Comparison of State-Of-The-Art Neural Network Survival Models With The Pooled Cohort Equations for Cardiovascular Disease Risk Prediction

Yu Deng
Northwestern University
Lei Liu
Washington University in St. Louis
Hongmei Jiang
Northwestern University
Yifan Peng
Cornell University
Yishu wei
Northwestern University
Yizhen Zhong
Northwestern University
Yun Zhao
University of California, Santa Barbara
Xiaoyun Yang
Northwestern University
Zhiyong Lu
National Center for Biotechnology Information
Abel kho
Northwestern University
Hongyan Ning
Northwestern University
Norrina Allen
Northwestern University
John Wilkins
Northwestern University
Kiang Liu
Northwestern University
Donald Lloyd-Jones
Northwestern University
Lihui Zhao (lihui.zhao@northwestern.edu)

Northwestern University

Jingzhi Yu

Northwestern University

Zhiyang zhou

University of Manitoba

Research Article

Keywords: Artificial intelligence, Cardiovascular disease, Cox regression, Deep learning, Machine learning, Neural network, Pooled Cohort Equations, Predictive modeling, Survival analysis

Posted Date: October 27th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-958135/v1

Comparison of State-of-the-Art Neural Network Survival Models with the Pooled Cohort Equations for Cardiovascular Disease Risk Prediction

Yu Deng[1], Lei Liu[2], Hongmei Jiang[3], Yifan Peng[4], Yishu Wei[3], Zhiyang Zhou[5], Yizhen Zhong[1], Yun Zhao[6], Xiaoyun Yang[1], Jingzhi Yu[1], Zhiyong Lu[7], Abel Kho[1], Hongyan Ning[1], Norrina B. Allen[1], John Wilkins[1], Kiang Liu[1], Donald Lloyd-Jones[1], Lihui Zhao[1]

[1]Feinberg School of Medicine, Northwestern University, Chicago, USA

[2]Division of Biostatistics, Washington University in St. Louis, St. Louis, USA

[3]Department of Statistics, Northwestern University, Chicago, USA

[4]Department of Population Health Sciences, Cornell University, Ithaca, USA

[5]Department of Statistics, University of Manitoba, Manitoba, Canada

[6]Department of Computer Science, University of California, Santa Barbara

[7]National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM), National Institute of Health (NIH), Bethesda, USA

Corresponding author: Lihui Zhao

Abstract

Background: The Pooled Cohort Equations (PCEs) are race- and sex-specific Cox PH-based models used for 10-year atherosclerotic cardiovascular disease (ASCVD) risk prediction with acceptable discrimination. In recent years, neural network models have gained increasing popularity with their success in image recognition and text classification. Various survival neural network models have been proposed by combining survival analysis and neural network architecture to take advantage of the strengths from both. However, the performance of these survival neural network models compared to each other and to PCEs in ASCVD prediction is unknown.

Methods: In this study, we used 6 cohorts from the Lifetime Risk Pooling Project and compared the performance of the PCEs in 10-year ASCVD risk prediction with an all two-way interactions Cox PH model (Cox PH-TWI) and three state-of-the-art neural network survival models including Nnet-survival, Deepsurv, and Cox-nnet. For all the models, we used the same 7 covariates as used in the PCEs. We fitted each of the aforementioned models in white females, white males, black females, and black males, respectively. We evaluated models' internal and external discrimination power and calibration.

Results: The training/internal validation sample comprised 23246 individuals. The average age at baseline was 57.8 years old (SD = 9.6); 16% developed ASCVD during average follow-up of 10.50 (SD = 3.02) years. Based on 10x10 cross-validation, the method that had the highest C-statistics was Cox PH-TWI (0.7372) for white males, PCE (0.7973) for white females, Cox PH-TWI (0.6989) for black males, and Deepsurv (0.7874) for black females. In the external validation dataset, PCE (0.7102), Deepsurv (0.7293), PCE (0.6907), and Nnet-survival (0.7243) had the highest C-statistics for white male, white female, black male, and black female

population, respectively. Calibration plots showed that in 10x10 validation, PCE had good calibration in white male, white female, black male but was outperformed by Deepsurv in black female. In external validation, all models overestimated the risk for 10-year ASCVD except for Deepsurv in black female.

Conclusions

We demonstrated the use of the state-of-the-art neural network survival models in ASCVD risk prediction. Neural network survival models and PCEs have generally comparable discrimination and calibration.

Keywords

Artificial intelligence, Cardiovascular disease, Cox regression, Deep learning, Machine learning, Neural network, Pooled Cohort Equations, Predictive modeling, Survival analysis

Background

Cox Proportional Hazards (Cox PH) model is widely used to quantify the effect of covariates in relation to time-to-event outcomes or to predict the survival time for a new individual [1]. Cox PH is a semi-parametric model, which consists of two main components: baseline hazard and hazard ratio. The estimates of its coefficients are obtained through optimization of the partial likelihood function, which depends on both censored and uncensored individuals.

