Improving 1-year mortality prediction in ACS patients using machine learning

Sebastian Weichwald[1,2], **Alessandro Candreva** [3,†], **Rebekka Burkholz**[1,†]
Roland Klingenberg[3,4,5,6], **Lorenz Räber**[7], **Dik Heg**[8], **Robert Manka**[3],
Baris Gencer[9], **François Mach**[9], **David Nanchen**[10], **Nicolas Rodondi**[11,12],
Stephan Windecker [7], **Reijo Laaksonen** [13,14], **Stanley L. Hazen**[15,16],
Arnold von Eckardstein[17], **Frank Ruschitzka** [3], **Thomas F. Lüscher**[18,19],
Joachim M. Buhmann[1,*] and **Christian M. Matter** [3,18,*]

[1]Department of Computer Science, Institute for Machine Learning, ETH Zurich, Switzerland; [2]Max Planck Institute for Intelligent Systems, Tübingen, Germany; [3]Department of Cardiology, University Heart Center, University Hospital of Zurich, Switzerland; [4]Kerckhoff Heart and Thorax Center, Department of Cardiology, Kerckhoff-Klinik, Bad Nauheim, Germany; [5]Campus of the Justus Liebig University of Giessen, Germany; [6]DZHK (German Center for Cardiovascular Research), Partner Site Rhine-Main, Bad Nauheim, Germany; [7]Department of Cardiology, Cardiovascular Center, University Hospital of Bern, Switzerland; [8]Clinical Trial Unit, University of Bern, Switzerland; [9]Department of Cardiology, Cardiovascular Center, University Hospital of Geneva, Switzerland; [10]Department of Ambulatory Care and Community Medicine, University of Lausanne, Switzerland; [11]Institute of Primary Health Care (BIHAM), University of Bern, Switzerland; [12]Department of General Internal Medicine, Inselspital, Bern University Hospital, University of Bern, Switzerland; [13]Zora Biosciences, Espoo, Finland; [14]Finnish Cardiovascular Research Center Tampere, Tampere University, Tampere, Finland; [15]Department of Cardiovascular and Metabolic Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA; [16]Department of Cardiovascular Medicine, Heart and Vascular Institute, Cleveland Clinic, Cleveland, OH, USA; [17]Institute for Clinical Chemistry, University Hospital Zurich, Zurich, Switzerland; [18]Center for Molecular Cardiology, University of Zurich, Switzerland; and [19]Cardiology, Royal Brompton & Harefield Hospitals, London, United Kingdom

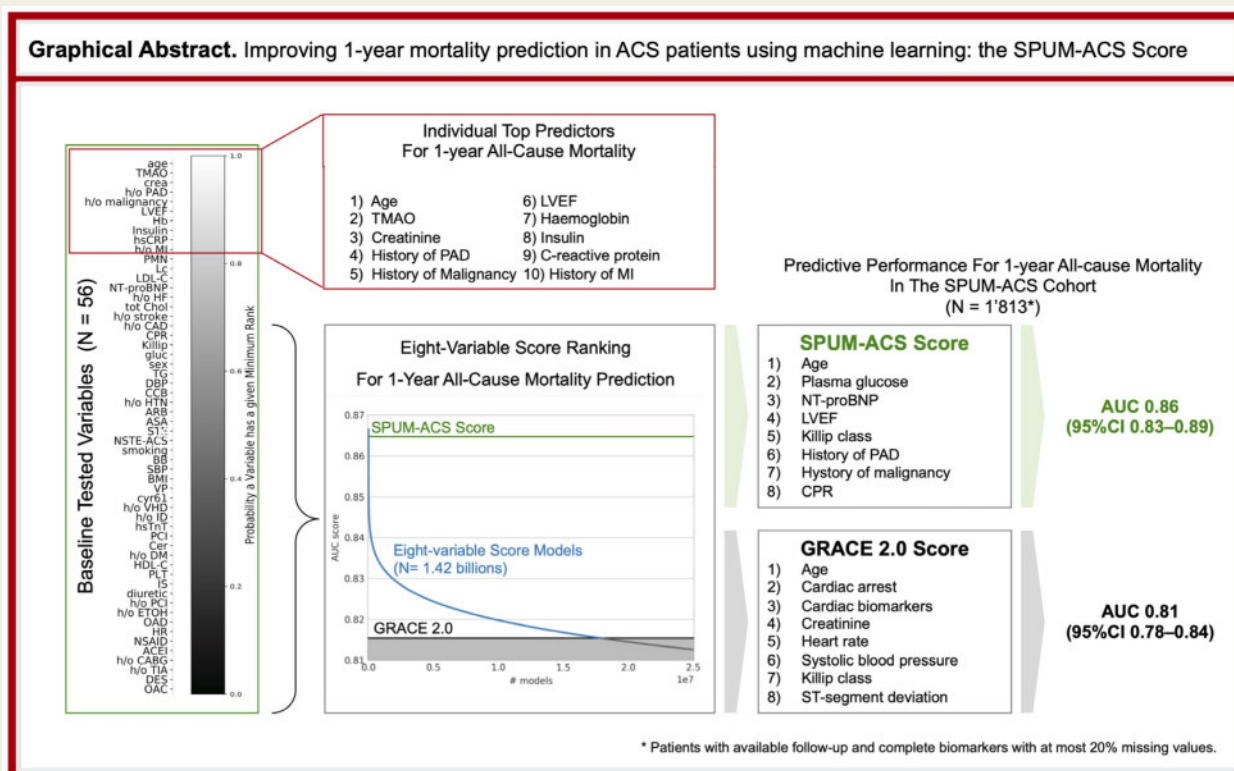|	
Background	The Global Registry of Acute Coronary Events (GRACE) score is an established clinical risk stratification tool for patients with acute coronary syndromes (ACS). We developed and internally validated a model for 1-year all-cause mortality prediction in ACS patients.
Methods	Between 2009 and 2012, 2'168 ACS patients were enrolled into the Swiss SPUM-ACS Cohort. Biomarkers were determined in 1'892 patients and follow-up was achieved in 95.8% of patients. 1-year all-cause mortality was 4.3% (n = 80). In our analysis we consider all linear models using combinations of 8 out of 56 variables to predict 1-year all-cause mortality and to derive a variable ranking.
Results	1.3% of 1'420'494'075 models outperformed the GRACE 2.0 Score. The SPUM-ACS Score includes age, plasma glucose, NT-proBNP, left ventricular ejection fraction (LVEF), Killip class, history of peripheral artery disease (PAD), malignancy, and cardio-pulmonary resuscitation. For predicting 1-year mortality after ACS, the SPUM-ACS Score outperformed the GRACE 2.0 Score which achieves a 5-fold cross-validated AUC of 0.81 (95% CI 0.78–0.84). Ranking individual features according to their importance across all multivariate models revealed age, trimethylamine N-oxide, creatinine, history of PAD or malignancy, LVEF, and haemoglobin as the most relevant variables for predicting 1-year mortality.
Conclusions	The variable ranking and the selection for the SPUM-ACS Score highlight the relevance of age, markers of heart failure, and comorbidities for prediction of all-cause death. Before application, this score needs to be externally validated and refined in larger cohorts.
Clinical Trial Registration	NCT01000701.

