Development of a neural network-based competing risk model for long-term prognostication in ischemic heart disease from a large database of electronic health records and clinical registries

Peter C. Holm (1), Amalie D. Haue (1, 3), Alex H. Christensen (3, 4), Henning Bundgaard (3), and Søren Brunak (1, 5)

1. Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen, Denmark
2. Department Obstetrics and Gynecology, Copenhagen University Hospital, Kettegård Alle 30, DK-2650 Hvidovre, Denmark
3. Department of Cardiology, Copenhagen University Hospital, Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Denmark
4. Department of Cardiology, Copenhagen University Hospital, Herlev-Gentofte Hospital, Borgmester Ib Juuls Vej 1, DK-2730 Herlev, Denmark
5. Copenhagen University Hospital, Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Denmark

Abstract

Background: Machine-learning based risk-stratification of patients with ischemic heart disease (IHD) can improve prognostication through identification of both high and low risk patients. However, most models are unable to handle competing risks which can affect usability and scope.

Methods: We present PMHnetV2, designed for joint modelling of competing risks in the presences of censoring. PMHnetV2 was developed using health data from 52,787 patients (train: 42,065, test: 10,722) with a diagnosis of IHD who underwent coronary angiography between 2006 and 2016. Follow-up information could be collected up until 2019. Model performance was evaluated using the index of prediction accuracy and time-dependent AUC.

Results: PMHnetV2 accurately predicts all-cause mortality, cardiovascular mortality, recurrent ischemic events, and hospital admissions with ischemic heart disease complications and sequela. The models are well-calibrated and improve on the baseline models.

Conclusion: modeling competing risks jointly can improve risk prediction. Utilizing this approach could offer patients and clinicians a post-angiography assessment that might aid in the decision-making process.

Funding: Novo Nordisk Foundation, NordForsk, and the Innovation Fund Denmark.

Introduction

In recent years, machine learning models, particularly neural networks, have emerged as pivotal tools in precision medicine (1). These models excel at identifying complex patterns in large datasets, a characteristic particularly helpful for analysis of large-scale modern electronic health records. This capability has paved the way for creation of new and advanced clinical decision support tools (2). In the field of cardiology, machine learning models are showing immense potential, often surpassing the performance of traditional methods. These models are especially leading in the area of prediction and prognostication, where time-to-event outcomes and censoring are prevalent (3–7). Here, neural network-based survival models represent the current state of the art (6,8–12). Our previous contribution to this evolving landscape is PMHnetV1, a neural network-based survival model developed to predict all-cause mortality in patients with ischemic heart disease (13).

A notable shortcoming in our previous work was the inability to differentiate between deaths related to complications of ischemic heart disease and those arising from completely unrelated causes, alongside predicting specific measures of disease progression. To address this, we need methods capable of handling competing risks (14), an attribute rarely found in existing neural network-based models. The most prominent exception is the "DeepHit" approach presented by Lee et al. (6) for survival analysis with competing risks. In the cardiological domain, Pieszko et al. leveraged this approach in developing a prediction model for personalized risk assessment following myocardial perfusion imaging (5). Although they did not present calibration metrics, their unified model surpassed traditional perfusion abnormality measures in discrimination, offering event-specific risk estimates for outcomes such as all-cause mortality, acute coronary syndrome, and revascularization. Yet, as Kvamme et al. highlighted, DeepHit-based models, while presenting excellent discrimination, tend to lag in terms of calibration compared to alternatives (11).

Recognizing both advancements and existing gaps in the current literature, this study seeks to build on our previous work by introducing PMHnetV2. This enhanced version extends the number of included predictors and incorporates prediction of cardiovascular mortality, recurrent myocardial ischemia, and future hospital admissions with ischemic heart disease complications and sequelae. To this end, we implemented and utilized a discrete-time logistic hazard model that allows jointly modelling competing risks. We compare our competing risk models that use this strategy against the simplistic strategy of treating such risks as censoring.

Materials and Methods

Sources of data

The data used for model development originates from a combination of national registries (CPR, DAR, LMSR, LPR), Eastern Danish Heart Registry (PATS), and an electronic health record database (BTH). Data sources were linked using the Danish unique national personal identification number ("CPR number"), which is assigned to all residents of the country.

	Description	Ref.
The Danish Central Person Register (CPR)	Contains date of birth, sex, person status (alive/dead/emigrated), time of status change, etc. on all persons living in Denmark.	(15)
The Causes of Death Register (DAR)	Contains date of death and information on direct and contributory causes of deaths coded using the ICD-10 classification system. Based on the physician-filled death certificates, that have been legally required since 1871.	(16)
The Register of Pharmaceutical Sales (LMSR)	Contains data on all prescription drugs sold in Danish pharmacies since 1994.	(17)
The Danish National Patient Registry (LPR)	Nationwide longitudinal registry with data on all admissions, examinations, treatments, procedures, diagnoses, etc. in Danish public hospitals since 1977. Have been using the ICD10 coding system for diagnosis codes since 1994.	(18)
The east-Danish heart registry (PATS)	Clinical quality database with data from coronary angiographies and percutaneous coronary interventions performed at specialized departments in the Capital Region of Denmark and Region Zealand.	(19,20)
BTH	Database with data from population-wide electronic health record system (2.6 million patients) from 1st of January 2006 to 31st of July 2016. Contains highly heterogenous data ranging from administrative data to laboratory test results and unstructured journal text.	-

Inclusion criteria

The target population is individuals with a diagnosis of ischemic heart disease. Patients were included if they had been subjected to a coronary angiography demonstrating 1/2/3-vessels disease or diffuse atheromatosis in the period 1st of January 2006 and 31st of December 2016 (both inclusive). We excluded patients without a prior ICD-10 code for ischemic heart disease (I20-25), below 18 years of age at time of the coronary angiography, and those not alive two days after the procedure (Figure 1A). For additional details, see *Supplementary Methods*.

Follow-up and endpoints

The time origin was the index coronary angiography, which was defined as the first coronary angiography for each patient that satisfied the other inclusion criteria. We defined four different major outcomes:

1. All-cause mortality (ACMO), which was identified from the central person register.
2. Cardiovascular mortality (CVMO), which was defined from the cause of death register as any death with an ICD-10 code of I00-99 registered as the underlying cause of death. Deaths from other causes were competing risks.
3. Cardiovascular complications (CVCO), which was defined as a composite outcome from the national patient register diagnosis or procedure codes. CVCO was defined as hospital admission with either “heart failure” (ICD-10: I50), atrial fibrillation or flutter (I48), “cardiac arrest” (I46), or “cerebrovascular accident” (I61, I63-64) as the primary diagnosis code; or implantation of pacemaker (SKS: BFCA0*) or cardioverter-defibrillator (BFCB0*). Admissions within four weeks after time origin were ignored (Fig S1 and *Supplementary Methods*). ACMO was treated as a competing risk.
4. Myocardial ischemic events (MIEV), which was identified from the national patient register as in-patients hospitalized for at least 24 hours with the primary diagnosis code ICD-10: I20-I25, or a procedure code for PCI or CABG. In both cases, events within eight weeks after time of origin were ignored (Fig S2 and *Supplementary Methods*). ACMO was treated as a competing risk.

Variables and features

In the following, we distinguish variables from features: we use variables when referring to the raw input data from the various data sources and features to describe the pre-processed inputs to the machine learning model. To illustrate, the categorical variable "tobacco usage" is transformed into four separate binary features through one-hot encoding: "former", "current", "never", and "missing".

In total, we used 1,860 different variables, belonging to five different categories, which after preprocessing amounted to 2,262 distinct features. The five categories, and the number of features belonging in each, were clinical variables: 80, procedure codes: 418, prescription data: 785, diagnoses: 504, and laboratory tests: 475. The full set of features can be seen in *Supplementary File 1* and additional details are found in *Supplementary Methods*.

Numerical variables included in the clinical category are site-specific (n=17) stenosis amount and the baseline values of systolic and diastolic blood pressures, pulse, age, height, and weight. Stenosis amount and age were non-missing and only needed scaling. The others were processed with median imputation before scaling, and we added an extra feature to indicate any missing values. As an example, the numerical variable "systolic blood pressure" is represented by two features: one indicating if data is missing and another containing the median-imputed and scaled blood pressure values. We used one-hot encoding for all categorical variables, designating a "missing" category where necessary to retain potentially useful information in the patterns of missingness. We set up the preprocessing pipeline using only the training data to prevent data leakage between the training and testing datasets.

Discrete-time survival analysis with competing risks

To model time-to-event data with censoring and competing risks, we implemented a discrete-time logistic hazard model that allows jointly modelling competing risks. This approach can be viewed as an extension of the methodology proposed by Gensheimer and Narasimhan (8). Our updated framework is a fully parametric discrete time survival model parameterized by a neural network which enables jointly modelling discrete-time survival data with competing risks.

Briefly, and largely following the book by Gerhard Tutz (21), we discretize the follow-up time to estimate the conditional hazard for individual subjects. This hazard represents the probability that the event of interest occurs at a specific time point, given that the subject is still at risk at the start of the interval. To facilitate this, we introduce the indicator δ_{ijr} defined as:

$$\delta_{ijr} = \begin{cases} \mathbb{1}\{r = 0\} & \text{if } t_{(j)} < t_i \\ \mathbb{1}\{r = r_i, t_{(j)} = t_i\} & \text{otherwise} \end{cases}$$

where t_i denotes the observed discrete survival time of subject i, r_i represents the observed event (with 0 indicating censoring), and $t_{(j)}$ refers to the time bin j. With this setup, we can write the negative log-likelihood as

$$loss = -\sum_{i=1}^{N}\sum_{s=1}^{t_i}\left(\sum_{r \in \kappa}\delta_{ijr}\log\lambda_r(s|x_i) + \delta_{ij0}\log\left(1 - \sum_{r \in \kappa}\lambda_r(s|x_i)\right)\right)$$

The model outputs conditional hazards as logits with an estimate for each risk (including survival) at each time interval.

