ESC European Society of Cardiology

European Heart Journal - Digital Health (2024) **5**, 702–711
https://doi.org/10.1093/ehjdh/ztae071

ORIGINAL ARTICLE
Artificial intelligence (machine learning, deep learning)

Validation of machine learning-based risk stratification scores for patients with acute coronary syndrome treated with percutaneous coronary intervention

Mitchel A. Molenaar [1,2,*], Jasper L. Selder[1,2], Amand F. Schmidt[1,2,3,4,5], Folkert W. Asselbergs[1,6], Jelle D. Nieuwendijk[1], Brigitte van Dalfsen[1], Mark J. Schuuring [1,7], Berto J. Bouma[1,2], Steven A.J. Chamuleau[1,2], and Niels J. Verouden[1,2,*]

[1]Department of Cardiology, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, The Netherlands; [2]Amsterdam Cardiovascular Sciences, Amsterdam University Medical Centers, Amsterdam, The Netherlands; [3]UCL British Heart Foundation Research Accelerator Centre, London, United Kingdom; [4]Department of Cardiology, Division Heart and Lungs, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX, Utrecht, The Netherlands; [5]Institute of Cardiovascular Science, Faculty of Population Health, University College London, London, United Kingdom; [6]Health Data Research UK and Institute of Health Informatics, University College London, London, United Kingdom; and [7]Department of Cardiology, Thorax Centrum Twente, Medisch Spectrum Twente, Koningsplein 1, 7512 KZ, Enschede, The Netherlands

Received 16 February 2024; revised 24 April 2024; accepted 13 August 2024; online publish-ahead-of-print 26 September 2024

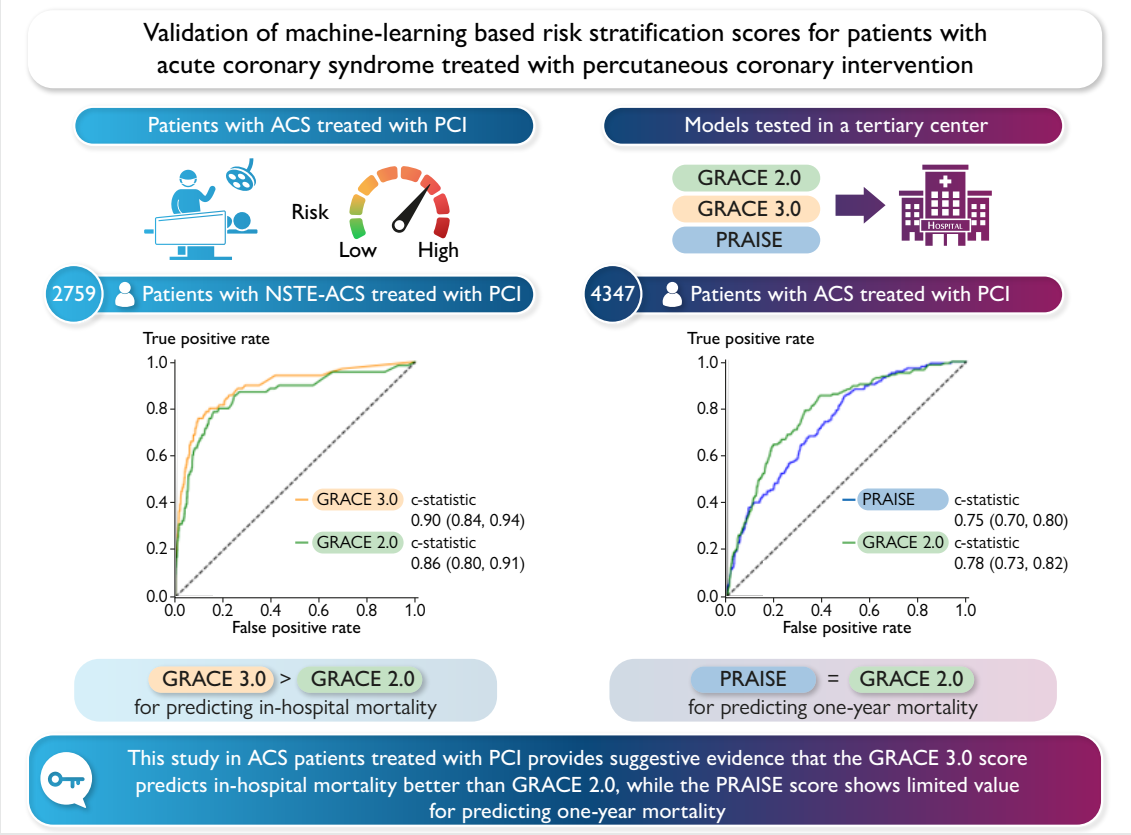
Aims This study aimed to validate the machine learning-based Global Registry of Acute Coronary Events (GRACE) 3.0 score and PRAISE (Prediction of Adverse Events following an Acute Coronary Syndrome) in patients with acute coronary syndrome (ACS) treated with percutaneous coronary intervention (PCI) for predicting mortality.

Methods and results Data of consecutive patients with ACS treated with PCI in a tertiary centre in the Netherlands between 2014 and 2021 were used for external validation. The GRACE 3.0 score for predicting in-hospital mortality was evaluated in 2759 patients with non-ST-elevation acute coronary syndrome (NSTE-ACS) treated with PCI. The PRAISE score for predicting one-year mortality was evaluated in 4347 patients with ACS treated with PCI. Both risk scores were compared with the GRACE 2.0 score. The GRACE 3.0 score showed excellent discrimination [c-statistic 0.90 (95% CI 0.84, 0.94)] for predicting in-hospital mortality, with well-calibrated predictions (calibration-in-the large [CIL] −0.19 [95% CI −0.45, 0.07]). The PRAISE score demonstrated moderate discrimination [c-statistic 0.75 (95% CI 0.70, 0.80)] and overestimated the one-year risk of mortality [CIL −0.56 (95% CI −0.73, −0.39)]. Decision curve analysis demonstrated that the GRACE 3.0 score offered improved risk prediction compared with the GRACE 2.0 score, while the PRAISE score did not.

Conclusion This study in ACS patients treated with PCI provides suggestive evidence that the GRACE 3.0 score effectively predicts in-hospital mortality beyond the GRACE 2.0 score. The PRAISE score demonstrated limited potential for predicting one-year mortality risk. Further external validation studies in larger cohorts including patients without PCI are warranted.

* Corresponding author. Mitchel A. Molenaar, Email: m.a.molenaar1@amsterdamumc.nl; Niels J. Verouden, Email: c.verouden@amsterdamumc.nl

Graphical Abstract



Keywords Acute Coronary Syndrome • Percutaneous Coronary Intervention • Mortality • Prognosis • Risk Assessment • Treatment Outcome

Introduction

Acute coronary syndrome (ACS) is defined as a range of conditions caused by sudden myocardial ischaemia. ACS affects more than 7 million people worldwide each year and approximately 8% of patients die within one year after admission.[1–4] To aid clinical decision making, risk assessment is performed in patients presenting with ACS.[5,6] Numerous risk scores have been developed to assess the risk of in-hospital and longer term mortality, of which the Global Registry of Acute Coronary Events (GRACE) scoring system is recommended by the European Society of Cardiology (ESC) guideline.[5,6] GRACE was a registry of patients presenting with ACS in the period 1999 to 2009 in a time before the widespread adoption of drug-eluting stents and contemporary medical therapy standards. Moreover, the generalizability of the GRACE 2.0 score remains limited, particularly in accurately predicting the probability of adverse events for the individual patient.[5,6] These limitations indicated a need for new and more personalized risk stratification tools.