* Correspondence: Christian M. Matter, MD; Department of Cardiology, University Heart Center, University Hospital Zurich and Center for Molecular Cardiology, University of Zurich, Raemistrasse 100, 8091 Zurich, Switzerland; E-mail: christian.matter@usz.ch Phone: +41 44 255 3871 and Joachim M. Buhmann, Dr. rer. nat.; Institute for Machine Learning, Department of Computer Science, ETH Zurich; Universitätstrasse 6, CAB G69.2, 8092 Zürich, Switzerland; Email: jbuhmann@inf.ethz.ch Phone: +41 44 632 3124
†equal contribution

Graphical Abstract

Graphical Abstract. Improving 1-year mortality prediction in ACS patients using machine learning: the SPUM-ACS Score

Summarizing scheme of GRACE 2.0 and ML-based risk score for 1-year mortality and putative impact. Analyses of baseline variables in ACS patients for risk of mortality at 1 year. Easily available variables from clinical assessment build the basis for the GRACE score. The ML-derived multivariate SPUM-ACS risk score improves risk prediction compared to the GRACE 2.0 score.

Introduction

Despite advances in the care of patients with acute coronary syndromes (ACS), the incidence of major adverse cardiovascular events (MACE) remains high after ACS.[1] Improved risk prediction by a multidimensional assessment of ACS patients upon admission remains an unmet clinical need. Recent guidelines of the European Society of Cardiology (ESC) recommend the Global Registry of Acute Coronary Events (GRACE) Score which provides "*the most accurate stratification of risk both on admission and at discharge*".[2] The GRACE risk calculator provides estimates for risk of death in hospital, at 6 months, 1 year and 3 years.[3] Baseline variables in the GRACE score include age, systolic blood pressure, heart rate, serum creatinine, Killip class, cardiac arrest, elevated cardiac biomarkers, and ST-deviation.

Improved pathophysiological insights into ACS and the discovery of novel biomarkers reflecting its central pathways highlight the need for updated prognostic risk scores. Natriuretic peptides (B-type natriuretic peptide, N-terminal pro-B-type natriuretic peptide [NT-proBNP] and mid-regional pro-A-type natriuretic peptide) are more predictive of risk after ACS than cardiac troponin.[4] The Finnish

Corogene and SPUM-ACS studies identified the prognostic value of distinct plasma ceramide ratios.[5] Plasma *Trimethylamine-N-Oxide* (TMAO) levels enabled improved prediction of cardiovascular risk after ACS.[6] Combining high-sensitivity cardiac Troponin T (hsTnT), NT-proBNP and high-sensitivity CRP (hsCRP) with the GRACE score improved its predictive performance for short- and long-term outcomes.[7]

The computational feasibility of a machine learning-based approach for adverse event prediction was recently demonstrated, leading to the proposal of a novel risk assessment tool able to predict clinical outcome after ACS with good accuracy.[8] However, its clinical application is impaired by the high number of clinical features needed to run the algorithm.

We present a complementary approach. Combining modern screening tools of the Special Program University Medicine Acute Coronary Syndromes and Inflammation (SPUM-ACS) Cohort and exhaustive Machine Learning (ML) analyses, we aimed to develop 1) prototypical models for improved risk prediction focusing on easily-available low-cost variables (herein termed "practical"), and 2) a comprehensive ranking of traditional and novel variables by their prognostic relevance, with an

analysis that distils information about the model space of all multivariate linear risk models.

Methods

Detailed Methods in supplemental online materials

Please refer to the supplemental material for details on how the candidate models are built and assessed (missing data, continuous variables, normalized AUC), the robust variable ranking based on a Generalized Mallows Model, the variable contribution analysis, and on the variable occurrence correlation analysis.

The SPUM-ACS Cohort

The SPUM-ACS Cohort is a prospective cohort study of 2'168 patients consecutively enrolled between December 2009 and October 2012 at four Swiss university hospitals (Bern, Geneva, Lausanne and Zurich). Included were patients presenting within 5 days (preferably 72 hours) after pain onset, with a clinical diagnosis of ST-segment elevation myocardial infarction (STEMI), non-ST-segment elevation myocardial infarction (NSTEMI), or unstable angina. Baseline demographics, anthropometry, ECG, imaging parameters (TTE and angiography), and blood samples for biomarker analyses were included. Blood samples were taken before the coronary intervention at the time of arterial sheath insertion for coronary angiography. Inclusion and exclusion criteria, clinical endpoints, and blinded biomarker and predictor assessment are explicated in a previous publication.[7] The GRACE Score 2.0 for 1-year outcomes was calculated using the calculator at https://www.outcomes-umassmed.org/grace/ Institutional review boards of all participating centres approved the study protocol. All participants gave informed consent in compliance with the Declaration of Helsinki. The study is listed at ClinicalTrials.gov (NCT01000701).

Outcome and Clinical Endpoints

A 1-year clinical visit was scheduled for each patient. Events were adjudicated by three independent experts using pre-specified event adjudication forms. All-cause mortality (cardiac, vascular, non-cardiovascular) at 1 year is the primary endpoint.[7] The secondary endpoint was a composite of all-cause mortality or non-fatal recurrent MI. Patients for whom clinical outcomes could not be adjudicated or biomarker data could not be collected were excluded from further analysis.

Statistical Analyses and Machine Learning Tools

Learning complex machine learning models requires larger amounts of data than are usually available in a clinical setting. Linear models can cope with less data and are more robust against overfitting. The interplay between model complexity and overfitting is known as the bias-variance trade-off in statistical learning theory and suggests that too complex models may overfit to patterns unique to the training data and as a consequence may not predict well on new cohorts.[9] Linear models are well interpretable and prevalent in clinical practice. To further improve their robustness and generalization ability in the context of small data, we leveraged machine learning principles for model selection in addition to criteria related to clinical practicality.[10] Our objective was not only to build an 8-variable score that outperforms the GRACE score but also to deduce robust information about which variables complement each other and are highly predictive in a multitude of models when combined. Step-wise feature selection procedures are driven by the performance of only a few models along the search path. Thus, iterative model development and the assessment of predictor importance within one particular model may fail to incorporate and reflect the multiplicity of models that perform equally well, a concept Breiman described as "Rashomon".[11] Following this school of thought, our analysis aggregates information about the entire set of competing models to arrive at a robust feature importance ranking and a model that is a prototypical choice among all well-performing models.

We explored the full model space of all possible 8-variable linear models based on combinations of 56 baseline variables. Considering availability early after hospital admission, reliability and expense of analyses, predictors were categorized as practical or impractical (Supplemental Table 1); assignment of the variables to these categories was performed subjectively by the main authors in order to stratify them for routine clinical use. Model performance in terms of the AUC (Area Under the Receiver Operating Characteristics Curve or C-Statistic) was evaluated and validated, using 80% of the data for model fitting, and predicting outcomes on the remaining 20% that were blinded during model fitting. This was repeated 5 times in a stratified cross-validation scheme so that each 20% fold was used once as validation data. We had to rely on cross-validation given the lack of an external validation cohort. Cross-validation is an established internal validation procedure for model selection and assessing model performance: mimicking the derivation of a model on some training data and its application to separate new held-out data, it asymptotically consistently estimates the generalization error of a model and asymptotically selects the optimal model for the true unknown data generating distribution under mild conditions.[12]

Results

SPUM-ACS Cohort

Medical history and baseline characteristics of the 2'168 patients (Supplemental Table 1) were previously described (in reference 7: Figure 1, Table 1, Table 2).[7] Biomarkers were analysed in 1'892 patients. Of those, 52.4% had STEMI, 43.3% NSTEMI, and 4.3% unstable angina. Revascularisation by PCI was performed in 91.8%, and by coronary bypass graft in 3.9% of patients. Clinical outcomes were independently adjudicated at 1 year. Analysed were 1'813 patients, those with available follow-up and complete biomarkers with at most 20% missing values. At 1 year, all-cause mortality was 4.3% (n = 80), 3.5% (n = 64) were of cardiac origin; 3.7% had a non-fatal MI (n = 67).