Model development

We used a ResNet-like architecture for tabular data similar to the one described in (22). The architecture depends on the number of input features m, the number of hidden neurons in each hidden layer h, and the number of output logits o, which is $|\text{time bins}| \times (|\text{risks}| + 1)$.

$$ResBlock(x) = x + \big(BatchNorm \circ Linear_{h,h} \circ Dropout \circ SiLU \circ Linear_{h,h} \circ Dropout\big)(x)$$

$$ResNet_{i,o}(x) = (Linear_{i,h} \circ ResBlock_h \circ \ldots \circ Resblock_h \circ Linear_{h,o})(x)$$

Predictions are obtained by passing the output logits through a SoftMax activation function, such that the conditional probability of each possible event (including survival) sums to 100%.

We used the holdout method and randomly split our derivation dataset into a training set (80%) and a test set (20%) for validation (Figure 1B). The training set was further subdivided into training (80%) and validation (20%) splits to have an unbiased estimate of model performance during hyperparameter optimization. All models were trained using the “AdamW” stochastic optimization algorithm (23) following the “super-convergence” training regimen previously described (24), which enables fast, accurate, and resource-efficient training of neural network models (25). We used a fixed batch size of 1024 and trained for a pre-specified number of epochs using the one-cycle learning rate scheduler (“OneCycleLR” in PyTorch).

Hyperparameters; which consisted of a) number of epochs b) max learning rate (upper boundary for the learning rate policy), c) dropout rate, d) weight decay, e) number of residual blocks, f) number of hidden units in each residual block, g) number of time bins in the discretization grid, upper limits on the retrospective inclusion windows for h) biochemical, i) diagnosis, j) procedure, and k) medication data, and whether to use l) skip-connections and m) batch-normalization; were optimized using Optuna (26). We used the index of prediction accuracy (IPA) of the primary outcome on the validation split as the tuning parameter, which is an R-squared type measure that reflects both calibration and discrimination (27). The IPA is time-dependent, so we calculated IPA at 50 evenly spaced timepoints from 0.5 to 5 years and numerically integrated the values to provide a single measure of performance. See *Supplementary Methods* for additional details.

Serving as benchmark, we trained naïve single-risk models where competing events was treated as censoring.

Statistics

For evaluation of the time-to-event prediction models, we used the time-dependent area under the receiver operating characteristic curve (AUC) and the IPA, a scaled version of the Brier score, as the main measures of performance (27–29). Model calibration was assessed graphically by comparing model estimates with pseudovalues of the actual outcomes (30,31). See *Supplementary Methods* for additional details.

We derived estimates for the incidence of the different endpoints using the Kaplan-Meier and the Aalen-Johansen estimators for endpoints without or with competing risks, respectively.

Software

We used PyTorch 1.13.0 (32), Lightning 2.0 (33), Scikit-Learn 1.3 (34), and Optuna 3.1 (26) for Python 3.10 for developing and training the neural network models. Our general purpose discrete-time competing risk framework that we developed for this application has been released on the python package index (PyPI) under the name “Discotime” and can also be found on GithHub at “peterchristofferholm/discotime”. We used version 0.1.0 of Discotime for this study.

Statistical analysis, data wrangling, and visualization was primarily performed using R (v. 4.1) (35) using the “riskregression” package for evaluation of prediction models (29), “survival“ for survival analysis (36), “ggplot2” for visualization (37), and the various packages in the “tidyverse” for ad-hoc data analysis (38).

The source code for the analyses in this study will be made available on GitHub after publication.

Conflicts of interest

Søren Brunak reports ownerships in Intomics, Hoba Therapeutics, Novo Nordisk, Lundbeck, and ALK; and managing board memberships in Proscion and Intomics. Henning Bundgaard reports ownership in Novo Nordisk and has received lecture fees from Amgen, BMS, MSD and Sanofi.

Data access and ethics

The study was approved by the Danish Data Protection Agency (ref: 514-0255/18-3000, 514-0254/18-3000, SUND-2016-50), The Danish Health Data Authority (ref: FSEID-00003724 and FSEID-00003092), and The Danish Patient Safety Authority (3-3013-1731/1/). Danish personal identifiers were pseudonymized prior to any analysis.

Study design, methods, and results were reported in agreement with the TRIPOD statement (39).

Data availability statement

Due to national and EU regulations, the source data used in this study cannot be made publicly available. Research groups with access to secure and dedicated computing environments can request access to the source data registries via application to the Danish Health Data Authority.

Funding

This study was funded by Novo Nordisk Foundation (grant agreements: NNF14CC0001 and NNF17OC0027594) – Hellerup, Denmark; NordForsk (PM Heart; grant agreement: 90580) – Oslo, Norge; and the Innovation Foundation (BigTempHealth; grant agreement: 5153-00002B) – Aarhus, Denmark.

Results

Table 1 shows the baseline characteristics of the 52,809 ischemic heart disease patients in our derivation cohort, which were randomly allocated into a training set (n = 42,048) and a test set (n = 10,761). Figures 1A and 1B illustrate patient inclusion and the temporal distribution of the index procedures, respectively. Both sets demonstrated considerable similarity in baseline characteristics. The baseline characteristics have been further stratified by sex, to enable comparison between men and women which are known to have different disease manifestations (40). In the training set, women comprised 31.1%, and men 68.9%; in the test set, the distribution was 30.8% women and 69.2% men. The median age was consistent at 66 (IQR: 15) years for men and 70 (IQR: 15) years for women in the training set and 70 (IQR: 16) years for women in the test set.

At the time of the index coronary angiography, women had a higher prevalence of diffuse coronary atherosclerosis compared to men, with 39.9% in the training set and 39.6% in the test set for women, versus 23.3% for men in the training set and 22.8% in the test set. Conversely, men exhibited a higher incidence of two or three-vessel disease, with 41.0% in the training set and 41.4% in the test set, compared to 27.1% in women for the training set and 27.3% for the test set. Tobacco usage was higher among men, with 28.5% of men in the training set and 28.7% in the test set identified as active smokers, in contrast to 25.1% of women in the training set and 26.3% in the test set.

Medication usage was prominent, with 61.8% of men and 64.6% of women in the training set taking lipid-lowering medications, and similar trends observed in the test set (61.9% of men and 65.1% of women). Anti-hypertensive medication usage was particularly substantial, with 76.6% of men and 86.1% of women in the training set, and 76.4% of men and 86.3% of women in the test set on these drugs. Hypertension was a common diagnosis recorded in 35.2% of men and 44.3% of women in the training set, and 34.7% of men and 44.3% of women in the test set. The hypertension diagnosis in this context refers to earlier hospital admissions where an I10 diagnosis was given, explaining the discrepancy with the proportion of patients on anti-hypertensive medication.

We defined four endpoints for our time-to-event prediction models: a) All-Cause Mortality (ACMO), b) Cardiovascular Mortality (CVMO), c) Other Cardiovascular Complications (CVCO), and d) New Myocardial Ischemia Events (MIEV). For endpoints other than ACMO, we also delineated the competing risks that would prevent patients from reaching the primary endpoint of interest. With follow-up data extending up to February 2019, we developed prediction models aimed at long-term risk stratification over a five-year prediction horizon.

Figure 1C showcases Kaplan-Meier estimates of survival probability for ACMO and Aalen-Johansen estimates for the cause-specific cumulative incidences of CVMO, MIEV, and CVCO. For ACMO, the 5-year survival probability estimate is 81.7% (CI: 81.3—82.1%) for the training set and 81.6% (CI: 80.9—82.4%) for the test set. In the case of CVMO, the estimated 5-year incidence of cardiovascular death is 8.1% (CI: 7.8—8.4%) in the training set and 8.3% (CI: 7.8—8.8%) in the test set. For MIEV, we estimated a 5-year incidence of new ischemic events to be 36.8% (CI: 36.3—37.2%) in the training set and 37.3% (CI: 36.3—38.2%) in the test set. Lastly, for CVCO, the 5-year incidence of complications was found to be 37.3% (CI: 36.9—37.8%) in the training set and 37.4% (CI: 36.5—38.4%) in the test set.

We set up neural network models for prediction of each of the four endpoints. To finetune key parameters of the models, we performed hyperparameter optimization. Figure S3 shows the overview of the hyperparameter sweeps. The best performing trial, based on the integrated cause-specific IPA, yielded scores of 22.2% for ACMO, 12.6% for CVMO, 23.5% for CVCO, and 8.0% for MIEV. These scores were all calculated using a validation split of the training set. For the IPA metric, a higher score is better: a perfect model has a score of 100%, a useful model has a score above 0%, and models with an IPA $\leq$ 0% are considered useless or harmful (27). In each of the sweeps, there were several trials with a performance in this "useless" range, which showcases the importance of careful parameter tuning. Figure S4 shows the relationship between the individual hyperparameter values and the average performance of trials using those specific values. Some hyperparameters were found to impact performance more than others. Enabling skip connections and batch normalization were in all cases associated with a better model performance. From visual inspection, learning rate, number of hidden units, number of blocks, and weight decay appears to be the parameters most critical to fine-tune for optimal performance. Table S1 shows the overview of the hyperparameters, specifying the search space used during optimization and the specific hyperparameter configuration for the best-performing trials. These final configurations were used in the training of the final models. Figure S5 shows the training history of the final models, tracking the negative log-likelihood for both the training and validation splits, as well as the IPA scores on the validation split at each training step. Across all models, we observe a concurrent increase in IPA as the negative log-likelihood decreases, with no signs of model overfitting.