In recent years, machine learning-based risk scores have enabled the identification of more complex patterns compared with traditional regression methods.[7,8] Machine learning has shown promising results for risk stratification in patients with ACS.[9,10] For example, the machine learning-based GRACE version 3.0 score[11] improved the prediction of in-hospital mortality among patients with non-ST-elevation acute coronary syndrome (NSTE-ACS) compared with the GRACE version 2.0 score.[11] The GRACE 3.0 score is calculated using nine variables of clinical presentation, which include the variables in the GRACE 2.0 score and sex. In external validation, the GRACE 3.0 score achieved a c-statistic of 0.91 (95% CI 0.89–0.92) in male patients and 0.87 (95% CI 0.84–0.89) in female patients.[11] In addition, the PRAISE risk score,[12] a machine learning-based risk score, was trained on ACS patients treated with percutaneous coronary intervention (PCI) and integrates 25 clinical, anatomical, and procedural features. The PRAISE risk score achieved a c-statistic of 0.92 (95% CI 0.90–0.93) in an external validation cohort. This score has demonstrated accurate discrimination between patients who are likely to die within one year and those who are not.[12]

PCI is a widely utilized treatment in patients with ACS.[13] Validation of these machine learning-based tools is essential for selecting the most effective risk stratification tools, needed to better understand the patient's prognosis after PCI, identify high-risk individuals, and guide clinicians in their decisions. An independent validation study for the GRACE 3.0 score and PRAISE score in patients with ACS treated with PCI at a Dutch tertiary centre is lacking. Therefore, the aim of this study was to validate the GRACE 3.0 score and PRAISE in patients with ACS treated with PCI for predicting mortality.

Methods

Study design and patient population

In this retrospective cohort study, patients diagnosed by the treating physician with ACS were consecutively selected from electronic health records of the Amsterdam University Medical Center, location VUmc, between 2015 and 2021. Patients were eligible if they were 18 years or older and underwent a PCI procedure. Patients were managed according to the ESC guidelines.[13–16] This study complies with the principles in the Declaration of Helsinki and received approval by the local human ethical review board. The study met the criteria for a waiver of the informed consent requirements.

Outcome

The primary outcomes were in-hospital mortality and one-year mortality.

Data collection

Baseline, treatment, and mortality data of patients who underwent PCI for ACS were collected from pseudonymized electronic health records and stored in a registry. Mortality data was verified using national registry data. The following variables were collected for each patient for calculation of the GRACE 2.0 score: age, heart rate, systolic blood pressure, Killip class, creatinine concentration, cardiac arrest, presence of ST-segment deviation, and troponin elevation. Calculation of the GRACE 3.0 score required the same variables, with addition of sex as variable. For calculation of the PRAISE score, 16 clinical variables (age, sex, diabetes, hypertension, hyperlipidemia, peripheral artery disease, estimated glomerular filtration rate [eGFR] using the Modification of Diet in Renal Disease [MDRD] equation,[17] previous myocardial infarction, previous PCI, previous coronary artery bypass grafting [CABG], previous stroke, previous bleeding, malignancy, STEMI presentation, haemoglobin, and left ventricular ejection fraction [LVEF]) were collected. The PRAISE score was calculated based on the collection of two procedural variables (vascular access and PCI with drug-eluting stent), two angiographic variables (multivessel disease and complete revascularization), and five therapeutic variables (treatment with β blockers, angiotensin-converting enzyme inhibitors or angiotensin-receptor blockers, statins, oral anticoagulation, and proton-pump inhibitors). The definitions of these variables are shown in Supplementary material online, *Table S1* and were in accordance with the definitions used in the studies that developed the machine learning-based risk scores.[11,12,18] The GRACE 2.0 (in-hospital and one-year mortality), GRACE 3.0 (in-hospital mortality), and PRAISE scores (one-year mortality) were calculated for each patient using the online web pages (https://www.outcomes-umassmed.org/grace/acs_risk2/index.html,[4] https://www.grace-3.com,[11,18] and https://praise.hpc4ai.it[12]).

Missing data

Missing values of variables in the risk scores were imputed by multiple imputation by chained equation (MICE) in accordance with Wenzl *et al.*[11] MICE is an iterative process that imputes missing values using modelling of the other variables in the dataset. In this study, each variable was imputed 20 times (20 iterations), and this process was repeated to create 20 imputed datasets.[19] Predictive mean matching, proportional odds, and polytomous logistic regression were applied for continuous data, ordered categorical data, and unordered categorical data, respectively. The outcome variables in-hospital mortality, six-month mortality, and one-year mortality were used as independent variables in the imputation process. Complete case analysis was performed to evaluate the effect of imputation on the results.

Evaluation of risk scores

The GRACE 3.0 score was evaluated in patients with NSTE-ACS treated with PCI for predicting in-hospital mortality.[11] The PRAISE score was evaluated in patients with ACS treated with PCI for predicting one-year mortality. In line with the PRAISE development cohort score, patients who died during hospitalization were excluded.[20,21] Additional analyses were conducted to investigate the performance of the GRACE 3.0 score and PRAISE score in patients with ACS and NSTE-ACS treated with PCI, in both male and female subgroups.

The discriminative performance of the machine learning-based risk scores was evaluated using concordance statistic (c-statistic).[22] The discriminative ability was classified as poor (c-statistic < 0.70), moderate (c-statistic 0.70–0.80), good (c-statistic 0.80–0.90), or excellent (c-statistic ≥ 0.90).[22,23] Rubin's Rules were used to combine the c-statistics from the imputed datasets and obtain an estimate with a 95% confidence interval.[24]

The calibration (agreement between predicted and actual observed risk), was assessed using calibration plots, calibration-in-the-large (CIL), and the calibration slope (CS).[25–27] CIL compares the average predicted risk with the observed risk, which is 0.00 in a perfectly calibrated risk score. A CIL lower than 0.00 indicates overestimation and a CIL greater than 0.00 indicates underestimation. CS evaluates the spread of the predicted risks, which is equal to 1.00 in a perfectly calibrated risk score. A CS lower than 1.00 indicates that the predicted risks are too high for patients at high risk and too low for patients at low risk. A CS greater than 1.00 indicates that the predicted risks are too low for high-risk patients and too high for low-risk patients.[26,28] The calibration of the GRACE 3.0 score is particularly clinically important around the 3% threshold, which guides the decision on early vs. a delayed invasive treatment in patients with ACS.[29]

Decision curve analysis was performed to evaluate the clinical utility of the machine learning-based risk scores for identifying high-risk and low-risk ACS patients of mortality after PCI.[30,31] In this analysis, the net benefit (a weighted combination of true and false positives, determined by a threshold probability) of the risk score is plotted against a range of threshold probabilities. The threshold probabilities can be interpreted as the percentage at which a clinician would opt for close monitoring.[32] This could be interpreted as the percentage of patients who require close monitoring needed to protect one patient from mortality. For example, setting a threshold at 0.05 (5%) means that to protect one patient from mortality, 20 patients need close monitoring. Similarly, for a threshold probability at 0.1 (10%), ten patients need close monitoring to prevent one patient from mortality. The net benefit across all thresholds of the machine learning risk scores is compared with the GRACE 2.0 score and close monitoring in all patients ('Always act'). The risk score with the highest net benefit at a certain threshold probability has the best trade-off between true positives and false positives and is the most clinically useful at that specific threshold probability. The maximum net-benefit is equal to the incidence of mortality, which is the case when all patients who will die are identified by the risk score without any false positives. Decision curve analysis combines both discrimination and calibration, which offers a valuable method to compare the performance of risk scores.