Performance of all Risk Prediction Models

The GRACE 2.0 Score achieved an AUC of 0.81 (95% CI 0.78–0.84) for predicting all-cause mortality at 1 year (80 events, 1'813 patients). The AUC of all 8-variable models (n = 1'420'494'075) is depicted in Figure 1. 17'857'817 (1.3%) models outperformed the GRACE 2.0 Score with improvements up to 0.866 ± 0.03 AUC. The GRACE 1.0 Score (AUC 0.78 ± 0.02) was outperformed by 214'256'356 (15.1%) models; using dichotomisation the GRACE performance falls below 0.75 as previously reported.[7] Similar results were obtained for the combined endpoint of mortality or non-fatal MI (Supplemental Figure 1).

A Practical Model with Robust Performance

Among all GRACE-outperforming models for 1-year all-cause mortality prediction, the best achieved an AUC of 0.87 ± 0.03. However, this model contained impractical variables and was thus ill-suited for clinical application. Favouring the practicality of the proposed score
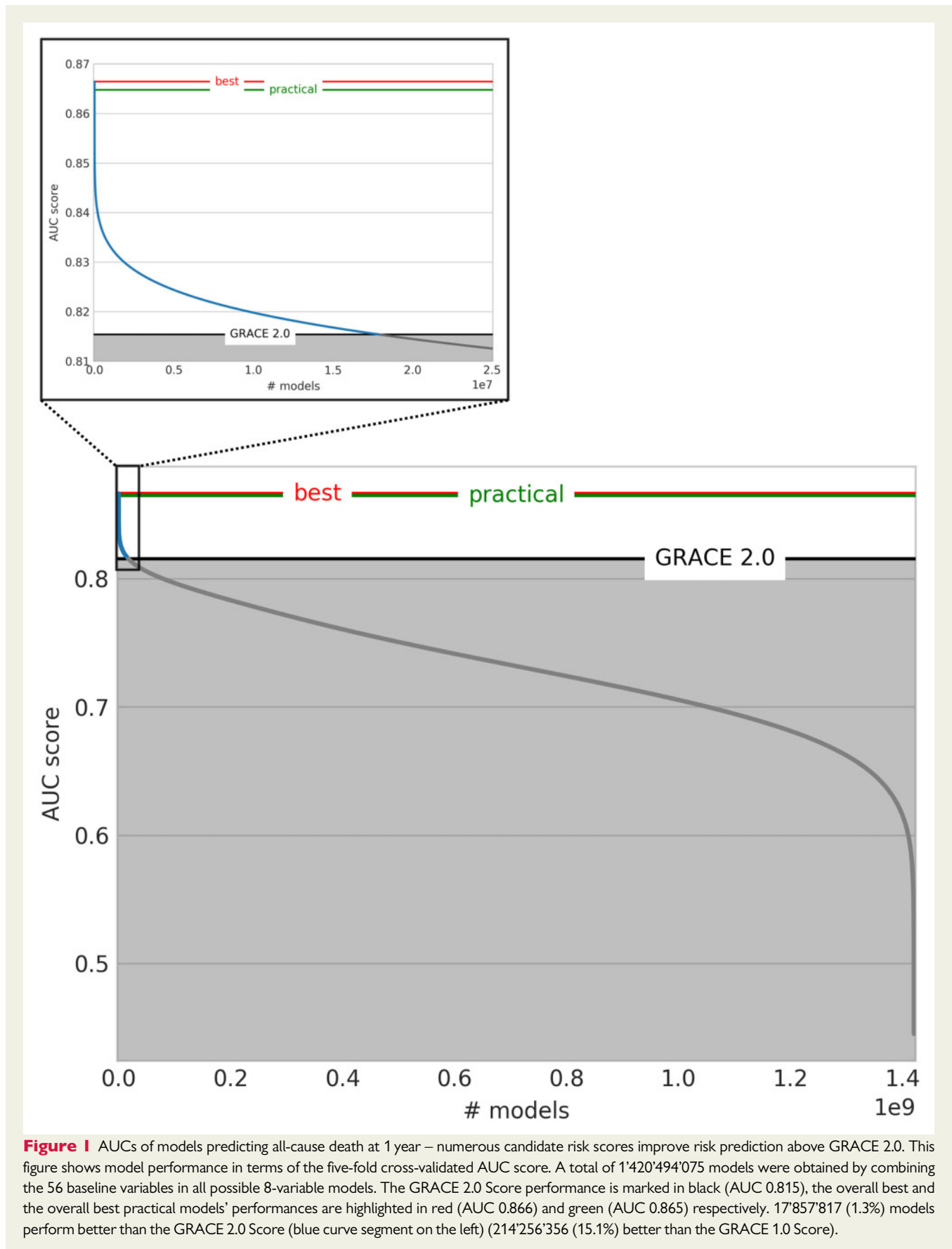
Figure 1 AUCs of models predicting all-cause death at 1 year – numerous candidate risk scores improve risk prediction above GRACE 2.0. This figure shows model performance in terms of the five-fold cross-validated AUC score. A total of 1'420'494'075 models were obtained by combining the 56 baseline variables in all possible 8-variable models. The GRACE 2.0 Score performance is marked in black (AUC 0.815), the overall best and the overall best practical models' performances are highlighted in red (AUC 0.866) and green (AUC 0.865) respectively. 17'857'817 (1.3%) models perform better than the GRACE 2.0 Score (blue curve segment on the left) (214'256'356 (15.1%) better than the GRACE 1.0 Score).

Table 1   The SPUM-ACS Score for predicting 1-year all-cause mortality after ACS

SPUM-ACS Score	GRACE 1.0 Score	GRACE 2.0 Score
Age	Age	Age
LVEF	History of MI	Cardiac arrest at admission
NT-proBNP	Elevated Cardiac Enzymes	Elevated Cardiac Enzymes
Plasma glucose	Creatinine	Creatinine
History of Malignancy	Heart Rate	Heart Rate
Killip Class	Systolic Blood Pressure	Systolic Blood Pressure
History of PAD	ST-segment Depression	ST-segment Deviation
Resuscitation	History of congestive HF	Killip Class
–	No in-hospital PCI	–
AUC = 0.86 ± 0.03	*AUC = 0.78 ± 0.02*	*AUC = 0.81 ± 0.03*

LVEF, left ventricular ejection fraction; NT-proBNP, N-terminal pro-brain natriuretic peptide; PAD, peripheral artery disease; HF, heart failure; MI, myocardial infarction.
Novel variables in the SPUM-ACS Score in bold.

Table 2   Practical application of the SPUM-ACS Score

SPUM-ACS SCORE		# 1: 79 yo NSTEMI	# 2: 69 yo Critical STEMI	# 3: 63 yo Benign STEMI
Risk	$0.04938 * (age - 63.47)$	0.7669 (79)	0.2731 (69)	-0.02321 (63)
	$0.2569 * (gluc - 6.287)$	-0.0994 (5.9)	1.8273 (13.4)	-0.1508 (5.7)
	$0.0002569 * (NT\text{-}proBNP - 389.3)$	3.167 (11264)	0.7329 (2906)	-0.01552 (336)
	$-0.044 * (LVEF - 50.87)$	-0.1817 (55)	0.6983 (35)	-0.402 (60)
	$0.971 * (Killip - 1)$	0 (1)	0 (1)	0 (1)
	$2.297 * h/o\ PAD$	0 (no)	0 (no)	0 (no)
	$2.108 * h/o\ malignancy$	0 (no)	0 (no)	0 (no)
	$1.354 * CPR$	0 (no)	1.354 (yes)	0 (no)
Sum		3.653	4.886	-0.592
Final Score [0,100]	$100/(1+exp(-sum + 0.55))$	96	99	24

Calculation coefficients for each variable in the SPUM-ACS Score are displayed. A score calculator is freely available as supplement online material (**Supplemental File "SPUM-ACS Score.html"**). For illustrative purposes, the SPUM-ACS Score was computed for three cohort patients, each representing a typical ACS scenario:
Example #1) Haemodynamically stable 79-year-old patient with NSTEMI and presenting on admission NT-proBNP level of 11264 ng/L, plasma glucose of 5.9 mmol/L;
Example #2) 69-year-old patient diagnosed with STEMI, presenting on admission elevated NT-proBNP level (2906 ng/L) and plasma glucose (13.4 mmol/L), with cardiac arrest;
Example #3) 63-year-old patient diagnosed with STEMI, presenting on admission slightly increased NT-proBNP level (336 ng/L) and plasma glucose of 5.7 mmol/L.
Final SPUM-ACS scores are displayed at the bottom.
In case of a missing variable, the corresponding row is to be excluded from the overall sum, i.e., a row holds a value 0 if the respective variable is missing.

for facilitating its application in clinical routine, we focussed on variables that are reliably available, early after admission, and with low expense (Supplemental Table 1).