The models were subsequently evaluated using the hitherto unseen test data. The prediction performance was assessed both qualitatively, through graphical representations (Figure 2), and quantitatively, using performance metrics AUC, Brier score, and IPA (Figure 3). The performance metrics are all time-dependent and were thus computed at 100 evenly spaced prediction horizons

ranging from 31 days to 5 years. For ease of presentation, we will primarily highlight the 1, 3, and 5-years predictions in the text.

Comparing the predicted t-year ACMO estimates across the observed outcomes (Figure 2), we observe the quantiles of the predicted mortality risk to be consistently higher for patients that experience the primary outcome (labeled “primary”) compared to those that do not (labeled “event-free”). This suggests good model discrimination, which is corroborated quantitatively by the AUC values. Specifically, the AUC of the ACMO model was 84.9% (CI: 83.3 - 86.6) at 1 year, 83.4% (CI: 82.3 - 84.6) at 3 years, and 83.4% (CI: 82.8 - 84.4) at 5 years (Figure 4). The IPA, which measures both discrimination and calibration, was 13.7% (CI: 7.4 - 20.1) at 1 year, 19.7% (CI: 16.0 - 23.5) at 3 years, and 25.8% (CI: 22.0 - 28.8) at 5 years. The IPA is a measure obtained by scaling the Brier score of the model with that of a null model based on observed incidence (27), and reflects both model discrimination and calibration. The Brier score for the ACMO model was significantly better (i.e., lower) than that of the null model (Kaplan-Meier) across all 100 evaluated timepoints (*Supplementary File 2*). To further assess calibration, we examined a calibration plot and found the model to be well-calibrated, with the calibration regression curve closely aligning with the 45-degree reference line (Figure 3) across all evaluated timepoints.

Examining the cause-specific CVMO predictions, we noted that the predicted risk was higher for patients experiencing cardiovascular mortality ("primary") and those with non-cardiovascular mortality ("competing") compared to those who remained event-free ("event-free") (Figure 2). Additionally, the risk quantiles for the "primary" group were consistently higher than for the "competing" group, although the difference was less pronounced. In terms of calibration, the model generally performed well, but there was some evidence of overestimation among those with the highest predicted risks (Figure 3). The number of patients at the highest end of predicted risks is however very low, so the calibration estimate there carries considerable uncertainty.

Quantifying the model performance, AUC for the CVMO model was 85.7% (CI: 83.7 - 87.7) at 1 year, 84.6% (CI: 83.1 - 86.1) at 3 years, and 82.8% (CI: 81.4 – 84.2) at 5 years. The IPA was 10.4% (CI: 1.7 – 19.0) at 1 year, 13.0% (CI: 7.1 – 19.0) at 3 years, and 14.2% (CI: 9.3 - 19.0) at 5 years. When comparing the Brier score of our PMHnet model to a null model based on observed incidence (Aalen-Johansen), our model outperformed the null model across all 100 evaluated timepoints.

We also included a single-risk naïve version of the CVMO model in the analysis, treating competing risks as censored events. This naïve model had worse AUC values at 56 of the 100 timepoints but was otherwise comparable to the competing risks model in terms of discrimination. However, the Brier score was worse at 97 out of the 100 evaluated timepoints.

For the MIEV predictions, the boxplots conditional on the t-year outcome reveal that cases undergoing with new ischemic events ("primary") generally have higher model predictions compared to those in the "competing" or "event-free" categories. However, the difference between the medians is relatively modest (Figure 2). This observation is supported by the AUC values, which are 70.1% (CI: 68.9 – 71.3) at 1 year, 69.1% (CI: 68.0 - 70.2) at 3 years, and 65.9% (CI: 64.8 - 67.1) at 5 years (Figure 4). The IPA further quantifies this with scores of 6.8% (CI: 3.8 - 9.0) at 1 year, 9.2% (CI: 7.3 - 11.1) at 3 years, and 6.8% (CI: 5.2 – 10.0) at 5 years.

The model's predictions for the test set are generally well-calibrated, except at the most extreme end of the predicted risks (Figure 3). Compared to a reference "null" model, the Brier score of the MIEV model was superior at 98 of the 100 evaluated timepoints. The same was observed for the “naïve” single-risk version of the model. However, compared to the competing risk model, this single-risk model had a worse AUC and Brier at 87 and 99 of the timepoints, respectively.

The CVCO model was also found to be well-calibrated, as seen from the calibration curve (Figure 3), and the Brier score (Figure 4). At 99 of the 100 timepoints the MACO model had significantly better brier score than the null model and 97 of 100 had better Brier score than the “naïve” single risk model. In terms of discrimination, the AUC was 79.6 (CI: 78.6-80.6) at 1 year, 79.9 (CI: 79.0-80.9) at 3 years, and 78.5 (CI: 77.5-79.4) at 5 years. Additionally, the IPA was 19.2% (CI: 16.7-21.7) at 1 year, 24.9% (CI: 23.0-26.8) at 3 years, and 23.9 (CI: 22.0-25.7) at 5 years.

Discussion

In this study, we presented the development of a collection of four different neural network-based time-to-event models for risk-prediction in ischemic heart disease. These models were developed to predict four different key outcomes following coronary angiography: all-cause mortality (ACMO), cardiovascular mortality (CVMO), cardiovascular complications (CVCO), and recurrent myocardial ischemic events (MIEV). We utilized a comprehensive dataset combining electronic health records and clinical registries, including more than 52 thousand Danish patients with ischemic heart disease and 2262 different features. Models were trained using a subset of 42 thousand patients, while validation was carried out on a separate subset of more than 10 thousand patients. From the evaluation of the model performances, we found that the cause and time-specific estimates of all models were well-calibrated and can be used to discriminate between patients experiencing the different endpoints and those who remain event-free, which points to the ability of the models to be of clinical utility.

This collection of models, which we refer to as PMHnetV2, represents an update to our previous contribution to field of machine learning-based prognostication in ischemic heart disease, PMHnetV1 (13), which was limited to the prediction of all-cause mortality. Enabling the development of the PMHnetV2 models, a key contribution of our study is the introduction of a new approach for construction of competing risk time-to-event models with neural networks. Our discrete-time approach can be viewed as an extension to the methodology first proposed by Gensheimer and Narasimhan (8) which enables jointly modelling competing risk data. The theoretical foundations of this extension is described in detail by Tutz and Schmid in the context of classical statistical analysis of discrete failure times (21), but we are, to the best of our knowledge, the first to apply this methodology to neural networks.

In the model evaluation, we compared the PMHnetV2 models based on this methodology to single-risk models that treat competing events as censoring. The single-risk version of the models was found to have good model discrimination and calibration, however in almost all cases, the models that are able account for competing risks significantly improved on both discrimination and calibration. In general, we found that the effect on calibration was the most pronounced. From this, we conclude that in prediction of endpoints with competing risks, it is more effective to adopt a methodology that allows for joint modelling of competing events, rather than the common practice of treating competing events as censoring.

In addition to our novel competing risk neural network-approach, a major strength of this study is the size of the dataset and its diversity, which enhances the applicability of our models to a broad patient population with ischemic heart disease. Where we in PMHnetV1 limited inclusion to only cover patients subject to their first ever coronary angiography, we in this study also included patients with one or more angiographies recorded prior to the inclusion period. Consequently, the patient population now include cases with more chronic manifestations of ischemic heart disease, as exemplified by the fact that 5,917 of patients had a history of one or more PCI/CABG procedures at the time of their index coronary angiography.

Despite its strengths, our study has several limitations worth highlighting. Presently, our model utilizes more than 2200 different features, many of which may be colinear and possibly redundant. There is a possibility that the model could be equally effective with a reduced number of features, however this is not an aspect that we have explored.

Moreover, most of our features are categorical and not continuous. We chose to discretize laboratory values according to their reference ranges, a decision aimed at enhancing the generalizability of feature processing and more closely aligning with clinical practice. However, an inherent limitation of our current framework is the lack of temporal resolution of features. For example, our models are unable discern whether an above-reference serum creatinine test occurred one week or one year prior to the assessment. To address this, we included hyperparameters defining the time-window for inclusion of features, specifying how far back in time data is considered. While this approach provides a partial solution, it is not ideal. Future studies should instead explore the integration of neural network architectures such as LSTM (41) or Transformers (42) for inclusion of sequential and time-resolved features.

Utilizing retrospective clinical data presented the challenge of missing variables, a phenomenon that can occur for several reasons. It could be a deliberate omission, such as the decision to forgo a specific blood test irrelevant to the diagnostic process. Alternatively, a clinician might have considered it unnecessary to document certain details, like a normal blood pressure measurement, at a given contact. Furthermore, there could be inadvertent gaps over time related to systematic issues with data storage, registration, or processing that often cannot be avoided in a dynamic clinical environment. Rather than presuming the missingness to be random and hence suitable for imputation through sophisticated methods, we chose to explicitly encode where data was missing. This approach preserves the integrity of the existing data and leverages the missing information to enhance the predictive capabilities of the model, resulting in a model that is both robust and reflective of real-world clinical scenarios.

Lastly, we found it very difficult to create the perfect algorithmic definition for determining the timing of disease recurrence and progression from large retrospective clinical. As a pragmatic work around, the endpoints cardiovascular complications (CVCO) and recurrent myocardial ischemic events (MIEV) were defined from surrogate markers, such as revascularization (PCI/CABG) and hospital admission with critical diagnoses as the pragmatic prediction target. These events, reliably identified from registry data, serve as indicators of development or worsening of symptoms in such a degree that it warranted clinical intervention.