Statistical analysis

Descriptive statistics were presented as median with interquartile range (IQR) to facilitate comparison with the cohorts upon which the risk scores (GRACE 2.0,[4] GRACE 3.0,[11] PRAISE[12]) were developed. Nominal or ordinal data were presented as numbers with percentages. Data processing and statistical analysis were performed in Python version 3.8. and R version 4.1.3. (R Core Team, Vienna, Austria). The c-statistics of the GRACE 3.0 and PRAISE scores were compared with the GRACE 2.0 score on significance using the DeLong's test for paired receiver operating characteristics (ROC) curves.[33] Rubin's Rules were used to aggregate the *P*-values obtained from the imputed datasets into a single estimate. A *P*-value less than 0.05 was considered statistically significant.

This study followed the TRIPOD (transparent reporting of multivariable prediction model for individual prognosis or diagnosis)[34] statement and met all CODE-EHR minimum framework standards for the use of healthcare data for clinical research.[35]

Results

Study population

A total of 10411 patients who underwent CAG or PCI were screened for inclusion. After excluding patients without ACS or PCI, 4471 patients were identified. Of these patients, 2759 patients (62%) presented with NSTE-ACS, in whom the GRACE 3.0 risk score was evaluated. A total of 124 patients with ACS treated with PCI (2.7%) died during

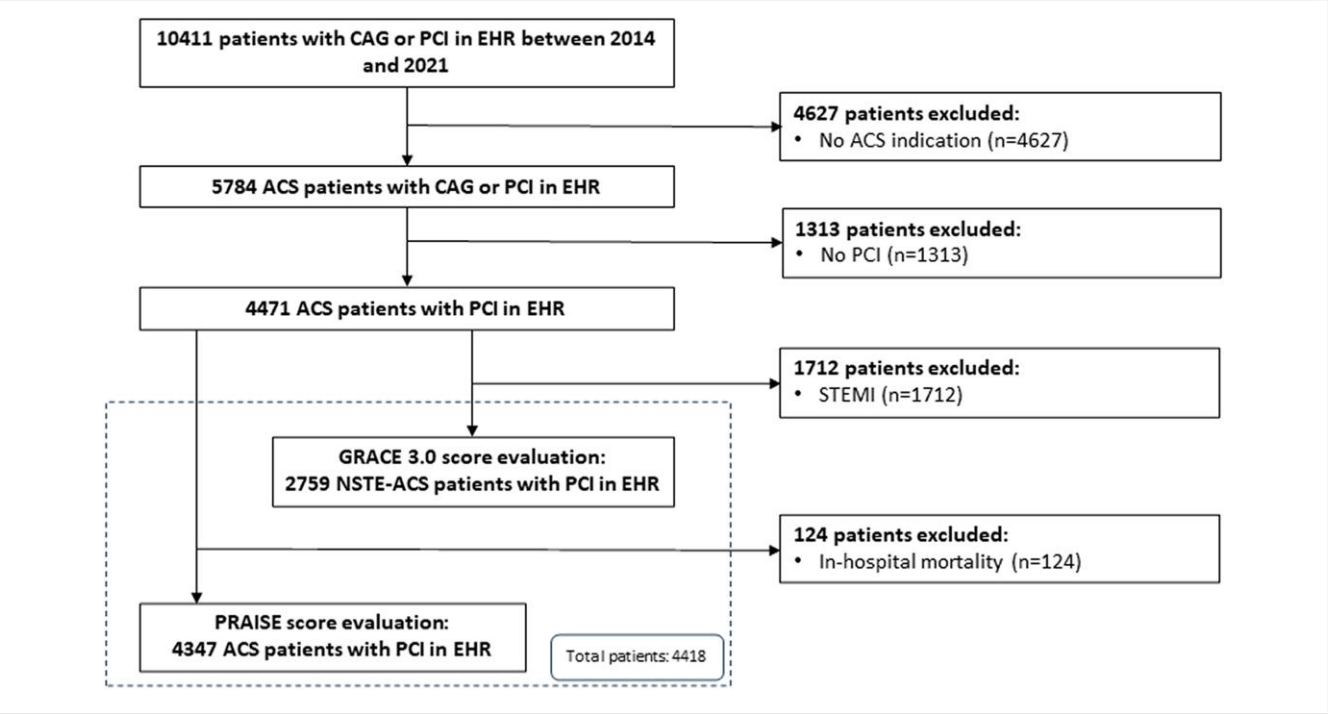


Figure 1 Inclusion flowchart. ACS = acute coronary syndrome; CAG = coronary angiography; EHR = electronic health records; NSTE-ACS = with non-ST-elevation acute coronary syndrome; PCI = percutaneous coronary intervention.

hospitalization, which resulted in 4347 patients with ACS in whom the PRAISE risk score was evaluated. A total number of 4418 patients with ACS treated with PCI were included in this study. The flow chart is depicted in *Figure 1*. In 24% of the patients, the data were not complete to calculate the GRACE 2.0 score. A total of 6.3% of the data were missing, as presented in Supplementary material online, *Table S2*. Ejection fraction had the highest rate of missing data, which was missing in 61% of the patients. No data were missing for the outcomes in-hospital and one-year mortality.

A total of 3178 (72%) of patients in this study were male, and the median age was 66 years. The baseline characteristics are presented in *Table 1*. There were differences in the patient characteristics with the original GRACE 2.0,[4] GRACE 3.0,[11] and PRAISE[12] cohort, as shown in Supplementary material online, *Table S3*. Patients in the GRACE 3.0 evaluation cohort had higher rates of cardiac arrest at admission and cardiogenic shock (Killip IV class) at presentation compared with patients in the GRACE 2.0 and 3.0 development cohort. PCI was performed in 35% of the females and 46% of the males in the original GRACE 3.0 cohort. The PRAISE development cohort had a higher prevalence of dyslipidemia, STEMI, and a higher median eGFR and ejection fraction, compared with patients in this study. All patients in the original PRAISE cohort underwent PCI.

Mortality

In the GRACE 3.0 evaluation cohort 71 patients (2.5%) died in the hospital, and 156 patients (5.6%) died within one year after admission, as shown in *Table 1*. In the PRAISE evaluation cohort, 145 patients (3.3%) died within one year after admission.

Performance: GRACE 3.0 score

The c-statistic of the GRACE 3.0 score for predicting in-hospital mortality was 0.90 (95% CI 0.84, 0.94), which exceeded that of the GRACE 2.0 score [0.86 (95% CI 0.80, 0.91), $P = 0.002$]. The ROC curves are shown in *Figure 2*. While the GRACE 2.0 score [CIL −0.31 (95% CI −0.56, −0.06)] showed, on average, a slight overestimation of the risk of mortality, GRACE 3.0 score was well-calibrated [CIL −0.19 (95% CI −0.45, 0.07), *Figure 3*].