We aimed to identify a model that both outperforms the GRACE score on our derivation cohort and can be expected to exhibit robust performance on new cohorts. Therefore, we first identified all 17 (out of 658'008) practical variable quintuplets that always outperform GRACE independent of which three variables they are combined with. That is, any model outperforms GRACE whenever at least one of these quintuplets is part of it. A model based on those quintuplets can be considered a robust representation of all practical GRACE-outperforming models. These variable quintuplets comprise the following 10 variables: *age, glucose, NT-proBNP, polymorphonuclear neutrophils (PMN), Killip class, peripheral artery disease (PAD), malignancy, left ventricular ejection fraction (LVEF), creatinine, and cardio-pulmonal reanimation (CPR)*. Among all 8-variable models comprised of those

10 variables we choose the one with highest cross-validated AUC score. This choice is robust and satisfies further selection criteria: it is the model with the highest normalised AUC score, comprises the 8 variables that appear most often among the 17 GRACE-outperforming variable quintuplets, and shares on average the most variables with all 17 GRACE-outperforming variable quintuplets.

Specifications of the SPUM-ACS Risk Score

The SPUM-ACS Score achieves a cross-validated AUC of 0.86 ± 0.03 (95% CI [0.83, 0.89]) for 1-year all-cause mortality prediction and comprises *age, glucose, NT-proBNP, Killip class, PAD, malignancy, LVEF, and CPR*. It outperforms GRACE 2.0 on all cross-validation folds. We termed this robust GRACE-outperforming practical model the **SPUM-ACS Score** (Table 1). The formula for computing the score (including logistic function clipping to the 0-100 range) is provided

Figure 2 SPUM-ACS Score calculator available as online supplemental file "SPUM-ACS Score.html".

(Table 2). A calculator is made available as supplement online material (Figure 2, **Supplemental File "SPUM-ACS Score.html"**).

Cut-off values identify three equally-sized patient groups as follows: *Low-risk* patients with SPUM-ACS Score between 0 and 33 and one-year all-cause mortality rate of 0.17%; the *intermediate-risk* group with cut-off values of 34 point and 66 points and one-year all-cause mortality rate of 1.8%; *high-risk* patients scoring above 66 points and with one-year all-cause mortality rate of 11.2% (Supplemental Table 2).

Our analyses focused on all-cause mortality at 1 year because of the increased complexity associated with predicting composite outcomes. We also analysed the combined outcome at 1 year (Supplemental Figure 1).

Robust Variable Importance Ranking for multivariate Prediction of 1-Year Mortality

Variables were ranked according to their importance across all multivariate 8-variable models for predicting 1-year all-cause mortality (Figure 3) (for the combined endpoint see Supplemental Figure 2). The probability for a variable to appear at a given rank is depicted by the grayscale in the respective cell, ranging from white (probability = 1) to black (probability = 0). *Age* was found to be the dominant factor, covering all top positions with probability close to 1 (Supplemental Table 3). The order among the remaining variables was more uncertain. TMAO, serum creatinine level, and PAD followed age as top predictors when combined with 7 variables for predicting 1-year all-cause mortality.

Variable Contribution in the SPUM-ACS Score

The SPUM-ACS Score performance depends on all 8 variables (Figure 4). The AUC-decrease analysis (see Supplemental Methods) highlights h/o PAD as most important for this model (AUC decrease ≈ 0.014), followed by h/o malignancy (AUC-decrease ≈ 0.012), while Killip class was least important (AUC decrease < 0.001).

Pairwise Variable Occurrence Correlations Indicating Good Variable Combinations
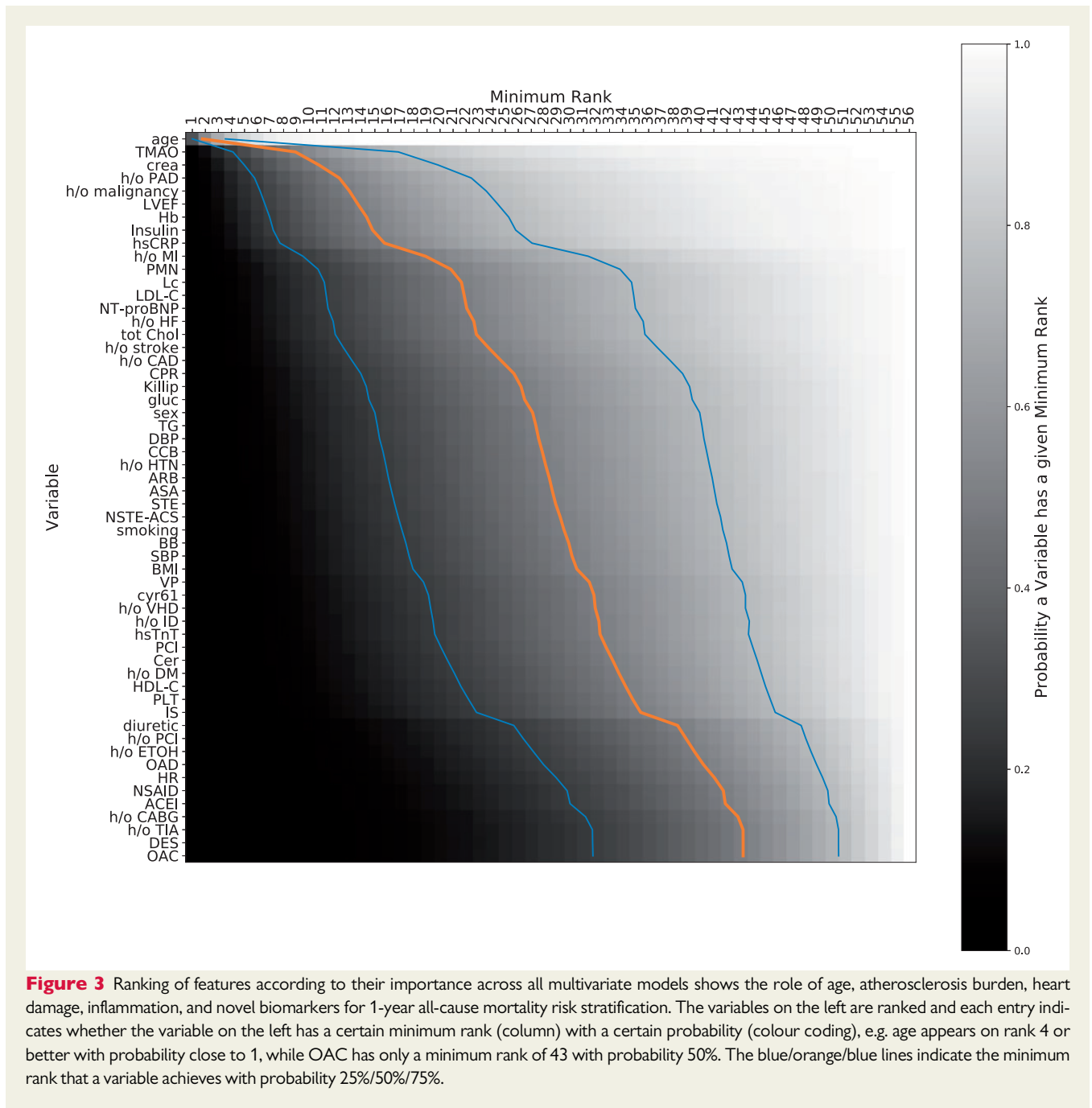
Certain variable pairs lead to more GRACE-outperforming models than expected given both their ranks (Figure 5). The two strongest correlations, CPR–NT-proBNP and h/o malignancy–NT-proBNP, are realized in our SPUM-ACS Score. NT-proBNP has a mediocre rank. Yet, it becomes valuable when complemented by h/o malignancy. It is also connected to CPR, which is linked to age, the top-ranked variable. Age combines well with many variables and defines a cluster. CPR bridges this important variable cluster to the one containing NT-proBNP.