In conclusion, our study makes a significant contribution to the field by implementing and testing a novel neural network approach for modeling survival data with competing risks in patients with ischemic heart disease.

Tables

		Training (n = 42,048)		Test (n = 10,761)	
Variable	Value	Male (n = 28,980)	Female (n = 13,068)	Male (n = 7,442)	Female (n = 3,319)
Age (years)	Median (IQR)	66 (15)	70 (15)	66 (15)	70 (16)
Height (cm)	Median (IQR) [NA%]	177 (9) [8.0%]	164 (8) [7.6%]	177 (9) [8.1%]	164 (8) [7.4%]
Weight (kg)	Median (IQR) [NA%]	85 (19) [4.7%]	70 (20) [4.9%]	85 (19) [5.1%]	70 (20) [4.9%]
Systolic blood pressure (mmHg)	Median (IQR) [NA%]	138 (32) [24.2%]	140 (32) [23.9%]	138 (31) [24.0%]	140 (33) [22.8%]
Diastolic blood pressure (mmHg)	Median (IQR) [NA%]	80 (18) [24.2%]	76 (18) [23.9%]	80 (19) [24.0%]	76 (19) [22.8%]
Coronary vessel pathology	Diffuse coronary atherosclerosis	6,760 (23.3%)	5,213 (39.9%)	1,696 (22.8%)	1,314 (39.6%)
	Single-vessel disease	10,318 (35.6%)	4,314 (33.0%)	2,669 (35.9%)	1,101 (33.2%)
	Two-vessel disease	5,808 (20.0%)	1,917 (14.7%)	1,471 (19.8%)	490 (14.8%)
	Three-vessel disease	6,094 (21.0%)	1,624 (12.4%)	1,606 (21.6%)	414 (12.5%)
Tobacco usage	Active	8,272 (28.5%)	3,283 (25.1%)	2,136 (28.7%)	873 (26.3%)
	Former	11,771 (40.6%)	4,204 (32.2%)	3,058 (41.1%)	1,035 (31.2%)
	Never	6,361 (21.9%)	4,505 (34.5%)	1,637 (22.0%)	1,103 (33.2%)
	Missing	2,576 (8.9%)	1,076 (8.2%)	611 (8.2%)	308 (9.2%)
Same day PCI		11,503 (39.7%)	4,340 (33.2%)	3,011 (40.5%)	1,061 (32.0%)
Previous PCI/CABG	0	25,488 (88.0%)	11,825 (90.5%)	6,582 (88.4%)	2,997 (90.3%)
	1	2,677 (9.2%)	986 (7.6%)	673 (9.0%)	255 (7.7%)
	2+	815 (2.8%)	257 (2.0%)	187 (2.5%)	67 (2.0%)
Medication	Lipid-lowering	17,908 (61.8%)	8,438 (64.6%)	4,610 (61.9%)	2,160 (65.1%)
	Anti-hypertensive	22,189 (76.6%)	11,252 (86.1%)	5,684 (76.4%)	2,863 (86.3%)
	Non-insulin glucose lowering medication	5,304 (18.3%)	2,219 (17.0%)	1,360 (18.3%)	587 (17.7%)
	Insulin	1,909 (6.6%)	994 (7.6%)	475 (6.4%)	239 (7.2%)
Diagnoses (ICD-10)	Diabetes (E10, E11)	4,650 (16.0%)	2,144 (16.4%)	1,153 (15.5%)	543 (16.4%)
	Hypertension (I10)	10,211 (35.2%)	5,794 (44.3%)	2,579 (34.7%)	1,470 (44.3%)
	COPD (J44)	2,017 (7.0%)	1,431 (11.0%)	512 (6.9%)	358 (10.8%)
	Heart failure (I50)	4,107 (14.2%)	1,611 (12.3%)	1,024 (13.8%)	440 (13.3%)
	Atrial fibrillation or atrial flutter (I44)	3,523 (12.2%)	1,556 (11.8%)	880 (11.8%)	378 (11.4%)

Table 1: Baseline statistics of the study population stratified by dataset split (training/test) and sex (male/female). Numbers in this table exclusively represent data available at the time origin. Medication was defined from the prescription database (LMSR) as a filled prescription of a drug belonging to the class of *lipid-lowering medication* (ATC: C10), *anti-hypertensive medication* (C02, C03, C07, C08, C09), *type-2 diabetes medicine* (A10B), or *insulin* (A10A). Numbers are reported as "median (IQR)", "median (IQR) [NA%]", or "count (%)". CABG: coronary artery bypass grafting, IQR: Interquartile range. PCI: percutaneous coronary intervention.

Figures

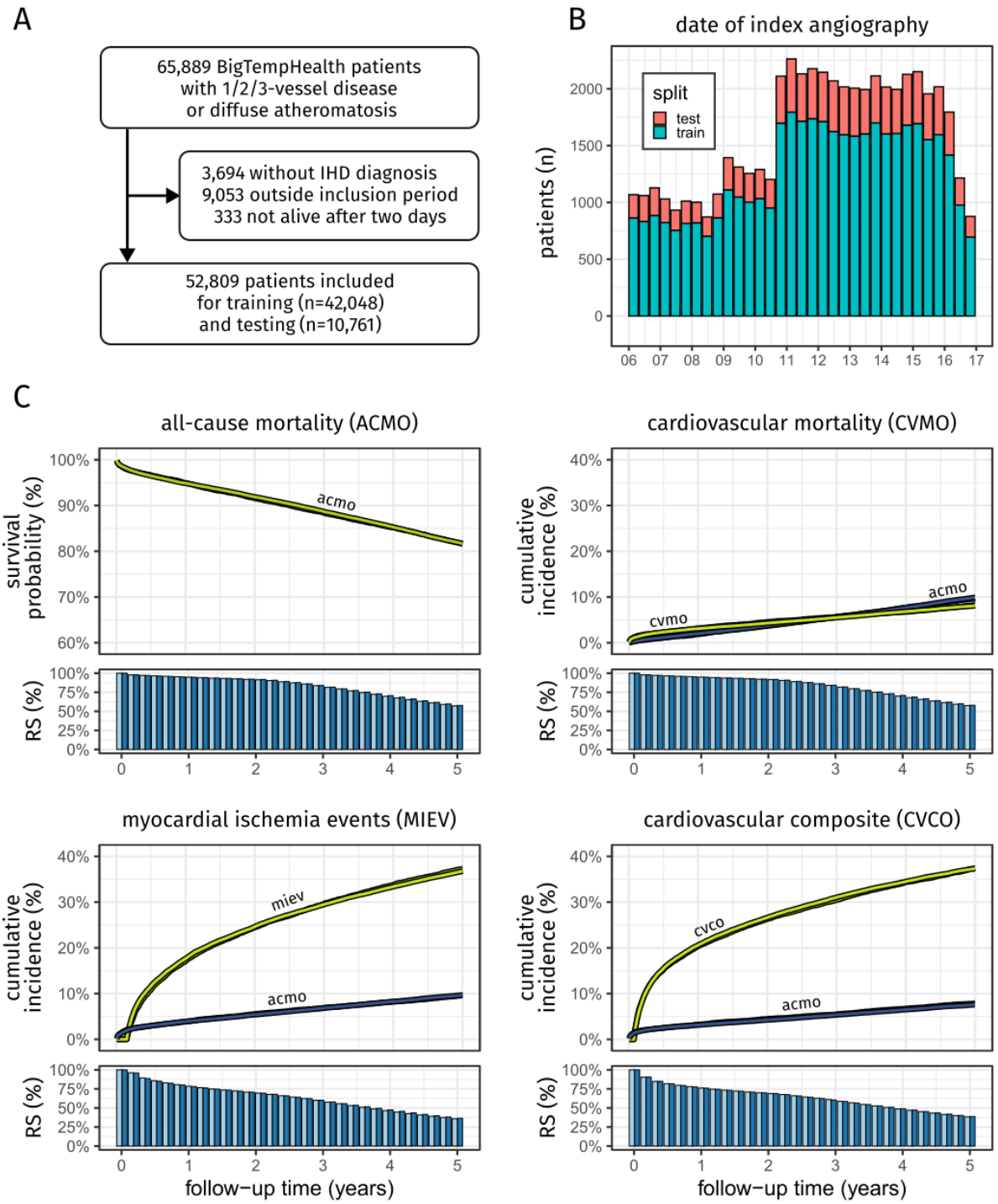


Figure 1: *Inclusion diagram (A), overview of inclusion dates (B), and observed survival and cumulative incidence curves across the four primary endpoints (C).*