The CS of the GRACE 3.0 score [CS 0.96 (95% CI 0.81, 1.12)] and GRACE 2.0 score [CS 1.08 (95% CI 0.90, 1.25)] were nearly perfect, which indicates that the risk scores are well-calibrated at extreme high and low predictions. The calibration plots are shown in *Figure 3*. Due to missing data, 1868 patients with NSTE-ACS treated with PCI were included in complete case analysis. Complete case analysis is shown in Supplementary material online, *Table S4*. In complete case analysis, 61 patients with NSTE-ACS treated with PCI (3%) died in the hospital. Complete case analysis yielded results similar to the analysis including all patients, except for the GRACE 2.0 [CIL −0.26 (95% CI −0.54, 0.01)] calibration, which was slightly better compared with the analysis of all patients.

Decision curve analysis showed that the GRACE 3.0 score was more effective in selecting patients at high- and low-risk of mortality for decision thresholds between 0% and 30%, compared with monitoring all patients closely and the GRACE 2.0 score. The decision curve analysis is shown in *Figure 4*.

In additional analyses, the c-statistic of the GRACE 3.0 risk score in male patients with NSTE-ACS treated with PCI ($n = 1982$) was 0.89 (95% CI 0.83, 0.95), and 0.93 (95% CI 0.79, 0.98) in female patients with NSTE-ACS treated with PCI ($n = 748$). These results are shown in Supplementary material online, *Table S5*.

Performance: PRAISE score

The c-statistic of the PRAISE score for predicting one-year mortality was 0.75 (95% CI 0.70, 0.80), which was lower but not significantly different from that of the GRACE 2.0 score [0.78 (95% CI 0.73,

Table 1 Baseline characteristics and adverse outcomes

Characteristics	PRAISE validation cohort (ACS, *n* = 4347)	GRACE 3.0 validation cohort (NSTE-ACS, *n* = 2759)
Age (years), median [Q1,Q3]	66.00 [57.00, 74.00]	68.00 [58.00, 75.00]
Male, *n* (%) or %	3127 (72.7)	1982 (72.6)
BMI (kg/m2), median [Q1,Q3]	26.50 [24.21, 29.39]	26.60 [24.30, 29.70]
Risk factors		
Hypertension, *n* (%) or %	2234 (53.5)	1514 (57.3)
Diabetes, *n* (%) or %	910 (21.3)	672 (24.8)
Dyslipidemia, *n* (%) or %	1252 (30.1)	857 (32.6)
Current or former smoker, *n* (%) or %	1589 (38.1)	804 (30.4)
Medical history		
Myocardial infarction, *n* (%) or %	719 (16.5)	571 (20.7)
Percutaneous coronary intervention, *n* (%) or %	834 (19.2)	651 (23.6)
Coronary artery bypass graft, *n* (%) or %	220 (5.1)	196 (7.1)
Peripheral artery disease, *n* (%) or %	157 (3.9)	124 (4.8)
Stroke or transient ischaemic attack, *n* (%) or %	178 (4.1)	138 (5.0)
Bleedings, *n* (%) or %	63 (1.4)	46 (1.7)
Clinical presentation		
STEMI, *n* (%) or %	1659 (38.2)	0
Unstable angina or non-STEMI, *n* (%) or %	2688 (61.8)	2759 (100)
Haemoglobin at admission (mg/dL), median [Q1,Q3]	13.86 [12.57, 14.82]	13.70 [12.25, 14.82]
Heart rate (bpm), median [Q1,Q3]	70.00 [60.00, 81.00]	69.00 [60.00, 80.00]
Systolic blood pressure (mmHg), median [Q1,Q3]	127.00 [112.00, 143.00]	130.00 [114.00, 146.00]
Creatinine (mg/dL), median [Q1,Q3]	0.93 [0.79, 1.11]	0.95 [0.81, 1.14]
ST-segment deviation, *n* (%) or %	2396 (57.7)	768 (30.0)
Abnormal cardiac enzymes, *n* (%) or %	2559 (60.5)	1184 (44.6)
Cardiac arrest at admission, *n* (%) or %	169 (4.0)	107 (4.0)
Killip class, *n* (%) or %		
I	3293 (81.0)	2090 (81.6)
II	630 (15.5)	354 (13.8)
III	114 (2.8)	87 (3.4)
IV	28 (0.7)	31 (1.2)
Ejection fraction (%), median [Q1,Q3]	47.00 [35.00, 62.00]	47.00 [35.00, 62.00]
eGFR (mL/min/1,73 m2), median [Q1,Q3]	63.44 [50.57, 80.76]	61.69 [48.96, 77.58]
Anatomy and procedural data		
Multivessel disease, *n* (%) or %	1981 (45.6)	1395 (50.6)
Percutaneous coronary intervention with DES implantation, *n* (%) or %	4347 (100.0)	2759 (100.0)
Complete revascularization, *n* (%) or %	3309 (76.1)	2142 (77.6)
Intervention		
Percutaneous coronary intervention, *n* (%) or %	4347 (100)	2759 (100)
Medical therapy at discharge		
Antiplatelet therapy, *n* (%) or %	3662 (84.2)	2345 (85.0)
Beta blockers, *n* (%) or %	2229 (51.3)	1214 (44.0)
ACE-inhibitor/ARB, *n* (%) or %	1870 (43.0)	1026 (37.2)
Statins, *n* (%) or %	2527 (58.1)	1270 (46.0)
Proton-pump inhibitor, *n* (%) or %	2190 (50.4)	1161 (42.1)
Oral anticoagulation, *n* (%) or %	466 (10.7)	309 (11.2)
Adverse outcomes		
In-hospital mortality, *n* (%) or %	0 (0)	71 (2.5)
One-year mortality, *n* (%) or %	145 (3.3)	156 (5.6)

Values are *n* (%), or median [interquartile range]. ACE, angiotensin-converting enzyme; ACS, acute coronary syndrome; ARB, angiotensin receptor blockers; BMI, body mass index; DES, drug-eluting stent; eGFR, estimated glomerular filtration rate [using the Modification of Diet in Renal Disease (MDRD) equation]; NSTE-ACS, non-ST-elevation acute coronary syndrome; STEMI, ST-elevation myocardial infarction.