Discussion

We propose a risk score for ACS patients having analysed more than 1.4 billion multivariate linear models (all 8-variable combinations of 56 baseline variables) for predicting 1-year all-cause mortality. Among all GRACE Score-outperforming models, we selected a prototypical choice of a top-performing 8-variable model comprising practical variables only and expected to exhibit robust performance on new cohorts. We propose this model as the **SPUM-ACS Score** for improved risk stratification of patients after ACS. Our comprehensive analysis enabled a robust variable ranking displaying the relative importance of variables within multivariate risk stratification models. Notably, due to multivariate interactions, lower-ranked variables may achieve high performance in some 8-variable models *when combined* with certain 7 variables while performance is low when combined with most other 7 variables.

The SPUM-ACS and the GRACE 2. 0 Score for Predicting 1-Year All-Cause Death

The SPUM-ACS Score comprises 8 variables: age, glucose, NT-proBNP, LVEF, Killip class, history of PAD, malignancy, and CPR. All

Figure 3 Ranking of features according to their importance across all multivariate models shows the role of age, atherosclerosis burden, heart damage, inflammation, and novel biomarkers for 1-year all-cause mortality risk stratification. The variables on the left are ranked and each entry indicates whether the variable on the left has a certain minimum rank (column) with a certain probability (colour coding), e.g. age appears on rank 4 or better with probability close to 1, while OAC has only a minimum rank of 43 with probability 50%. The blue/orange/blue lines indicate the minimum rank that a variable achieves with probability 25%/50%/75%.

are known outcome predictors in ACS. Yet, the combination of both traditional (age, Killip class, history of PAD, and CPR) and more recent risk markers (glucose, LVEF, NT-proBNP, and history of malignancy) in our proposed risk model for prediction of 1-year all-cause mortality after ACS is less common. Parameters describing ventricular performance such as Killip class (clinical pulmonary congestion), LVEF, and NT-proBNP (right or left heart filling pressure) provide complementary information. Moreover, the SPUM-ACS Score does not differentiate between ST-elevated and non ST-elevated ACS, in contrast to the TIMI Score.[13] A recent monocentric study obtained similar individual risk predictors, performing extensive phenotyping and ML for predicting 6-month mortality in 9066 ACS patients in

Finland.[14] Yet, we focussed on the prediction of 1-year all-cause mortality, a hard clinical endpoint that is unaffected by trial design or clinical interpretation. Recently D'Ascenzo and colleagues[8] proposed the Prospective Randomized Amlodipine Survival Evaluation (PRAISE) score for the prediction of 1-year all-cause death after ACS. Complementing their analysis, we tested a multivariate 8-variable score comparable to GRACE 2.0, and included NT-proBNP, LVEF, and novel biomarkers (e.g. Cyr61, TMAO, ceramide). Our exhaustive model space analysis focuses on extracting information about predictor variables that are stable across a multitude of simple models instead of relative feature importance within a selection of more complex machine learning models. Moreover, even if externally
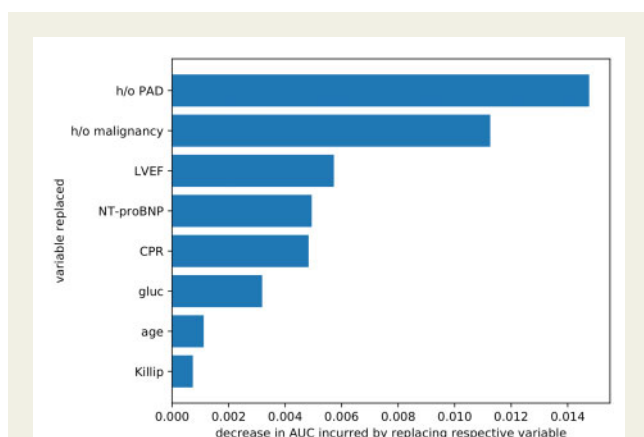
Figure 4 Variable contribution to the performance of the SPUM-ACS Score. The SPUM-ACS Score model achieves AUC 0.86 while GRACE 2.0 achieves 0.81 (GRACE 1.0 achieves 0.78); the bars indicate how much the performance drops if the respective variable is being replaced by the next best practical variable, e.g. the model where we replace the LVEF variable is 0.006 AUC worse than the original model, while replacing Killip class in the SPUM-ACS model incurs a lesser drop in model performance, hence LVEF plays a more important role for this model.

Figure 5 Pairwise correlations of variable occurrence in models outperforming GRACE 2.0. This scheme illustrates the degree to which variables complement each other in good models: Each node in the network corresponds to one variable. Its colour indicates a cluster membership. Clusters mark variables that are *better combinable than expected* based on both their ranks. Gray links mark good combinations of variables in different clusters, while coloured links highlight good inter-cluster combinations. The thicker a line, the higher is the correlation to appear together in good models. We only visualize the biggest 5% of positive correlations. Links that are realized in our score are highlighted with black borders. The two strongest correlations between CPR—NT-proBNP and h/o malignancy—NT-proBNP are among them.

validated, the PRAISE score was not challenged against the GRACE scores, while reaching an AUC of 0.82 in the external validation cohort , comparable to the performance of the GRACE 2.0 Score in our population. Finally, the PRAISE score including 25 clinical features did not represent a valid simplification in the clinical decision making. The SPUM-ACS score can be instead computed with only 8 variables largely available in clinical routine.

The SPUM-ACS Score was cross-validated within our Swiss multicentric biomarker cohort. Compared to the state-of-the-art GRACE 2.0 Score, prediction of 1-year all-cause mortality improved. 17 quintuplets outperformed the GRACE Score independent of which three variables they were combined with. Thus, using one of those 17 GRACE-outperforming quintuplets, competitive performance can still be achieved using five variables only (for example, using the best-performing quintuplet comprising the variables age, glucose, NT-proBNP, malignancy, and PAD). Our exhaustive search over more than 1.4 billion linear models also puts the solid baseline performance of the traditional GRACE 1.0 score into context: it was outperformed by *only* 15.1% of models that, in contrast to the GRACE score, were developed on the SPUM-cohort (the GRACE 2.0 Score was outperformed by 1.3% of models).

Value of Age and Metabolic Risk Markers

Age is a well-established risk predictor. The top 10 ranked variables include several non-traditional predictors: TMAO, creatinine plasma level, PAD, malignancy, LVEF, haemoglobin, Type 2 diabetes mellitus (T2DM), hsCRP, and history of MI. These parameters describe different pathophysiological contexts, such as cardiovascular metabolism, immune response, heart failure, or cardiomyocyte injury.