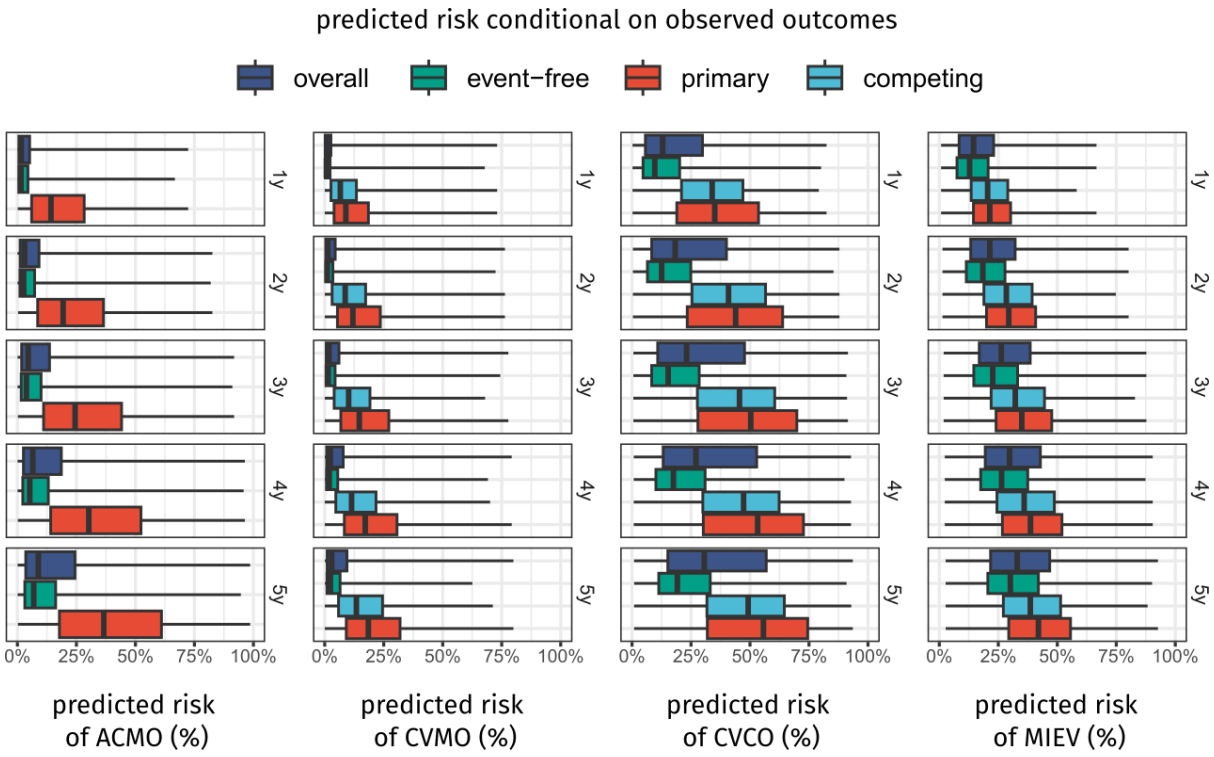
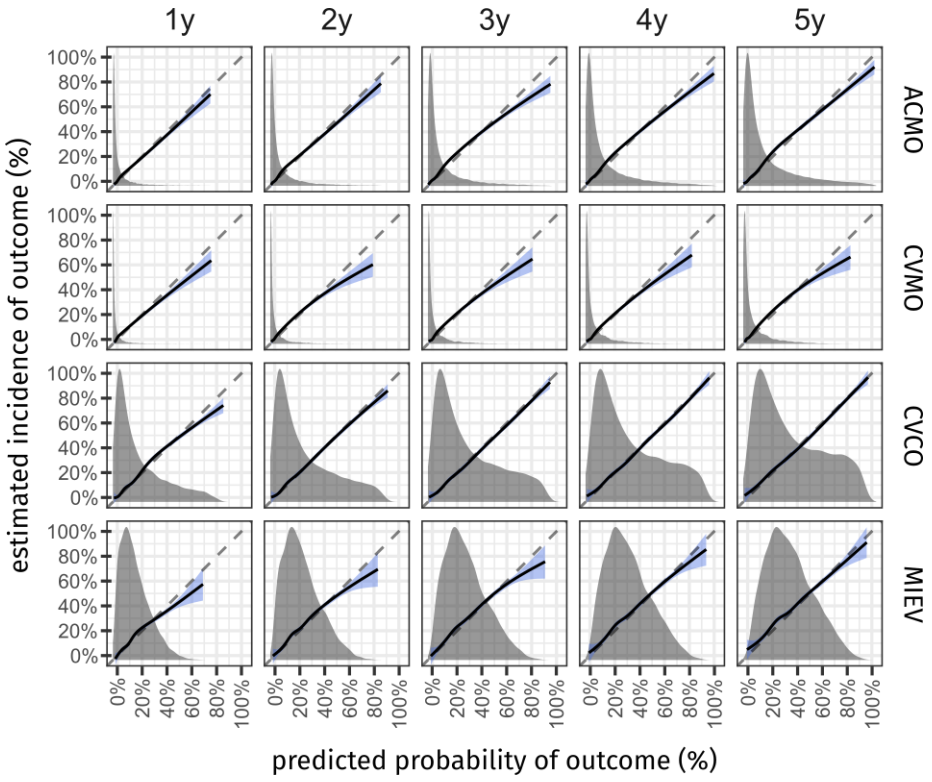


Figure 2: *Visual assessment of test-set discrimination. Boxplots showing the quantiles of predicted t-year risks of the primary endpoints conditional on the observed t-year outcome. The "overall" boxplot shows the observed quantiles in the test set, and boxplots labelled "primary", "competing", and "event-free" are estimated using inverse probability of censoring weighting.*



406

407 ***Figure 3:*** *Visual assessment of test-set calibration. Calibration plots showing t-year risk estimates*
408 *incidences observed in the test set. The regression line is obtained by fitting a natural cubic spline*
409 *with 6 degrees of freedom on pseudo-values obtained through jackknife resampling of the Kaplan-*
410 *Meier (ACMO) or Aalen-Johansen (CVMO, RVSC, and MACO) estimates. The dashed 45-degree*
411 *reference lines indicate perfect calibration. The density curve in the background of each panel shows*
412 *the distribution of the model estimates.*

413

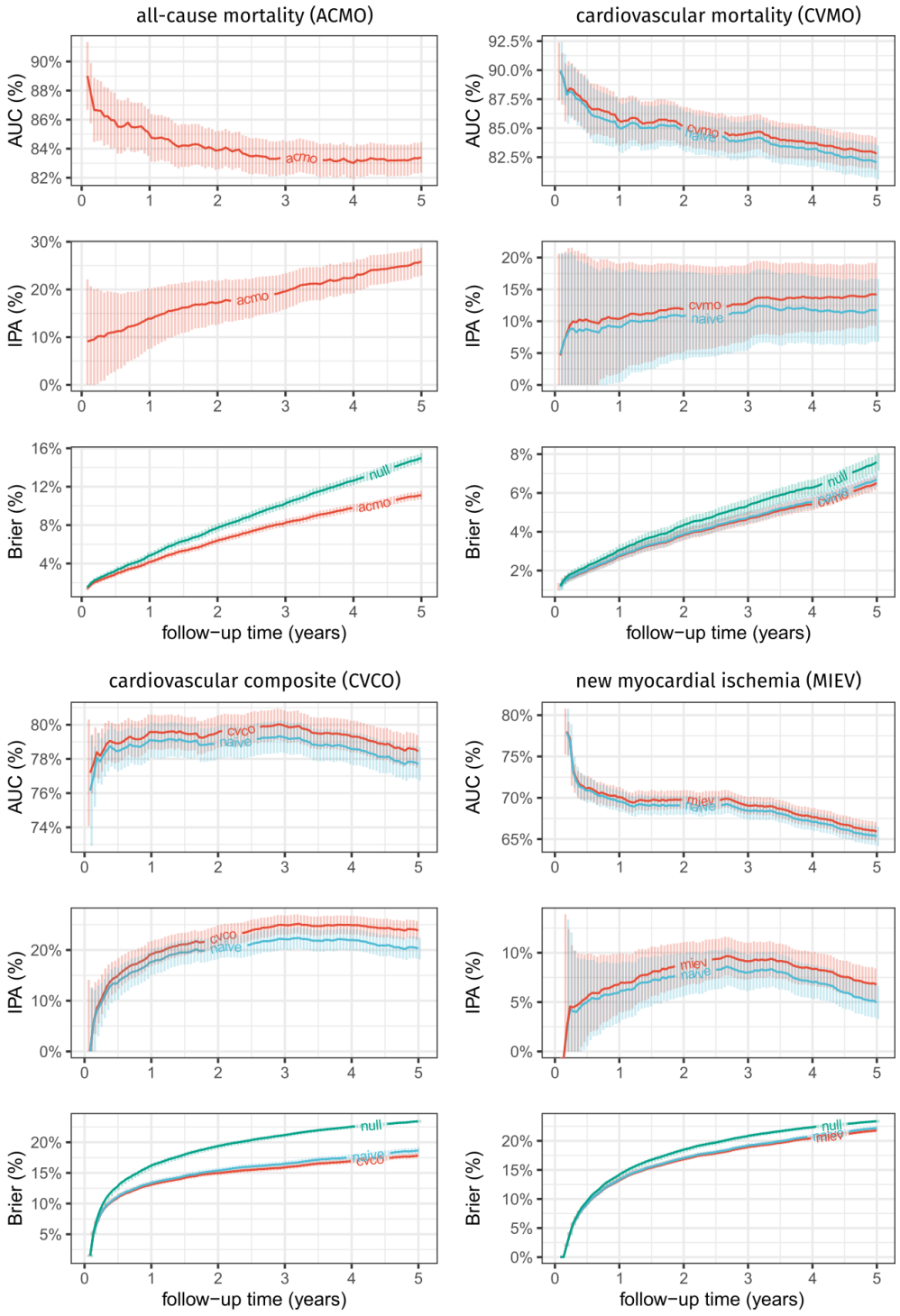


***Figure 4**: Evaluation of test-set performance across the four outcomes. Panels in A, B, C, and D display performance metrics for the all-cause mortality model (ACMO), cardiovascular mortality model (CVMO), de-novo revascularization model (RVSC), and the major complications model (MACO), respectively. Metrics include the time-dependent AUC (AUC), the index of prediction accuracy (IPA), and the Brier score. These metrics are calculated at 100 evenly spaced intervals ranging from 31 days to 10 years. In the Brier score panel, the "null" curve represents the Brier score based on either the Kaplan-Meier estimate for ACMO or the Aalen-Johansen estimate for the primary outcomes in CVMO, RVSC, and MACO. For models with competing risks (CVMO, RVSC, MACO), an additional curve shows the performance of a naïve model that treats competing events as censored.*

References

1. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019 Jan;25(1):44–56.

2. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. N Engl J Med. 2019 Apr 4;380(14):1347–58.

3. Steinfeldt J, Buergel T, Loock L, Kittner P, Ruyoga G, Belzen JU zu, et al. Neural network-based integration of polygenic and clinical information: development and validation of a prediction model for 10-year risk of major adverse cardiac events in the UK Biobank cohort. Lancet Digit Health. 2022 Feb 1;4(2):e84–94.

4. D'Ascenzo F, Filippo OD, 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 Jan 16;397(10270):199–207.

5. Pieszko K, Shanbhag AD, Singh A, Hauser MT, Miller RJH, Liang JX, et al. Time and event-specific deep learning for personalized risk assessment after cardiac perfusion imaging. Npj Digit Med. 2023 May 1;6(1):1–11.

6. Lee C, Zame W, Yoon J, Schaar M van der. DeepHit: A Deep Learning Approach to Survival Analysis With Competing Risks. Proc AAAI Conf Artif Intell [Internet]. 2018 Apr 26 [cited 2023 Sep 8];32(1). Available from: https://ojs.aaai.org/index.php/AAAI/article/view/11842

7. Steele AJ, Denaxas SC, Shah AD, Hemingway H, Luscombe NM. Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease. Singh TR, editor. PLOS ONE. 2018 Aug 31;13(8):e0202344.

8. Gensheimer MF, Narasimhan B. A scalable discrete-time survival model for neural networks. PeerJ. 2019 Jan 25;7:e6257.

9. Nielsen AB, Thorsen-Meyer HC, Belling K, Nielsen AP, Thomas CE, Chmura PJ, et al. Survival prediction in intensive-care units based on aggregation of long-term disease history and acute physiology: a retrospective study of the Danish National Patient Registry and electronic patient records. Lancet Digit Health. 2019 Jun 1;1(2):e78–89.

10. Zhao L, Feng D. Deep Neural Networks for Survival Analysis Using Pseudo Values. IEEE J Biomed Health Inform. 2020 Nov;24(11):3308–14.

11. Kvamme H, Borgan Ø. Continuous and discrete-time survival prediction with neural networks. Lifetime Data Anal. 2021 Oct 1;27(4):710–36.

12. Kvamme H, Borgan Ø, Scheel I. Time-to-Event Prediction with Neural Networks and Cox Regression [Internet]. arXiv; 2019 [cited 2023 Sep 8]. Available from: http://arxiv.org/abs/1907.00825

13. Holm PC, Haue AD, Westergaard D, Röder T, Banasik K, Tragante V, et al. Development and validation of a neural network-based survival model for mortality in ischemic heart disease [Internet]. medRxiv; 2023 [cited 2023 Sep 8]. p. 2023.06.16.23291527. Available from: https://www.medrxiv.org/content/10.1101/2023.06.16.23291527v1

14. Kleinbaum DG, Klein M. Survival Analysis: A Self-Learning Text, Third Edition. 3rd edition. New York, NY: Springer; 2011. 715 p.

15. Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. Eur J Epidemiol. 2014 Aug 1;29(8):541–9.

16. Helweg-Larsen K. The Danish Register of Causes of Death. Scand J Public Health. 2011 Jul;39(7 Suppl):26–9.

17. Kildemoes HW, Sørensen HT, Hallas J. The Danish National Prescription Registry. Scand J Public Health. 2011 Jul;39(7 Suppl):38–41.

18. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol. 2015 Nov;449.

19. Özcan C, Juel K, Flensted Lassen J, von Kappelgaard LM, Mortensen PE, Gislason G. The Danish Heart Registry. Clin Epidemiol. 2016 Oct 25;8:503–8.

20. Schmidt M, Andersen LV, Friis S, Juel K, Gislason G. Data Resource Profile: Danish Heart Statistics. Int J Epidemiol. 2017 Oct 1;46(5):1368–1369g.

21. Tutz G, Schmid M. Modeling Discrete Time-to-Event Data. 1st ed. 2016 edition. New York, NY: Springer; 2016. 257 p.

22. Gorishniy Y, Rubachev I, Khrulkov V, Babenko A. Revisiting Deep Learning Models for Tabular Data [Internet]. arXiv; 2023 [cited 2023 Sep 14]. Available from: http://arxiv.org/abs/2106.11959

23. Loshchilov I, Hutter F. Decoupled Weight Decay Regularization [Internet]. arXiv; 2019 [cited 2023 Sep 14]. Available from: http://arxiv.org/abs/1711.05101

24. Smith LN, Topin N. Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates [Internet]. arXiv; 2018 [cited 2023 Sep 14]. Available from: http://arxiv.org/abs/1708.07120

25. Coleman C, Narayanan D, Kang D, Zhao T, Zhang J, Nardi L, et al. DAWNBench: An End-to-End Deep Learning Benchmark and Competition.

26. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework [Internet]. arXiv; 2019 Jul. Report No.: 1907.10902. Available from: https://arxiv.org/abs/1907.10902

27. Kattan MW, Gerds TA. The index of prediction accuracy: an intuitive measure useful for evaluating risk prediction models. Diagn Progn Res. 2018 May 4;2(1):7.

28. Schumacher M, Graf E, Gerds T. How to Assess Prognostic Models for Survival Data: A Case Study in Oncology. Methods Inf Med. 2003;42(5):564–71.

29. Gerds TA, Kattan MW. Medical risk prediction models: with ties to machine learning. 1st ed. Boca Raton: CRC Press; 2021. (Chapman & hall/crc biostatistics series).

30. Andersen PK, Perme MP. Pseudo-observations in survival analysis. Stat Methods Med Res. 2010 Feb;19(1):71–99.

31. Geloven N van, Giardiello D, Bonneville EF, Teece L, Ramspek CL, Smeden M van, et al. Validation of prediction models in the presence of competing risks: a guide through modern methods. BMJ. 2022 May 24;377:e069249.

32. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library [Internet]. arXiv; 2019 [cited 2023 Sep 18]. Available from: http://arxiv.org/abs/1912.01703

33. Falcon W, The PyTorch Lightning team. PyTorch Lightning [Internet]. 2019 [cited 2023 Sep 18]. Available from: https://github.com/Lightning-AI/lightning

34. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12(85):2825–30.

35. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2019. Available from: https://www.R-project.org/

36. Therneau TM. A Package for Survival Analysis in S.

37. Wickham H. ggplot2. WIREs Comput Stat. 2011;3(2):180–5.

38. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. J Open Source Softw. 2019 Nov 21;4(43):1686.

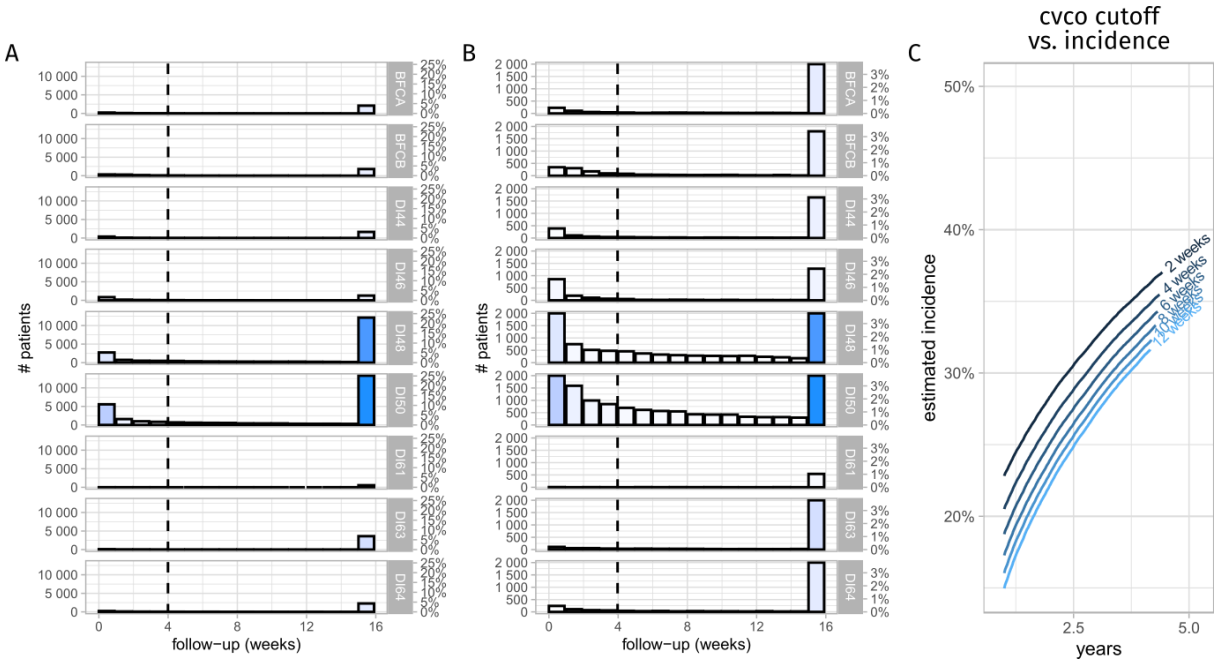
39. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD). Circulation. 2015 Jan 13;131(2):211–9.

