatory performance (AUROC >0.7) for EuroSCORE II.²⁶⁻³¹ These findings reaffirm the robust predictive performance of EuroSCORE II for mortality in patients undergoing cardiac surgery.

In addition to European populations, our findings align with those of validation studies conducted in Asian cohorts.^{15-17,23} Specifically, Liu et al¹⁵ demonstrated strong discriminatory performance for EuroSCORE II, with an AUROC of 0.792 in a single-centre setting. This concordance further supports the consistency and reliability of EuroSCORE II as a mortality prediction tool in Asian cardiac surgery populations.

However, Kurniawaty et al¹⁹ reported considerably different findings, demonstrating only fair discriminatory performance, with evidence of miscalibration and underprediction in an Indonesian population. This discrepancy may be attributable to differences in patient age. Both our cohort and the European cohorts had substantially higher median (63 years) or mean (64.6 years)⁶ ages compared with the mean age in the Indonesian cohort (44 years)¹⁹. Given the younger age profile and lower prevalence of risk factors included in the EuroSCORE II model among Indonesian patients, its predictive performance may be limited in that population. Accordingly, these findings may be less generalisable to the Hong Kong population.

For the overall cohort, EuroSCORE II demonstrated superior performance in both discrimination and calibration compared with the logistic EuroSCORE. This difference may reflect the tendency of the logistic EuroSCORE to overestimate mortality risk, particularly in high-risk emergency patients.¹⁰ Consequently, EuroSCORE II appears to provide more accurate risk stratification than the

TABLE 2. Multivariable analysis of risk factors associated with postoperative mortality in the overall cohort (n=4180)

	Adjusted OR (95% CI)	P value
Patient-related factors		
Age	1.029 (1.009-1.049)	0.004
Female sex	0.925 (0.625-1.369)	0.696
COPD	0.986 (0.571-1.704)	0.961
Extracardiac arteriopathy	1.427 (0.895-2.276)	0.135
History of poor mobility	0.734 (0.236-2.286)	0.594
Previous cardiac surgery	2.045 (1.190-3.513)	0.010
Active endocarditis	1.313 (0.597-2.886)	0.499
Critical preoperative condition	1.454 (0.821-2.574)	0.199
Renal impairment requiring dialysis	3.401 (1.707-6.776)	<0.001
Insulin-dependent diabetes mellitus	1.093 (0.494-2.417)	0.827
BMI	0.972 (0.929-1.017)	0.214
Current smoker	0.8 (0.494-1.296)	0.365
Hypercholesterolaemia	0.913 (0.624-1.335)	0.639
Hypertension	1.338 (0.897-1.994)	0.153
Neurological dysfunction	1.156 (0.456-2.929)	0.76
Cardiac-related factors		
CCS class IV	1.742 (1.072-2.83)	0.025
LVEF ≤30%	1.324 (0.66-2.656)	0.429
Left main stem disease	1.068 (0.542-2.108)	0.849
Recent MI within 90 days	2.090 (1.173-3.724)	0.012
Pulmonary hypertension	0.976 (0.618-1.542)	0.918
NYHA class ≥III	1.870 (1.277-2.737)	<0.001
Operation-related factors		
Surgery on the thoracic aorta	1.825 (1.062-3.137)	0.029
Emergency surgery	1.863 (1.12-3.098)	0.017
Type of surgery		
1 coronary procedure		<0.001
1 non-coronary procedure	0.224 (0.104-0.483)	
2 procedures	0.461 (0.263-0.809)	
≥3 procedures	0.7 (0.404-1.214)	
Logistic EuroSCORE	1.002 (0.984-1.021)	0.806
EuroSCORE II	1.022 (0.997-1.048)	0.085

Abbreviations: 95% CI = 95% confidence interval; BMI = body mass index; CCS = Canadian Cardiovascular Society; COPD = chronic obstructive pulmonary disease; EuroSCORE = European System for Cardiac Operative Risk Evaluation; LVEF = left ventricular ejection fraction; MI = myocardial infarction; NYHA = New York Heart Association; OR = odds ratio

logistic EuroSCORE.