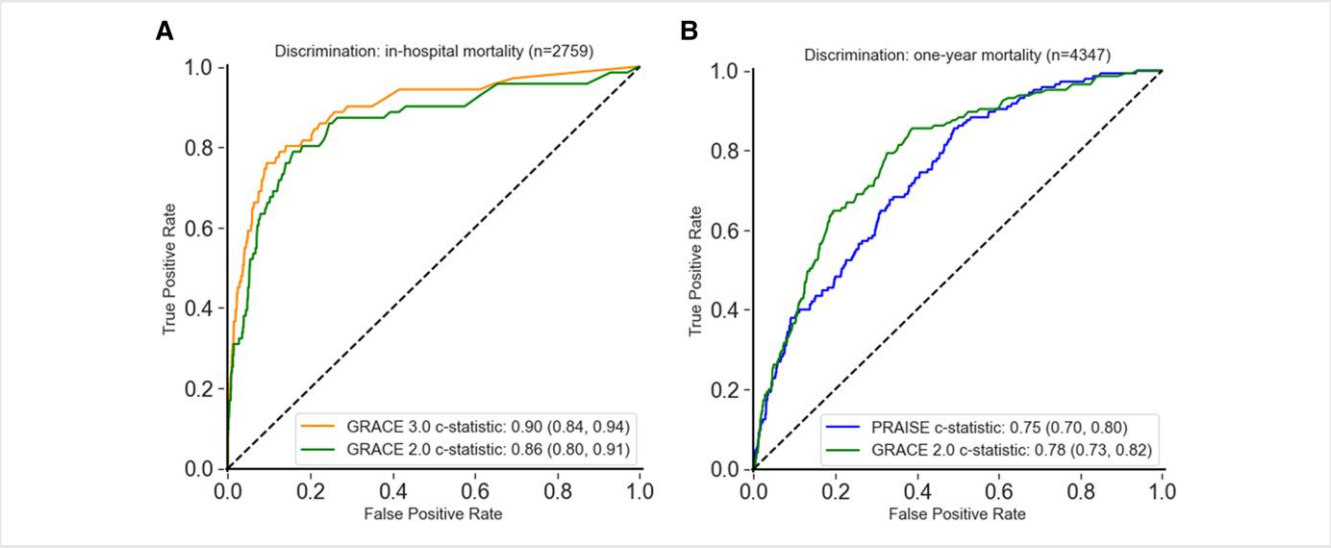


Figure 2 Discriminative performance of the GRACE 3.0 risk score, PRAISE risk score, and GRACE 2.0 risk score for predicting mortality. Legend: The ROC curves of the GRACE 3.0 and GRACE 2.0 scores for predicting in-hospital mortality (Plot A), as well as the PRAISE score and GRACE 2.0 score for predicting one-year mortality (Plot B), are depicted. The c-statistics are reported with a 95% confidence interval. GRACE = The Global Registry of Acute Coronary Events; PRAISE = Prediction of Adverse Events following an Acute Coronary Syndrome.

0.82), $P = 0.2$], as shown in *Figure 2*. The PRAISE score [CIL −0.56 (95% CI −0.73, −0.39)] showed near-perfect calibration for low mortality probabilities (0–5%). For predicted probabilities beyond 5%, the PRAISE score overestimated the risk of mortality. The GRACE 2.0 risk score overestimated the risk of mortality across the entire range of predicted mortality probabilities [CIL −1.12 (95% CI −1.30, −0.95)].

Decision curve analysis of the PRAISE score did not show any improvement in risk prediction compared with the GRACE 2.0 score, as depicted in *Figure 4*.

In complete case analysis only 1174 ACS patients treated with PCI (27% of the patients) were included. Of these patients, a total of 63 (5%) died within one year after presentation. The c-statistic of the PRAISE score was 0.79 (95% CI 0.71, 0.85) and the model was well-calibrated [CIL 0.05 (−0.21, 0.32)].

In additional analyses, the c-statistic of the PRAISE risk score in male ACS patients treated with PCI ($n = 3127$) was 0.77 (95% CI 0.71, 0.82), and 0.71 (95% CI 0.62, 0.80) in female ACS patients treated with PCI ($n = 1177$). These results are shown in Supplementary material online, *Table S6*.

Discussion

In this study, two machine learning-based risk scores (GRACE 3.0 and PRAISE) were validated in a population of patients with ACS who were treated with PCI in a tertiary centre. The GRACE 3.0 risk score, evaluated in 2759 patients with NSTE-ACS treated with PCI, showed excellent discriminative performance for predicting in-hospital mortality. The GRACE 3.0 risk score was more clinically useful as a risk prediction tool compared with the GRACE 2.0 score, as shown by the decision curve analysis. The PRAISE score, evaluated in 4347 ACS patients treated with PCI, showed moderate discrimination for predicting one-year mortality. The PRAISE score overestimated the risk of one-year mortality for patients with ACS treated with PCI with a predicted risk greater than 5% and did not provide a significant benefit over the GRACE 2.0 score.

The GRACE 3.0 score was developed using prospective data from 386,591 patients with NSTE-ACS in England, Wales, and Northern Ireland. Wenzl *et al.* demonstrated comparable discriminatory performance to that observed in this study during validation in cohorts of external centres.[11,36] The high discriminative ability can be explained by the machine learning model (XGBoost) that uses multiple decision trees to achieve optimal classification. In addition to the GRACE 2.0 score, the GRACE 3.0 score incorporates sex by utilizing separate machine learning models for male and female patients, which provides a more personalized sex-specific output. Previous studies have shown that the GRACE 2.0 score has suboptimal agreement between the predicted and observed risk of in-hospital mortality and one-year mortality.[6,11,37] The results of our study provide suggestive evidence that the GRACE 3.0 score could be a valuable tool for clinicians in assessing in-hospital mortality risk.

The PRAISE score was developed and validated by Ascenzo *et al.*,[12] who found a high discriminatory performance [c-statistic 0.92 (95% CI 0.90, 0.93)] for this score in an external cohort. Differences in the patient population between our study and the study by Ascenzo *et al.* may explain the limited ability of the risk score to generalize to a new population. In particular, the median eGFR and ejection fraction were higher in the patients of the study of Ascenzo *et al.*, which are important predictive factors of mortality in the PRAISE score.[12] Shi *et al.*[38] validated the PRAISE score in an Asian population (6412 ACS patients treated with PCI) and demonstrated that the PRAISE had a slightly greater net benefit compared with the GRACE 2.0 score. In line with Shi *et al.*, the results of our study suggest that retraining of the PRAISE model on our dataset is required to improve the prediction of mortality for this risk score.

The PRAISE score was calculated using 25 variables, including anatomical and procedural data. Anatomical and procedural data have been previously incorporated into mortality prediction scores,[39] such as vascular access site,[40] number of diseased coronary arteries,[41–43] and location of significant lesions.[43] The impact of anatomical and procedural data in a machine learning model for risk prediction has been investigated in only a limited number of studies. Zack *et al.*[44] demonstrated that their model outperformed logistic regression (c-statistic 0.88 vs. c-statistic 0.81) for predicting long term (180-day) mortality using 410 variables, including angiographic and interventional details. In addition, Mori *et al.*[45] trained a XGBoost model on angiographic