40. Aggarwal NR, Patel HN, Mehta LS, Sanghani RM, Lundberg GP, Lewis SJ, et al. Sex Differences in Ischemic Heart Disease. Circ Cardiovasc Qual Outcomes. 2018 Feb;11(2):e004437.

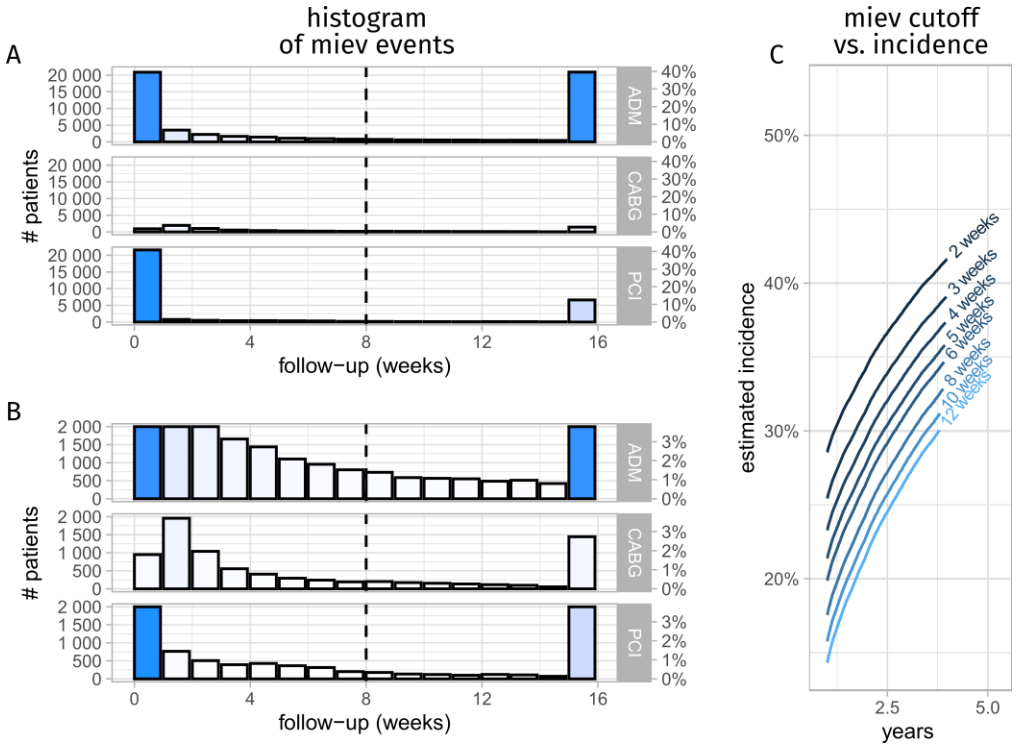
41. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Comput. 1997 Nov;9(8):1735–80.

42. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is All you Need. In: Advances in Neural Information Processing Systems [Internet]. Curran Associates, Inc.; 2017 [cited 2023 Dec 22]. Available from: https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

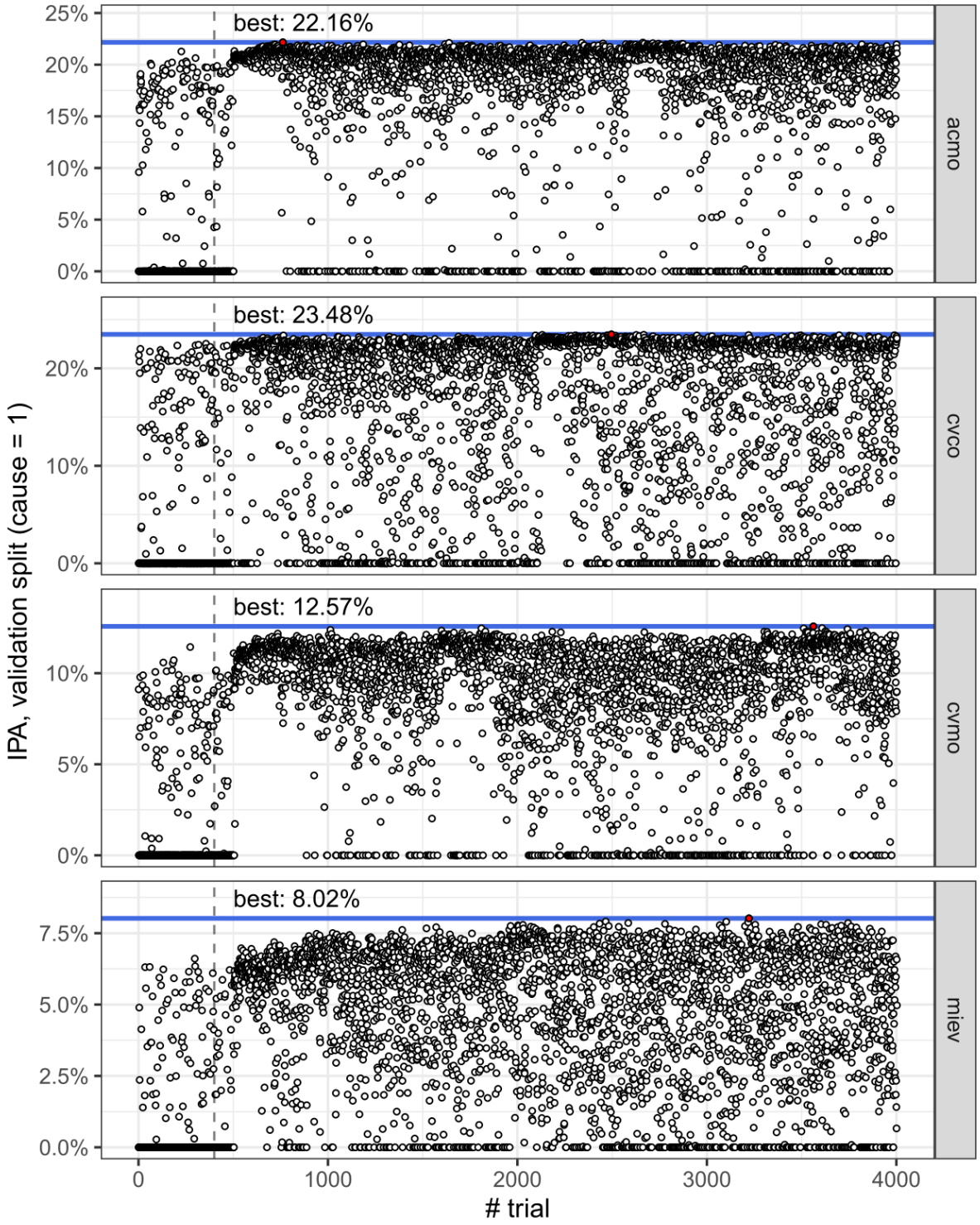
Supplementary figures



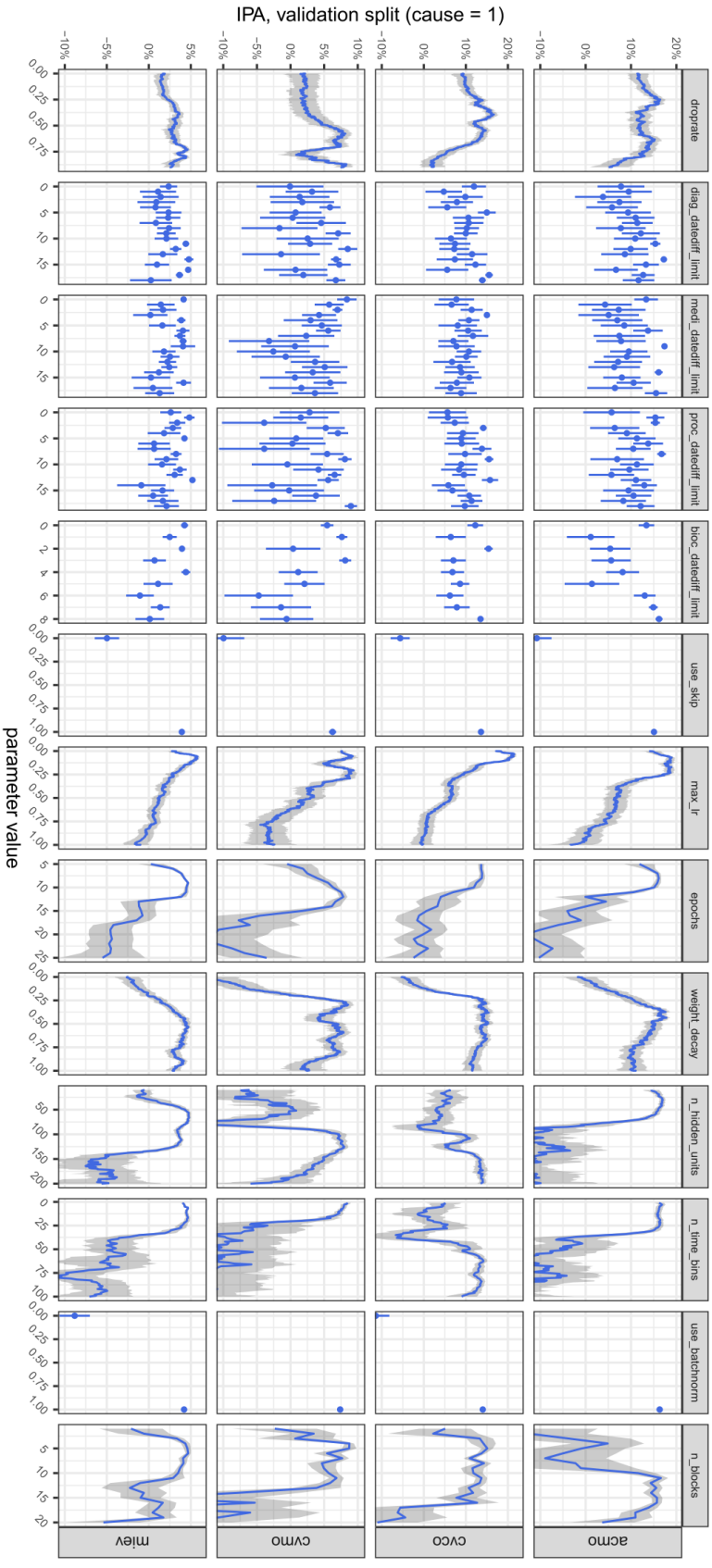
Supp. Figure 1: Time to any of the cardiovascular composite (CVCO) endpoints after index coronary angiography. A) shows the overall counts, and B) provides a zoomed in view of the same data. Bars show the number of unique patients with a CVCO procedure (BFCA, BFCB) or admission (I44, I46, I48, I50, I61, I63, and I64) in each week following the coronary angiography (same day events also included). X-axis have been limited to 16 weeks, and events after that have been aggregated. The dashed vertical lines show the 4-week blanking window used to separate index-related events from progression-related events (the CVCO endpoint). C) Show the effect of the varying the duration of the blanking window.