For isolated CABG procedures, the discriminatory performance of EuroSCORE II was strong in our study, supported by a non-significant Hosmer–Lemeshow statistic, consistent with findings from a large UK validation cohort.⁷ Studies conducted in Finland³² (AUROC=0.852) and China¹⁶ (AUROC=0.762) similarly demonstrated robust discriminatory performance of EuroSCORE

II in predicting operative mortality among high-risk isolated CABG patients and those undergoing CABG with or without concomitant major cardiac surgery. However, a study from Singapore reported poor discrimination and calibration, particularly in moderate- and high-risk cohorts.³³ Comparable findings were reported in studies from Indonesia²² and Malaysia,²³ which demonstrated fair discrimination but underestimation of mortality after isolated CABG. These discrepancies suggest that additional caution may be warranted when applying EuroSCORE II in isolated CABG populations. Differences in demographic characteristics or study design may contribute to variability in model performance, warranting further investigation.

For aortic procedures, EuroSCORE II demonstrated higher AUROC values and more favourable Hosmer–Lemeshow P values than the logistic EuroSCORE. Nevertheless, caution is warranted because the model does not incorporate specific procedural variables (eg, open surgery vs minimally invasive approaches) as risk factors, which may limit precision in mortality prediction for aortic surgery.⁷

The adoption of contemporary machine learning and artificial intelligence techniques, rather than logistic regression, may offer more effective modelling approaches for capturing complex, non-linear interactions among established risk factors. Furthermore, incorporating the statistically significant variable identified through multivariate analysis of the PWH database, specifically dialysis, into a future EuroSCORE III model may further enhance its predictive performance.

Strengths

First, the robustness of this validation study is supported by its substantial sample size (n=4180), which increases statistical power, enables detection of smaller effects, and enhances generalisability. Second, the absence of missing data strengthens measurement completeness and the credibility of the validation process, reduces information bias, and facilitates a more precise evaluation of EuroSCORE II predictive performance within this large cohort.

Limitations

First, reliance on data from a single institution may introduce sampling bias. Therefore, multi-centre analyses should be conducted in future, provided sufficient resources are available. Second, the retrospective design limited the study by precluding long-term follow-up after patient discharge. Consequently, the analysis did not capture longer-term outcomes that may be influenced by baseline EuroSCORE II risk estimates. Additionally, the cohort demonstrated a skewed distribution across risk categories, with a substantial proportion (>85.2%)

categorised as low or intermediate risk, thereby limiting generalisability to high-risk populations.

Future research

First, in aortic surgery, the discrepancy in EuroSCORE II performance observed between Hong Kong and the UK indicates a need for further investigation.¹⁰ A meta-analysis focusing on validation of EuroSCORE II in aortic procedures could help refine risk assessment in this subgroup. Second, although EuroSCORE II is a valuable risk stratification tool in cardiac surgery, minimally invasive cardiac procedures³⁴ and certain established risk factors (eg, diffuse coronary artery disease and aortic calcification)¹⁵ are not included in the model. Accordingly, there may be a need for in-depth evaluation of their relevance to EuroSCORE II calculation. Third, multi-centre studies would enable validation of these findings on a broader scale. Collaboration with the other two cardiac centres in Hong Kong would enhance generalisability and support more robust conclusions.

Conclusion

In our cohort, EuroSCORE II demonstrated strong discriminatory performance and good calibration for predicting 30-day postoperative mortality among patients undergoing cardiac surgery. It also shows superior calibration and comparable or improved discrimination in the three principal subgroups— isolated CABG, isolated valve surgery, and aortic surgery—compared with the logistic EuroSCORE. Accordingly, EuroSCORE II represents a risk stratification tool superior to the logistic EuroSCORE and is well suited for use in Hong Kong.