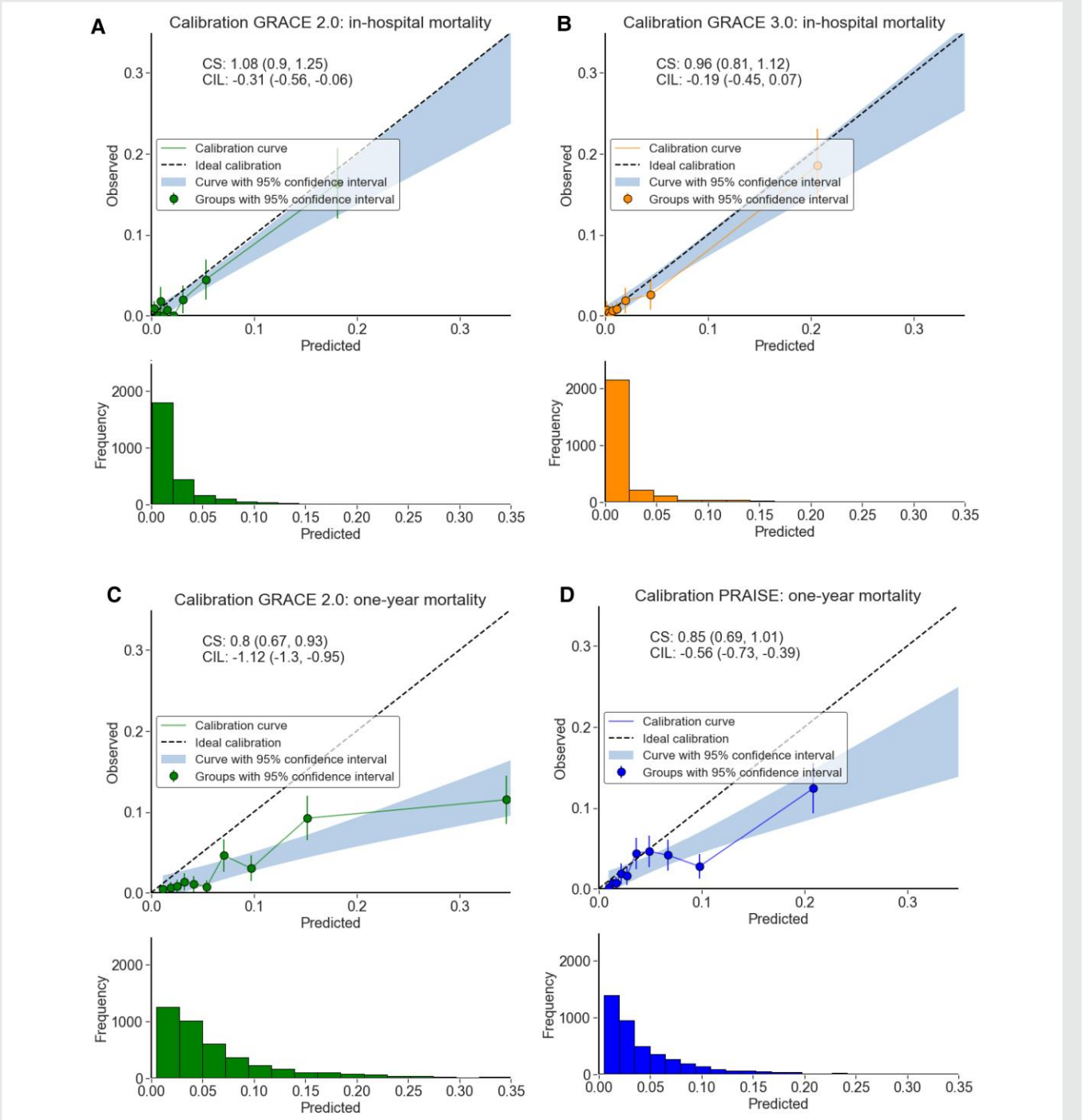


Figure 3 Calibration plots of the GRACE 3.0 risk score, PRAISE risk score, and GRACE 2.0 risk score. Legend: The calibration (agreement between predicted and actual observed risk) and frequency (number of patients falling within each predicted risk category) plots are depicted for the GRACE 2.0 (in-hospital mortality, Plot A), GRACE 3.0 (in-hospital mortality, Plot B), GRACE 2.0 (one-year mortality, Plot C), and PRAISE (one-year mortality, Plot D). Numbers are reported with 95% confidence intervals. CIL = calibration-in-the-large; CS = calibration slope.

data of 378 572 patients treated with CABG to predict multiple outcomes and demonstrated improved risk stratification compared with machine learning models trained on clinical data alone. These findings suggest that further research is needed to establish a set of important anatomical and procedural features to further explore the potential of machine learning for optimizing risk stratification in ACS patients treated non-conservatively. It is important to note that calculating a risk score involving a large number of variables via calculators can increase clinicians' workload. Automatic calculation and integration in electronic health records can help clinicians predict the risk for individual patients.

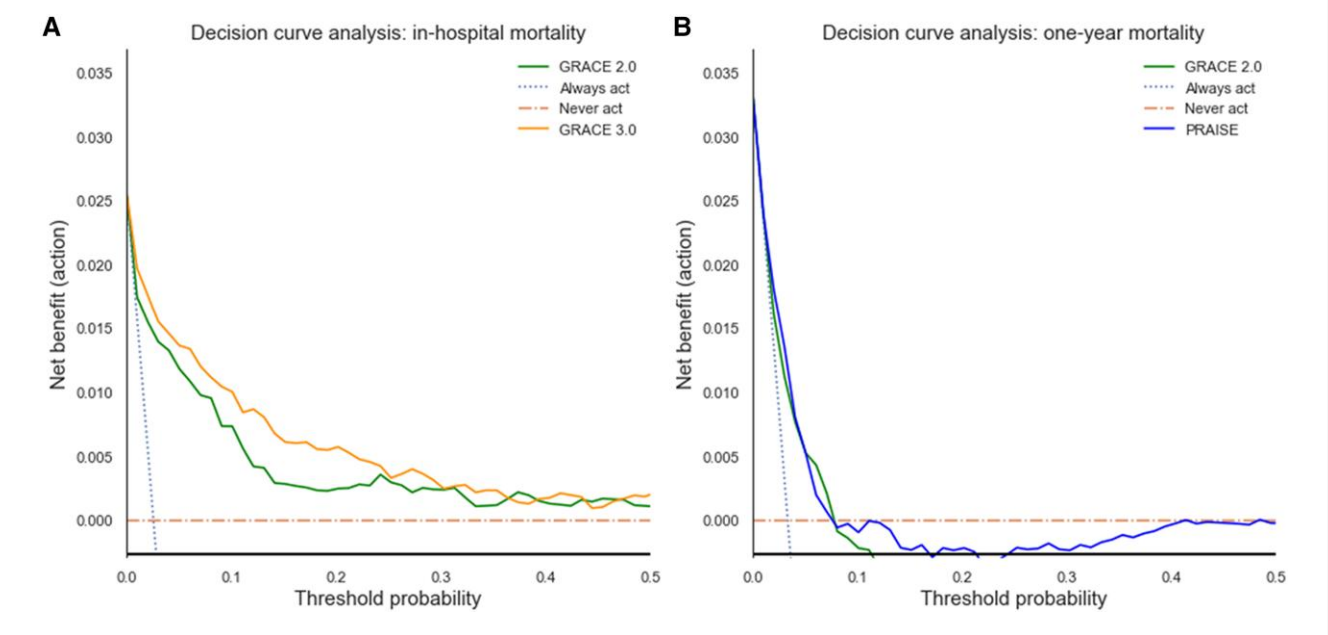


Figure 4 Decision curve analysis for the GRACE 3.0 risk score, GRACE 2.0 risk score, and the PRAISE risk score. Decision curve analysis of the GRACE 3.0 and GRACE 2.0 (Plot A), and PRAISE and GRACE 2.0 (Plot B). The *y*-axis represents the net benefit (a weighted combination of true and false positives, determined by the threshold probability) and the *x*-axis represents the threshold probability. A higher net benefit at a certain threshold probability indicates that the risk score is more clinically useful at that threshold probability for predicting mortality. Plot A demonstrates that the GRACE 3.0 risk score was more effective in selecting patients who require close monitoring compared with monitoring all patients closely ('Always act') and compared with the GRACE 2.0 risk score. Plot B demonstrates that the PRAISE score did not provide a significant benefit over the GRACE 2.0 score across different threshold probabilities.