Supp. Figure 2: Time to recurrent myocardial ischemic events (MIEV), A) shows the overall counts, and B) provides a zoomed in view of the same data. Bars shows the number of unique patients with an MIEV admission (I20-I25) or revascularization procedure (PCI/CABG) in each week following the index coronary angiography. X-axis have been limited to 16 weeks, and events after that have been aggregated. The last bar thus shows how many distinct patients have a future admission with each of the complications. The dashed vertical lines show the 8-week cutoff we use to separate index-related admissions from disease progression complications (the MIEV endpoint). C) Shows the effect of varying the duration of the washout/cutoff period.

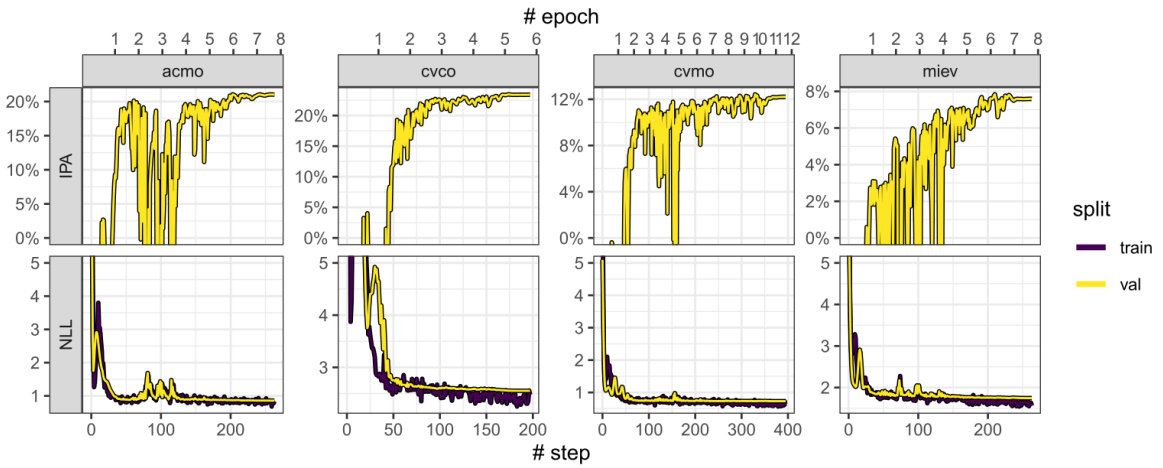


Supp. Figure 3: performance of trials in hyperparameter sweeps for the four different models. The optimization metric is the integrated IPA calculated on a validation split of the training set. Trials with a negative IPA have been clipped to 0%. The blue line shows the performance of the best performing trials. The dashed vertical line shows the transition from random sampler to TPE sampler after the 400 startup trials.



Supp. Figure 4: relationship between hyperparameter values and objective values of trials in the hyperparameter sweeps. Each panel shows the rolling average IPA (±SD) (y-axis) with a sliding window covering one decile of the hyperparameter values (x-axis).

.



Supp. Figure 5: Training history for the final models using the fine-tuned hyperparameters. Plot shows the negative log-likelihood (NLL) and cause-specific integrated IPA (IPA) calculated after each batch (step).

Supplementary tables

Hyperparameter	Search space	ACMO	CVMO	CVCO	MIEV
Enable skip connections	True or False	True	True	True	True
Enable batch normalization	True or False	True	True	True	True
Maximum Learning Rate	1e-3 to 1	0.204	0.072	0.064	0.076
Number of Epochs	5 to 25	8	12	6	8
Weight Decay	1e-5 to 1	0.3774	0.5269	0.2412	0.7359
Dropout rate	0% to 90%	28.4%	86.9%	25.9%	78.7%
Number of ResBlocks	1 to 20	15	5	6	7
Number of hidden units in ResBlocks	10 to 100	30	105	147	104
Number of time bins	1 to 60	23	6	84	9
Biochemical cutoff	0.5 to 5 years (stepsize: 0.5)	4.5	1	4.5	1.5
Diagnosis cutoff	0.5 to 10 years (stepsize: 0.5)	7.5	7.5	9.5	6.5
Medication cutoff	0.5 to 10 years (stepsize: 0.5)	5	1.5	2	4
Procedure cutoff	0.5 to 10 years (stepsize: 0.5)	7.5	5	2	4.5

Supp. Table 1: Overview of hyperparameters, the search space explored, and the best performing configurations found following hyperparameter optimization.

Supplementary Methods

Inclusion criteria

From the Eastern Danish Heart Registry (PATS), we identified all coronary angiographies where a coronary pathology of diffuse atheromatosis or 1/2/3-vessel disease was diagnosed, and then ascertained that patients had a prior diagnosis code (ICD10) of I20-25 (ischemic heart disease), R07.4 (chest pain), Z03.4 (suspected myocardial infarction), or Z03.5C (suspected stable angina) using the Danish National Patient Registry (DNPR). We then limited the coronary angiographies to those performed between 1st of January 2006 and 31st of December 2016. All but the first of the remaining procedures for each patient were then discarded and we limited the study population to individuals above 18 years of age. Finally, we discarded patients that were not alive two days after the procedure.

Endpoint definition

All-cause mortality (ACMO) was defined from the "t_person" table in the CPR where it is recorded as "C_STATUS = 90" and used "D_STATUS_HEN_START" as the event date. For patients recorded as being alive, "C_STATUS = 0", we used the last update date of the CPR as the censoring date.

Cardiovascular mortality (CVMO) was defined from the Cause of Death Register (DAR) as any death with an ICD-10 code of I00-99 registered as the underlying cause of death. We used the date from "D_STATDATO" to determine the date. We used the latest update date of the DAR, 2019-02-15, as the censoring date. Deaths from other causes were treated as competing risks.

Cardiovascular complications (CVCO) were defined from the National Patient Register (DNPR) and included hospital admissions and procedures. The hospital admissions) of interest were here any in-patient admission ("C_PATTYPE = 0") a primary diagnosis of "heart failure" (ICD-10: I50), "atrial fibrillation or flutter" (I48), "cardiac arrest" (I46), or "cerebrovascular accident" (I61, I63-64). In addition, procedure codes corresponding to implantation of pacemaker (SKS: BFCA0*) or cardioverter-defibrillator (BFCB0*) was also included. To differentiate between admissions and procedures related to the index coronary angiography and unplanned admissions and procedures that represent disease progression, we introduced a wash-out or blanking period, which we set to 4-weeks (Fig S1).

Recurrent myocardial ischemia events (MIEV) were likewise defined from in-hospital admissions registered in the DNPR. We defined recurrent events as a) hospitalizations with a duration longer than 24 hours with a primary diagnosis of ischemic heart disease (ICD-10: I20-25) and b) unplanned

percutaneous coronary intervention (PCI) and coronary bypass grafting (CABG) procedures (PCI: "^KFNG.*" and CABG: "^KFN[A-E].*") as recorded in the "t_sksopr" table. For a) we used admission date + 1 day as the event time and for b) we used the procedure date. Similar to the CVCO outcome, we wanted to differentiate between admissions directly related to the initial coronary angiography and unplanned ones representing disease progression. As a pragmatic approach, we decided on using a blanking-window here used an 8-week cutoff (Fig S3). All events prior to that cutoff were therefore ignored, and the earliest admission with any of the MIEV events was used as the primary endpoint.

Hyperparameter optimization

We performed hyperparameter optimization using the Optuna hyperparameter optimization package for python (26). For hyperparameter sweeps, we ran 4,000 different trials using the Tree-structured Parzen Estimator sampling algorithm included in Optuna. We used 400 startup trials and otherwise relied on the default settings of Optuna. Since we constrained our models to only train for a maximum of 25 epochs and training thus did not take long, we did not set up pruning in any of our sweeps.

Statistics

Calibration plots were generated by fitting natural cubic splines with six degrees of freedom to the model's estimates and pseudo-values, which were derived via jackknife estimation of the marginal cumulative incidence. This methodology is described in BMJ 2022;377:e069249 and DOI: 10.1177/0962280209105020). To aid interpretation, we overlaid the regression curve on a kernel density estimate of the model predictions at the specific prediction horizons.