Our analyses also included ceramides, i.e. the Cer(d18:1/16:0)/Cer(d18:1/24:0) ratio, a novel molecular lipid species whose
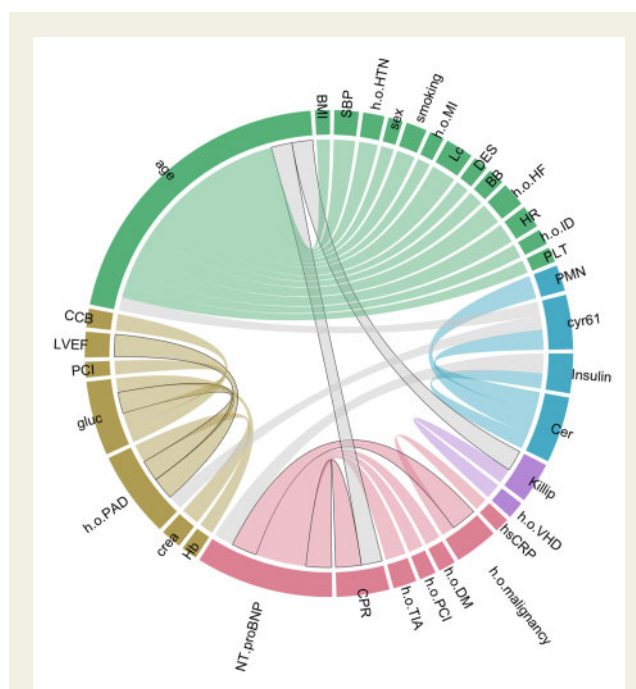
prognostic reclassification power was previously validated in the Corogene, BECAC, and SPUM-ACS cohort.[5] Interestingly, ceramides were the highest ranked variable for 1-year combined events, overtaking *age and LDL-C*.

TMAO is an intestinal metabolite of choline and phosphatidylcholine, produced by gut microbiota. Increased blood TMAO levels may confer pro-atherogenic and pro-thrombotic properties.[15] In our cohort, acute TMAO levels were a major prognostic determinant for both outcomes.

Independently from a pre-existing metabolic dysregulation by diabetes mellitus, hyperglycaemia upon admission for ACS is associated with worse outcomes.[16] Whether hyperglycaemia is a marker of the activation of the sympathetic nervous system at presentation or a mediator of the disease process remains unclear.[17]

Predictive Role of Acute Myocardial Injury, Impaired LV Function, and Inflammation

Cysteine-rich angiogenic inducer 61 (Cyr61) turned out non-inferior to hs-cTnT for improving prediction of all-cause mortality after ACS when added to the GRACE 1.0 score at 1 year.[18,19] Here, Cyr61

scored high for the combined outcome at 1 year. NT-proBNP levels upon admission are powerful prognostic markers in ACS patients.[20] Combined with hs-cTnT and hsCRP, NT-proBNP can improve risk prediction of the GRACE Score.[7] Within the SPUM-ACS Score, NT-proBNP contributes to the overall performance (AUC-drop $\approx$ 0.005). Killip Class showed higher-order interactions with many variables and acts as easily available surrogate marker of cardiac performance in our score. Among the surrogate markers of acute inflammation, counts of PMN were among the top-performing quintuplets.

Cardiovascular Risk Related to History of Malignancy, Renal Function, Anaemia, and Peripheral Artery Disease

Cancer survivors and patients with active cancer carry an increased cardiovascular risk. This association of the two dominant causes of death in Western countries is a clinical conundrum, where shared risk factors (e. g. smoking, diabetes, obesity), common pathogenic pathways (e. g. IL1β),[21,22] and cardiovascular side effects of chemo-, radio- or immunotherapies offer only partial explanations. The prognostic impact of history of malignancy was significant after 1 year.

Renal function assessed by serum creatinine plasma level, is featured in the GRACE Score 1.0 and 2.0 and is known to predict cardiovascular events.[23] Creatinine measurement was a powerful risk predictor for all-cause death at 1 year in our cohort. Renal function does not feature in our ML-based ACS Score, likely due to its strong pathogenic overlap with age and HF.

Anaemia may be trigger, consequence, or prognostic bystander of heart failure and ACS.[24] Our variable ranking confirmed the value of baseline haemoglobin for 1-year mortality prediction.

PAD captures relevant patient history that provides complementary information to the history of MI reflecting plaque burden in a different arterial bed. Both were strong predictors of overall mortality at 1 year and part of our ML-based ACS Score.

Robustness of the Model

Model performance evaluation by cross validation is less prone to overfitting to specific characteristics of the SPUM Cohort and bias than the commonly reported in-sample AUC.[12] To increase robustness of variable ranking and model selection, the AUCs were normalized within each validation fold accounting for the variation in AUC across training-test splits. By fitting a Generalized Mallows Model (GMM) to the five count-based rankings a consensus ranking was obtained. The SPUM-ACS Score combines the best GRACE-outperforming quintuplets and is a robust representative of all *practical* GRACE-outperforming models.

Pairwise Clustering Implies Complex Inter-Variable Correlations

Higher order correlations between variables may explain why some of the top-ranked variables (e.g. creatinine level) are not present in our score, whereas lower-ranked variables are. Clusters reflect variables that combine well. Overlap in prognostic information of variables in different clusters may explain why they are not suitable to build a strong risk score. The SPUM-ACS Score utilizes the two strongest variable combinations from the pairwise analyses, i. e. CPR—NT-proBNP and h/o malignancy—NT-proBNP.

Furthermore, CPR combines well with age, and age combines well with Killip class, again two variable combinations that are represented in the SPUM-ACS Score. This variable selection strategy explains why the 8 top-ranked parameters in the variable ranking differ from those featured in our robust risk score model.

Limitations

Our study tests 56 variables associated with outcomes in ACS patients. However, not all currently known risk markers were collected or analysed. Further, variables were categorized into practical and impractical in line with the experience of the clinical co-authors. As such, this categorization is subjective and researchers in different clinical environments may categorize differently. This limits transfer to different clinical contexts with different availability and cost constraints, while at the same time delineates the importance and possibility to take such considerations into account already during model development. This way, one can prevent purely data-driven machine learning pipelines to yield models that are impractical for the intended clinical use and instead refine the space of possible models to one's needs. The proposed SPUM-ACS Score was focused on models that outperformed the GRACE 2.0 Score in predicting all-cause death at 1 year, without considering other less robust clinical outcomes or shorter time windows such as in-hospital mortality.

The relatively low number of events in our cohort limits the reliability of our models when predicting events from baseline variables in other cohorts. Our study reports a 1-year all-cause mortality rate which may be in the lower range, but appears in line with other studies (4.1–4.9%),[25,26] likely reflecting different patient characteristics at baseline and therapy of patients. We searched for external validation cohorts but could not identify a cohort that the corresponding authors were willing to share and that provided sufficient matching of the key variables needed for our SPUM-ACS Score.

Conclusions and Perspectives

Our findings highlight the computational feasibility of large-scale exhaustive ML-based analyses for improved risk stratification in cardiovascular medicine: the proposed SPUM-ACS Score is a representative and robust choice among all candidate models that outperform the GRACE 2.0 Score for predicting 1-year all-cause mortality and among which our internal cross-validation cannot further resolve. The score comprises age, glucose, NT-proBNP, LVEF, Killip class, and history of PAD, malignancy, and CPR. The variable importance ranking highlights the relevance of age, heart failure/damage (NT-proBNP, creatinine, Cyr61, hs-cTnT), lipid metabolism (Cer), stress (glucose), atherosclerosis burden (history of PAD and MI), and immune response (PMN, hsCRP, history of malignancy) across all multivariate models for risk prediction.