Author contributions

Concept or design: KHL Ng, T Fujikawa, K Wang, RHL Wong. Acquisition of data: KHL Ng, MWT Kwok, JYK Ho, SCY Chow, JWY Chan, K Lim, ATC Chang, ICH Siu, T Fujikawa, RHL Wong.

Analysis or interpretation of data: KHL Ng, T Fujikawa, K Wang, RHL Wong.

Drafting of the manuscript: KHL Ng, T Fujikawa, K Wang, RHL Wong.

Critical revision of the manuscript for important intellectual content: KHL Ng, T Fujikawa, RHL Wong.

All authors had full access to the data, contributed to the study, approved the final version for publication, and take responsibility for its accuracy and integrity.

Conflicts of interest

All authors have disclosed no conflicts of interest.

Acknowledgement

The authors thank Dr Simon KS Yau from Department of Family Medicine of the New Territories East Cluster for his

insightful contributions to data interpretation and manuscript revision.

Declaration

This research was presented at The Hospital Authority Convention 2025 (26 May 2025, Hong Kong).

Funding/support

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics approval

This research was approved by the Institutional Review Board of The Chinese University of Hong Kong/Hospital Authority New Territories East Cluster, Hong Kong (Ref No.: 2024.571). The requirement for informed patient consent was waived by the Board due to the retrospective nature of the study.

Supplementary material

The supplementary material was provided by the authors and some information may not have been peer reviewed. Accepted supplementary material will be published as submitted by the authors, without any editing or formatting. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by the Hong Kong Academy of Medicine and the Hong Kong Medical Association. The Hong Kong Academy of Medicine and the Hong Kong Medical Association disclaim all liability and responsibility arising from any reliance placed on the content. To view the file, please visit the journal online (<https://doi.org/10.12809/hkmj2514125>).