The number of clinical prediction models is growing exponentially. However, only a fraction of these models is validated in an external cohort. Wessler *et al.* demonstrated that only 42% of the cardiovascular clinical prediction models were validated in an external cohort.[46] External validation is needed to determine the generalizability of a model, especially in models that are trained on small datasets. External validation is a prerequisite before implementing a model in clinical practice.[47] In the context of patients with ACS, a risk model can have implications for treatment decisions by, for example, selecting the right patients who may require close monitoring, aggressive management of risk factors, extended hospital stay, additional interventions, or for prioritizing treatment. Therefore, validation of these machine learning models in diverse clinical settings is important, to which our study contributes.

Several remarks can be made about this study. First, patients were retrospectively included in our study, which resulted in missing data. Complete case analysis and all-case analysis did not show significant differences in the performance of the GRACE 3.0 score. The PRAISE score overestimated the risk of one-year mortality in all-case analysis, while the risk score demonstrated nearly perfect calibration in complete case analysis. These findings may be explained by the substantial number of patients who had missing data, with variables missing in up to 61% of the patients. Excluding patients with missing data from analysis resulted in a different risk distribution compared with the entire population, with higher rates of mortality (5.0% vs. 3.3%). Second, the GRACE 3.0 score appeared to be more clinically useful compared with the GRACE 2.0 score. The GRACE 3.0 score was originally developed on a population of NSTE-ACS patients in which 36% females and 64% males underwent PCI. It was not specifically tailored for patients who underwent PCI, as in our study. It is possible that we have primarily selected patients with a lower clinician-assessed risk of poor outcomes compared with the original GRACE 3.0 cohort of Wenzl *et al.* due to the risk-treatment paradox,[6,48,49] in which clinicians might be hesitant to perform an invasive procedure in patients with a high risk of adverse outcomes. Further assessment is needed in an untreated ACS population to confirm the generalizability of the GRACE 3.0 score for risk stratification. Third, the available data limited our ability to investigate the performance of the PRAISE risk score (recurrent ACS, major bleeding) beyond all-cause mortality and to evaluate other risk scores (e.g. PARIS and PRECISE-DAPT risk scores). Fourth, the combination of the single-centre design and the low number of in-hospital mortality events necessitates larger datasets to conclusively assess the performance of the evaluated risk scores. Fifth, a strength of the study is that the machine learning-based risk scores were compared with the GRACE 2.0 score, which is recommended by the ESC guidelines for risk assessment in patients with ACS.[29]

Conclusion

In conclusion, this study in ACS patients treated with PCI provides suggestive evidence that the GRACE 3.0 score effectively predicts in-hospital mortality beyond the GRACE 2.0 score. In our dataset, the PRAISE score showed limited potential for predicting one-year mortality and did not enhance clinical decision making compared with the GRACE 2.0 score. Further external validation of GRACE 3.0 and PRAISE in larger prospective multi-centre patient cohorts including patients without PCI is warranted.

Supplementary material

Supplementary material is available at *European Heart Journal – Digital Health*.

Acknowledgements

The authors thank Matthijs Schakel for his support in this project.

Funding

The study was supported by the internal funding institution.

Conflict of interest: AFS has received funding from New Amsterdam Pharma for unrelated projects. FWA is supported by UCL Hospitals NIHR Biomedical Research Centre, EU Horizon (AI4HF 101080430 and DataTools4Heart 101057849), Dutch Research Council (MyDigiTwin 628.011.213).

Data availability

The original contributions presented in the study are included in the article/Supplementary material online. Further inquiries can be directed to the corresponding author.