Our data underscore the need for a comprehensive and expedient management of ACS patients in these areas for improving prognosis in ACS patients. Confirmation, refinement, and external validation of our proposed exhaustive model development procedure is needed.

Acknowledgements and funding

Conflict of interest

FM has received research grants to the institution from Amgen, AstraZeneca, Boston Scientific, Biotronik, Medtronic, MSD, Eli Lilly and St. Jude Medical including speaker or consultant fees. SWi has received research grants to the institution from Abbott, Boston Scientific, Biosensors, Biotronik, the Medicines Company, Medtronic and St. Jude Medical and honoraria from Abbott, Astra Zeneca, Eli Lilly, Boston Scientific, Biosensors, Biotronik, Medtronic and Edwards. TFL received research grants to the institution from AstraZeneca, Bayer Healthcare, Biosensors, Biotronik, Boston Scientific, Eli Lilly, Medtronic, MSD, Merck, Roche and Servier, including speaker fees by some of them. CMM received research grants to the institution from Eli Lilly, AstraZeneca, Roche, Amgen and MSD including speaker or consultant fees. LR received speaker fees and research grants to the institution from St. Jude Medical. CTU Bern, which is part of the University of Bern, has a staff policy of not accepting honoraria or consultancy fees. SLH is named as co-inventor on pending and issued patents held by the Cleveland Clinic relating to cardiovascular diagnostics or therapeutics, and reports having the right to receive royalty payment for inventions or discoveries related to cardiovascular diagnostics or therapeutics from Cleveland Heart Laboratory Inc, Quest Diagnostics and Procter & Gamble. SLH also reports having been paid as a consultant for Procter & Gamble, and receiving research funds from Astra Zeneca, Procter & Gamble, Pfizer Inc, and Roche Diagnostics. All other authors declared no competing interests.

References

1. Szummer K, Wallentin L, Lindhagen L, Alfredsson J, Erlinge D, Held C, James S, Kellerth T, Lindahl B, Ravn-Fischer A, Rydberg E, Yndigegn T, Jernberg T. Improved outcomes in patients with ST-elevation myocardial infarction during the last 20 years are related to implementation of evidence-based treatments: experiences from the SWEDEHEART registry 1995-2014. *Eur Heart J* 2017;**38**: 3056–3065.

2. Ibanez B, James S, Agewall S, Antunes MJ, Bucciarelli-Ducci C, Bueno H, Caforio ALP, Crea F, Goudevenos JA, Halvorsen S, Hindricks G, Kastrati A, Lenzen MJ, Prescott E, Roffi M, Valgimigli M, Varenhorst C, Vranckx P, Widimský P, Collet J-P, Kristensen SD, Aboyans V, Baumbach A, Bugiardini R, Coman IM, Delgado V, Fitzsimons D, Gaemperli O, Gershlick AH, Gielen S, Harjola V-P, Katus HA, Knuuti J, Kolh P, Leclercq C, Lip GYH, Morais J, Neskovic AN, Neumann F-J, Niessner A, Piepoli MF, Richter DJ, Shlyakhto E, Simpson IA, Steg PG, Terkelsen CJ, Thygesen K, Windecker S, Zamorano JL, Zeymer U, Windecker S, Aboyans V, Agewall S, Barbato E, Bueno H, Coca A, Collet J-P, Coman IM, Dean V, Delgado V, Fitzsimons D, Gaemperli O, Hindricks G, Iung B, Jüni P, Katus HA, Knuuti J, Lancellotti P, Leclercq C, McDonagh T, Piepoli MF, Ponikowski P, Richter DJ, Roffi M, Shlyakhto E, Simpson IA, Zamorano JL, Chettibi M, Hayrapetyan HG, Metzler B, Ibrahimov F, Sujayeva V, Beauloye C, Dizdarevic-Hudic L, Karamfiloff K, Skoric B, Antoniades L, Tousek P, Terkelsen PJ, Shaheen SM, Marandi T, Niemelä M, Kedev S, Gilard M, Aladashvili A, Elsaesser A, Kanakakis IG, Merkely B, Gudnason T, Iakobishvili Z, Bolognese L, Berkinbayev S, Bajraktari G, Beishenkulov M, Zake I, Lamin HB, Gustiene O, Pereira B, Xuereb RG, Ztot S, Juliebø V, Legutko J, Timóteo AT, Tatu-Chiţoiu G, Yakovlev A, Bertelli L, Nedeljkovic M, Studenčan M, Bunc M, García de Castro AM, Petursson P, Jeger R, Mourali MS, Yildirir A, Parkhomenko A, Gale CP, ESC Scientific Document Group. Group ESCSD. 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.).

3. Fox KA, Dabbous OH, Goldberg RJ, Pieper KS, Eagle KA, Van de Werf F, Avezum A, Goodman SG, Flather MD, Anderson FA, Jr., Granger CB. Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome: prospective multinational observational study (GRACE). *BMJ* 2006;**333**:1091.

4. Thygesen K, Mair J, Mueller C, Huber K, Weber M, Plebani M, Hasin Y, Biasucci LM, Giannitsis E, Lindahl B, Koenig W, Tubaro M, Collinson P, Katus H, Galvani M, Venge P, Alpert JS, Hamm C, Jaffe AS. Study Group on Biomarkers in Cardiology of the ESCWGoACC. Recommendations for the use of natriuretic peptides in acute cardiac care: a position statement from the Study Group on Biomarkers in Cardiology of the ESC Working Group on Acute Cardiac Care. *Eur Heart J* 2012;**33**:2001–2006.

5. Laaksonen R, Ekroos K, Sysi-Aho M, Hilvo M, Vihervaara T, Kauhanen D, Suoniemi M, Hurme R, März W, Scharnagl H, Stojakovic T, Vlachopoulou E, Lokki M-L, Nieminen MS, Klingenberg R, Matter CM, Hornemann T, Jüni P, Rodondi N, Räber L, Windecker S, Gencer B, Pedersen ER, Tell GS, Nygård O, Mach F, Sinisalo J, Lüscher TF. Plasma ceramides predict cardiovascular death in patients with stable coronary artery disease and acute coronary syndromes beyond LDL-cholesterol. *Eur Heart J* 2016;**37**:1967–1976.

6. Li XS, Obeid S, Klingenberg R, Gencer B, Mach F, Raber L, Windecker S, Rodondi N, Nanchen D, Muller O, Miranda MX, Matter CM, Wu Y, Li L, Wang Z, Alamri HS, Gogonea V, Chung YM, Tang WH, Hazen SL, Luscher TF. Gut microbiota-dependent trimethylamine N-oxide in acute coronary syndromes: a prognostic marker for incident cardiovascular events beyond traditional risk factors. *Eur Heart J* 2017;**38**:814–824.

7. Klingenberg R, Aghlmandi S, Räber L, Gencer B, Nanchen D, Heg D, Carballo S, Rodondi N, Mach F, Windecker S, Jüni P, von Eckardstein A, Matter CM, Lüscher TF. Improved risk stratification of patients with acute coronary syndromes using a combination of hsTnT, NT-proBNP and hsCRP with the GRACE score. *Eur Heart J Acute Cardiovasc Care* 2018;**7**:129–138.