References

- Vervoort D, Swain JD, Pezzella AT, Kpodonu J. Cardiac surgery in low- and middle-income countries: a state-of-the-art review. *Ann Thorac Surg* 2021;111:1394-400.
- 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;18:40.
- Society for Cardiothoracic Surgery in Great Britain and Ireland. Blue Books. Available from: <https://scts.org/professionals/reports/resources/>. Accessed 10 Sep 2024.
- Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European System for Cardiac Operative Risk Evaluation (EuroSCORE). *Eur J Cardiothorac Surg* 1999;16:9-13.
- Roques F, Michel P, Goldstone AR, Nashef SA. The logistic EuroSCORE. *Eur Heart J* 2003;24:881-2.
- Nashef SA, Roques F, Sharples LD, et al. EuroSCORE II. *Eur J Cardiothorac Surg* 2012;41:734-44.
- Chalmers J, Pullan M, Fabri B, et al. Validation of EuroSCORE II in a modern cohort of patients undergoing cardiac surgery. *Eur J Cardiothorac Surg* 2013;43:688-94.
- Zheng Z, Li Y, Zhang S, et al. The Chinese coronary artery bypass grafting registry study: how well does the EuroSCORE predict operative risk for Chinese population? *Eur J Cardiothorac Surg* 2009;35:54-8.
- Yap CH, Reid C, Yip M, et al. Validation of the EuroSCORE model in Australia. *Eur J Cardiothorac Surg* 2006;29:441-6.
- Grant SW, Hickey GL, Dimarakis I, et al. Performance of the EuroSCORE models in emergency cardiac surgery. *Circ Cardiovasc Qual Outcomes* 2013;6:178-85.
- 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;98:1568-72.
- Roques F, Nashef SA, Michel P, et al. Risk factors and outcome in European cardiac surgery: analysis of the EuroSCORE multinational database of 19030 patients. *Eur J Cardiothorac Surg* 1999;15:816-22.
- Gogbashian A, Sedrakyan A, Treasure T. EuroSCORE: a systematic review of international performance. *Eur J Cardiothorac Surg* 2004;25:695-700.
- Nashef SA, Roques F, Hammill BG, et al. Validation of European System for Cardiac Operative Risk Evaluation (EuroSCORE) in North American cardiac surgery. *Eur J Cardiothorac Surg* 2002;22:101-5.
- 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: a single-center study in Taiwan. *Acta Cardiol Sin* 2022;38:495-503.
- Shen L, Chen X, Gu J, Xue S. Validation of EuroSCORE II in Chinese patients undergoing coronary artery bypass surgery. *Heart Surg Forum* 2018;21:E036-9.
- Zhang GX, Wang C, Wang L, et al. Validation of EuroSCORE II in Chinese patients undergoing heart valve surgery. *Heart Lung Circ* 2013;22:606-11.
- Prince of Wales Hospital, The Chinese University of Hong Kong. Cardiac Surgery Report 2021-22. Available from: https://www.surgery.cuhk.edu.hk/cts/Cardiac_Surgery_Report_2021-22.pdf. Accessed 4 Sep 2024.
- Kurniawaty 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;20:491-6.
- Sembiring YE, Ginting A, Puruhito I, Budiono K. Validation of EuroSCORE II to predict mortality in post-cardiac surgery patients in East Java tertiary hospital. *Med J Indones* 2021;30:54-9.
- Atashi A, Amini S, Tashnizi MA, et al. 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;33:40-6.
- Zahara R, Soeharto DE, Widyantoro B, Sugisman, Herlambang B. Validation of EuroSCORE II scoring system on isolated CABG patient in Indonesia. *Egypt Heart J* 2023;75:86.
- 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;7:534.
- Riley RD, Ensor J, Snell KI, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;368:m441.
- Dendrite Clinical Systems. Databases for hospitals and clinics. Available from: <https://www.e-dendrite.com/hospital-systems>. Accessed 5 Sep 2024.
- Paparella D, Guida P, Di Eusanio G, et al. Risk stratification for in-hospital mortality after cardiac surgery: external

- validation of EuroSCORE II in a prospective regional registry. *Eur J Cardiothorac Surg* 2014;46:840-8.
27. Stavridis G, Panaretos D, Kadda O, Panagiotakos DB. Validation of the EuroSCORE II in a Greek cardiac surgical population: a prospective study. *Open Cardiovasc Med J* 2017;11:94-101.
 28. Nezic D, Spasic T, Micovic S, et al. Consecutive observational study to validate EuroSCORE II performances on a single-center, contemporary cardiac surgical cohort. *J Cardiothorac Vasc Anesth* 2016;30:345-51.
 29. Garcia-Valentin A, Mestres CA, Bernabeu E, et al. Validation and quality measurements for EuroSCORE and EuroSCORE II in the Spanish cardiac surgical population: a prospective, multicentre study. *Eur J Cardiothorac Surg* 2016;49:399-405.
 30. Koszta G, Sira G, Szatmári K, Farkas E, Szerafin T, Fülesdi B. Performance of EuroSCORE II in Hungary: a single-centre validation study. *Heart Lung Circ* 2014;23:1041-50.
 31. Barili F, Pacini D, Capo A, et al. Does EuroSCORE II perform better than its original versions? A multicentre validation study. *Eur Heart J* 2013;34:22-9.
 32. 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;93:1930-5.
 33. Luo HD, Teoh LK, Gaudino MF, Fremes S, Kofidis T. The Asian system for cardiac operative risk evaluation for predicting mortality after isolated coronary artery bypass graft surgery (ASCORE-C). *J Card Surg* 2020;35:2574-82.
 34. Ilcheva L, Risteski P, Tudorache I, et al. Beyond conventional operations: embracing the era of contemporary minimally invasive cardiac surgery. *J Clin Med* 2023;12:7210.