References

1. Reed GW, Rossi JE, Cannon CP. Acute myocardial infarction. *The Lancet* 2017;**389**:197–210.
2. Steg PG, Goldberg RJ, Gore JM, Fox KAA, Eagle KA, Flather MD, *et al.* Baseline characteristics, management practices, and in-hospital outcomes of patients hospitalized with acute coronary syndromes in the global registry of acute coronary events (GRACE) **further information about the project, along with a complete list of the study participants, can be found at www.outcomes.org/grace. *Am J Cardiol* 2002;**90**:358–363.
3. Bhatt DL, Lopes RD, Harrington RA. Diagnosis and treatment of acute coronary syndromes: a review. *JAMA* 2022;**327**:662–675.
4. Fox KAA, FitzGerald G, Puymirat E, Huang W, Carruthers K, Simon T, *et al.* Should patients with acute coronary disease be stratified for management according to their risk? Derivation, external validation and outcomes using the updated GRACE risk score. *BMJ Open* 2014;**4**:e004425.
5. Ono M, Kawashima H, Hara H, Gamal A, Wang R, Gao C, *et al.* External validation of the GRACE risk score 2.0 in the contemporary all-comers GLOBAL LEADERS trial. *Catheter Cardiovasc Interv* 2021;**98**:E513–E522.
6. van der Sangen NMR, Azzahhafi J, Yin DRPPCP, Peper J, Rayhi S, Walhout RJ, *et al.* External validation of the GRACE risk score and the risk–treatment paradox in patients with acute coronary syndrome. *Open Heart* 2022;**9**:e001984.
7. Oikonomou EK, Williams MC, Kotanidis CP, Desai MY, Marwan M, Antonopoulos AS, *et al.* A novel machine learning-derived radiotranscriptomic signature of perivascular fat improves cardiac risk prediction using coronary CT angiography. *Eur Heart J* 2019;**40**:3529–3543.
8. Molenaar MA, Bouma BJ, Asselbergs FW, Verouden NJ, Selder JL, Chamuleau SAJ, *et al.* Explainable machine learning using echocardiography to improve risk prediction in patients with chronic coronary syndrome. *Eur Heart J Digit Health* 2024;**5**:170–182.
9. Molenaar MA, Selder JL, Nicolas J, Claessen BE, Mehran R, Bescós JO, *et al.* Current state and future perspectives of artificial intelligence for automated coronary angiography imaging analysis in patients with ischemic heart disease. *Curr Cardiol Rep* 2022;**24**:365–376.
10. Gill SK, Karwath A, Uh H-W, Cardoso VR, Gu Z, Barsky A, *et al.* Artificial intelligence to enhance clinical value across the spectrum of cardiovascular healthcare. *Eur Heart J* 2023;**44**:713–725.
11. Wenzl FA, Kraler S, Ambler G, Weston C, Herzog SA, Räber L, *et al.* Sex-specific evaluation and redevelopment of the GRACE score in non-ST-segment elevation acute coronary syndromes in populations from the UK and Switzerland: a multinational analysis with external cohort validation. *The Lancet* 2022;**400**:744–756.
12. D'Ascenzo F, De Filippo O, Gallone G, Mittone G, Deriu MA, Iannaccone M, *et al.* Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): a modelling study of pooled datasets. *The Lancet* 2021;**397**:199–207.
13. Steg P, James SK, Atar D, Badano LP, Lundqvist CB, Borger MA, *et al.* ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force on the management of ST-segment elevation acute myocardial infarction of the European Society of Cardiology (ESC). *Eur Heart J* 2012;**33**:2569–2619.
14. Windecker S, Kolh P, Alfonso F, Collet J-P, Cremer J, Falk V, *et al.* 2014 ESC/EACTS guidelines on myocardial revascularization: the task force on myocardial revascularization of the European Society of Cardiology (ESC) and the European association for cardio-thoracic surgery (EACTS)Developed with the special contribution of the European association of percutaneous cardiovascular interventions (EAPCI). *Eur Heart J* 2014;**35**:2541–2619.
15. Roffi M, Patrono C, Collet J-P, Mueller C, Valgimigli M, Andreotti F, *et al.* 2015 ESC guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: task force for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation of the European Society of Cardiology (ESC). *Eur Heart J* 2016;**37**:267–315.
16. Ibanez B, James S, Agewall S, Antunes MJ, Bucciarelli-Ducci C, Bueno H, *et al.* 2017 ESC guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: the task force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC). *Eur Heart J* 2018;**39**:119–177.
17. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. *Ann Intern Med* 1999;**130**:461–470.
18. Wenzl FA, Lüscher TF. Application of a sex-specific GRACE score in practice – authors' reply. *The Lancet* 2023;**401**:23.
19. van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw* 2011;**45**:1–67.
20. D'Ascenzo F, Biondi-Zoccai G, Moretti C, Bollati M, Omedè P, Sciuto F, *et al.* TIMI, GRACE and alternative risk scores in acute coronary syndromes: a meta-analysis of 40 derivation studies on 216,552 patients and of 42 validation studies on 31,625 patients. *Contemp Clin Trials* 2012;**33**:507–514.
21. De Filippo O, D'Ascenzo F, Raposeiras-Roubin S, Abu-Assi E, Peyracchia M, Bocchino PP, *et al.* P2y12 inhibitors in acute coronary syndrome patients with renal dysfunction: an analysis from the RENAMI and BleeMACS projects. *Eur Heart J Cardiovasc Pharmacother* 2020;**6**:31–42.
22. Assessing the Fit of the Model. *Applied logistic regression*. Hoboken, NJ: John Wiley & Sons, Ltd; 2000. p.143–202.
23. Lloyd-Jones DM. Cardiovascular risk prediction: basic concepts, current status, and future directions. *Circulation* 2010;**121**:1768–1777.
24. Rubin DB. Inference and missing data. *Biometrika* 1976;**63**:581–592.
25. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014;**35**:1925–1931.
26. Van Calster B, McLernon DJ, Smeden M van, Wynants L, Steyerberg EW, Bossuyt P, *et al.* Calibration: the achilles heel of predictive analytics. *BMC Med* 2019;**17**:230.
27. Riley RD, Archer L, Snell KIE, Ensor J, Dhiman P, Martin GP, *et al.* Evaluation of clinical prediction models (part 2): how to undertake an external validation study. *BMJ* 2024;**384**:e074820.
28. Cox DR. Two further applications of a model for binary regression. *Biometrika* 1958;**45**:562–565.
29. Byrne RA, Rossello X, Coughlan JJ, Barbato E, Berry C, Chieffo A, *et al.* 2023 ESC guidelines for the management of acute coronary syndromes: developed by the task force on the management of acute coronary syndromes of the European Society of Cardiology (ESC). *Eur Heart J* 2023;**44**:3720–3826.
30. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res* 2019;**3**:18.
31. Vickers AJ, Calster BV, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016;**352**:i6.
32. Vickers AJ, Van Claster B, Wynants L, Steyerberg EW. Decision curve analysis: confidence intervals and hypothesis testing for net benefit. *Diagn Progn Res* 2023;**7**:11.
33. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;**44**:837–845.
34. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med* 2015;**13**:1.
35. Kotecha D, Asselbergs FW, Achenbach S, Anker SD, Atar D, Baigent C, *et al.* CODE-EHR best practice framework for the use of structured electronic healthcare records in clinical research. *Eur Heart J* 2022;**43**:3578–3588.
36. Wenzl FA, Bruno F, Kofoed KF, Raeber L, Roffi M, Stellos K, *et al.* Validation of the GRACE 3.0 score and redefinition of the risk threshold for early invasive treatment in non-ST-segment elevation acute coronary syndromes: a modelling study from five countries. *Eur Heart J* 2023;**44**:ehad655.1539.
37. Hung J, Roos A, Kadesjö E, McAllister DA, Kimenai DM, Shah ASV, *et al.* Performance of the GRACE 2.0 score in patients with type 1 and type 2 myocardial infarction. *Eur Heart J* 2020;**42**:2552–2561.
38. Shi B, Wang H, Liu J, Cai Z, Song C, Yin D, *et al.* Prognostic value of machine-learning-based PRAISE score for ischemic and bleeding events in patients with acute coronary syndrome undergoing percutaneous coronary intervention. *J Am Heart Assoc* 2023;**12**:e025812.
39. Hizoh I, Domokos D, Banhegyi G, Becker D, Merkely B, Ruzsa Z. Mortality prediction algorithms for patients undergoing primary percutaneous coronary intervention. *J Thorac Dis* 2020;**12**:1706–1720.
40. Hizoh I, Gulyas Z, Domokos D, Banhegyi G, Majoros Z, Major L, *et al.* A novel risk model including vascular access site for predicting 30-day mortality after primary PCI: the ALPHA score. *Cardiovasc Revasc Med* 2017;**18**:33–39.

41. De Luca G, Suryapranata H, van't Hof AWJ, de Boer M-J, Hoorntje JCA, Dambrink J-HE, *et al.* Prognostic assessment of patients with acute myocardial infarction treated with primary angioplasty. *Circulation* 2004;**109**:2737–2743.
42. Halkin A, Singh M, Nikolsky E, Grines CL, Tcheng JE, Garcia E, *et al.* Prediction of mortality after primary percutaneous coronary intervention for acute myocardial infarction: the CADILLAC risk score. *J Am Coll Cardiol* 2005;**45**:1397–1405.
43. de Mulder M, Gitt A, van Domburg R, Hochadel M, Seabra-Gomes R, Serruys PW, *et al.* EuroHeart score for the evaluation of in-hospital mortality in patients undergoing percutaneous coronary intervention. *Eur Heart J* 2011;**32**:1398–1408.
44. Zack CJ, Senecal C, Kinar Y, Metzger Y, Bar-Sinai Y, Widmer RJ, *et al.* Leveraging machine learning techniques to forecast patient prognosis after percutaneous coronary intervention. *JACC Cardiovasc Interv* 2019;**12**:1304–1311.
45. Mori M, Durant TJS, Huang C, Mortazavi BJ, Coppi A, Jean RA, *et al.* Toward dynamic risk prediction of outcomes after coronary artery bypass graft: improving risk prediction with intraoperative events using gradient boosting. *Circ Cardiovasc Qual Outcomes* 2021;**14**:e007363.
46. Wessler BS, Nelson J, Park JG, McGinnes H, Gulati G, Brazil R, *et al.* External validations of cardiovascular clinical prediction models: a large-scale review of the literature. *Circ Cardiovasc Qual Outcomes* 2021;**14**:e007858.
47. Bleeker SE, Moll HA, Steyerberg EW, Donders ART, Derksen-Lubsen G, Grobbee DE, *et al.* External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 2003;**56**:826–832.
48. Hall M, Bebb OJ, Dondo TB, Yan AT, Goodman SG, Bueno H, *et al.* Guideline-indicated treatments and diagnostics, GRACE risk score, and survival for non-ST elevation myocardial infarction. *Eur Heart J* 2018;**39**:3798–3806.
49. Saar A, Marandi T, Ainla T, Fischer K, Blöndal M, Eha J. The risk-treatment paradox in non-ST-elevation myocardial infarction patients according to their estimated GRACE risk. *Int J Cardiol* 2018;**272**:26–32.