8. D'Ascenzo F, De Filippo O, Gallone G, Mittone G, Deriu MA, Iannaccone M, Ariza-Solé A, Liebetrau C, Manzano-Fernández S, Quadri G, Kinnaird T, Campo G, Simao Henriques JP, Hughes JM, Dominguez-Rodriguez A, Aldinucci M, Morbiducci U, Patti G, Raposeiras-Roubin S, Abu-Assi E, De Ferrari GM, Piroli F, Saglietto A, Conrotto F, Omedé P, Montefusco A, Pennone M, Bruno F, Bocchino PP, Boccuzzi G, Cerrato E, Varbella F, Sperti M, Wilton SB, Velicki L, Xanthopoulou I, Cequier A, Iniguez-Romo A, Munoz Pousa I, Cespon Fernandez M, Caneiro Queija B, Cobas-Paz R, Lopez-Cuenca A, Garay A, Blanco PF, Rognoni A, Biondi Zoccai G, Biscaglia S, Nunez-Gil I, Fujii T, Durante A, Song X, Kawaji T, Alexopoulos D, Huczek Z, Gonzalez Juanatey JR, Nie S-P, Kawashiri M-A, Colonnelli I, Cantalupo B, Esposito R, Leonardi S, Grosso Marra W, Chieffo A, Michelucci U, Piga D, Malavolta M, Gili S, Mennuni M, Montalto C, Oltrona Visconti L, Arfat Y. Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): a modelling study of pooled datasets. *Lancet* 2021;**397**:199–207.

9. von Luxburg U, Schölkopf B. Statistical Learning Theory: Models, Concepts, and Results. *Handbook of the History of Logic 2011* 2011;**10**:651–706. Volume Pages

10. Chapelle O, Vapnik V, Bengio Y. Model Selection for Small Sample Regression. *Machine Learning. Machine Learning* 2002;**48**:9–23. volume pages

11. Breiman L. Statistical modeling: The two cultures. *Statistical Science* 2001;**16**: 199–215.

12. Dudoit S, van der Laan M. Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. *Statistical Methodology* 2005 July 2005;**2**:131–154. Volume Issue Pages

13. Ramsay G, Podogrodzka M, McClure C, Fox KA. Risk prediction in patients presenting with suspected cardiac pain: the GRACE and TIMI risk scores versus clinical evaluation. *QJM* 2006;**100**:11–18.

14. Hernesniemi JA, Mahdiani S, Tynkkynen JA, Lyytikainen LP, Mishra PP, Lehtimaki T, Eskola M, Nikus K, Antila K, Oksala N. Extensive phenotype data and machine learning in prediction of mortality in acute coronary syndrome - the MADDEC study. *Ann Med* 2019;**51**:156–163.

15. Tang WH, Wang Z, Levison BS, Koeth RA, Britt EB, Fu X, Wu Y, Hazen SL. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med* 2013;**368**:1575–1584.

16. Capes SE, Hunt D, Malmberg K, Gerstein HC. Stress hyperglycaemia and increased risk of death after myocardial infarction in patients with and without diabetes: a systematic overview. *Lancet* 2000;**355**:773–778.

17. De Caterina R, Madonna R, Sourij H, Wascher T. Glycaemic control in acute coronary syndromes: prognostic value and therapeutic options. *Eur Heart J* 2010;**31**:1557–1564.

18. Hilfiker A, Hilfiker-Kleiner D, Fuchs M, Kaminski K, Lichtenberg A, RothköTter H-J, Schieffer B, Drexler H. Expression of CYR61, an angiogenic immediate early gene, in arteriosclerosis and its regulation by angiotensin II. *Circulation* 2002;**106**:254–260.

19. Klingenberg R, Aghlmandi S, Liebetrau C, Räber L, Gencer B, Nanchen D, Carballo D, Akhmedov A, Montecucco F, Zoller S, Brokopp C, Heg D, Jüni P, Marti Soler H, Marques-Vidal P-M, Vollenweider P, Dörr O, Rodondi N, Mach F, Windecker S, Landmesser U, von Eckardstein A, Hamm CW, Matter CM, Lüscher TF. Cysteine-rich angiogenic inducer 61 (Cyr61): a novel soluble biomarker of acute myocardial injury improves risk stratification after acute coronary syndromes. *Eur Heart J* 2017;**38**:3493–3502.

20. de Lemos JA, Morrow DA, Bentley JH, Omland T, Sabatine MS, McCabe CH, Hall C, Cannon CP, Braunwald E. The prognostic value of B-type natriuretic peptide in patients with acute coronary syndromes. *N Engl J Med* 2001;**345**:1014–1021.

21. Ridker PM, MacFadyen JG, Thuren T, Everett BM, Libby P, Glynn RJ, Ridker P, Lorenzatti A, Krum H, Varigos J, Siostrzonek P, Sinnaeve P, Fonseca F, Nicolau J, Gotcheva N, Genest J, Yong H, Urina-Triana M, Milicic D, Cifkova R, Vettus R, Koenig W, Anker SD, Manolis AJ, Wyss F, Forster T, Sigurdsson A, Pais P, Fucili A, Ogawa H, Shimokawa H, Veze I, Petrauskiene B, Salvador L, Kastelein J, Cornel JH, Klemsdal TO, Medina F, Budaj A, Vida-Simiti L, Kobalava Z, Otasevic P, Pella D, Lainscak M, Seung K-B, Commerford P, Dellborg M, Donath M, Hwang J-J, Kultursay H, Flather M, Ballantyne C, Bilazarian S, Chang W, East C, Everett B, Forgosh L, Glynn R, Harris B, Libby P, Ligueros M, Thuren T, Bohula E, Charmarthi B, Cheng S, Chou S, Danik J, McMahon G, Maron B, Ning M, Olenchock B, Pande R, Perlstein T, Pradhan A, Rost N, Singhal A, Taqueti V, Wei N, Burris H, Cioffi A, Dalseg AM, Ghosh N, Gralow J, Mayer T, Rugo H, Fowler V, Limaye AP, Cosgrove S, Levine D, Lopes R, Scott J, Thuren T, Ligueros M, Hilkert R, Tamesby G, Mickel C, Manning B, Woelcke J, Tan M, Manfreda S, Ponce T, Kam J, Saini R, Banker K, Salko T, Nandy P, Tawfik R, O'Neil G, Manne S, Jirvankar P, Lal S, Nema D, Jose J, Collins R, Bailey K, Blumenthal R, Colhoun H, Gersh B, Glynn RJ. Effect of interleukin-1beta inhibition with canakinumab on incident lung cancer in patients with atherosclerosis: exploratory results from a randomised, double-blind, placebo-controlled trial. *Lancet* 2017;**390**:1833–1842.

22. Libby P, Kobold S. Inflammation: a common contributor to cancer, aging, and cardiovascular diseases-expanding the concept of cardio-oncology. *Cardiovasc Res* 2019;**115**:824–829.

23. Mann J. Renal failure and cardiovascular risk. Increased borderline serum creatinine–a warning sign?. *MMW Fortschr Med* 2001;**143**:30–34.

24. Park JY, Choi BG, Rha SW, Kang TS. Five-year outcomes in patients with anemia on admission undergoing a coronary intervention for acute myocardial infarction in Koreans: propensity score matching analysis. *Coron Artery Dis* 2018;

25. Rossello X, Dorresteijn JA, Janssen A, Lambrinou E, Scherrenberg M, Bonnefoy-Cudraz E, Cobain M, Piepoli MF, Visseren FL, Dendale P. Risk prediction tools in cardiovascular disease prevention: A report from the ESC Prevention of CVD Programme led by the European Association of Preventive Cardiology (EAPC) in collaboration with the Acute Cardiovascular Care Association (ACCA) and the Association of Cardiovascular Nursing and Allied Professions (ACNAP). *Eur J Prev Cardiol* 2019;**26**:1534–1544.

26. Wallentin L, Becker RC, Budaj A, Cannon CP, Emanuelsson H, Held C, Horrow J, Husted S, James S, Katus H, Mahaffey KW, Scirica BM, Skene A, Steg PG, Storey RF, Harrington RA, Investigators P, Freij A, Thorsen M. Ticagrelor versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med* 2009;**361**:1045–1057.