With the availability of large datasets and high-speed computational power, neural network algorithms have become increasingly popular. Neural networks have been successful when applied to unstructured data such as image recognition and text classification [2–5]. Compared to Cox PH, standard neural network architectures focus on predicting outcomes as a binary classification problem at a specific follow-up point. However, it is common in medical studies that individuals are lost to follow-up (censored data) before the failure or event time. Standard neural network models cannot train or test on these individuals, which leads to sample size reduction. In 1995, Faraggi-Simon first combined neural network architectures with the Cox PH model to make use of censored information as well as to model non-linear features-outcome relations [6]. Since then, there has been increasing interest in incorporating neural network architectures in survival analysis. In current literature, there are two main ways of modeling time-to-event using neural networks: (i) adapting Cox PH model and using partial likelihood loss, e.g., Cox-nnet [7] and Deepsurv [8]; or (ii) discretizing survival time and using a heuristic loss function, e.g., Nnet-survival [9].

Atherosclerotic cardiovascular disease (ASCVD) is the leading cause of death globally [10]. Currently, some commonly used prediction models for ASCVD are based on Cox PH, such as the Framingham CHD risk score and its derivatives [11]. In recent years, the American College

of Cardiology (ACC)/American Heart Association (AHA) guidelines developed new equations, i.e., the Pooled Cohort Equations (PCEs), to estimate 10-year ASCVD risk in non-Hispanic whites and African Americans [12]. The equations are developed based on datasets from several community-based epidemiology cohort studies. The PCEs are four race-, sex-specific and Cox PH based models. It is unclear whether neural network survival models can outperform PCEs for 10-year ASCVD risk prediction. In addition, it is unclear how different architectures of neural network survival models perform compared to each other. In this study, we compared the four race- and sex-specific PCEs with race- and sex-specific state-of-the-art neural network survival models: Nnet-survival, Deepsurv, and Cox-nnet [7–9] in primary ASCVD risk prediction. For fair comparison, we also included Cox PH models with all significant two-way interactions since this enables Cox PH to capture more complex relationships. For all models, we used the same seven predictors as in the PCEs. Our study is the first study to compare the state-of-the-art neural network survival models with PCEs in incident ASCVD prediction.

Methods

Model I, II: Pooled Cohort Equations, all two-way interaction Cox PH

PCEs are four Cox PH based models, each of which is for a specific race and sex group (white male, white female, black male, black female). Cox PH models the probability an individual experiences the event during a small-time interval given the individual is free of an event at the beginning of the time interval [1], which is also known as hazard rate. Specifically, the hazard function can be expressed as the follows:

$$\lambda_i(t) = \lambda_0(t)\exp\left(\beta_1 X_{i1} + \cdots + \beta_p X_{ip}\right) = \lambda_0(t)\exp(\boldsymbol{X}_i^T\boldsymbol{\beta}), \quad (1)$$

where t is the survival time, $\lambda_0(t)$ is the baseline hazard risk at time t, $\boldsymbol{X}_i$ is the covariates for individual i, $\boldsymbol{\beta}$ is the regression coefficient vector. The hazard function consists of two parts: baseline hazard $\lambda_0(t)$ and a hazard ratio or risk function $\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})$. Cox PH assumes that the relative risk for each covariate ($\boldsymbol{\beta}$ in the equation) is constant over time. The estimate of $\boldsymbol{\beta}$ is obtained by optimizing the Cox partial likelihood function as defined below:

$$l(\beta) = \sum_{i:\Delta_i=1}\left(\boldsymbol{X}_i^T\boldsymbol{\beta} - log\sum_{j:Y_j\geq Y_i}\exp\left(\boldsymbol{X}_j^T\boldsymbol{\beta}\right)\right) \quad (2)$$

where Δ_i is the indicator for the occurrence of event, Y_j is follow-up time for individual j.

In the PCEs, seven predictors were selected based on demonstrated statistical utility using prespecified criteria [12]. These predictors include age at baseline, systolic blood pressure (SBP), diabetes medical history, treatment for hypertension, current smoker, high density cholesterol and total cholesterol. The interactions between age at baseline and the other predictors were tested based on p-values. Only interactions that had significant p-values ($p<0.05$) were kept in the model. The PCEs demonstrated acceptable performance in derivation samples, with C-statistics for 10-year risk prediction of 0.80 in white females, 0.76 in white males, 0.81 in black females, and 0.70 in black males in 10x10 cross-validation [12].

To capture more complex non-linear relationships between predictors and ASCVD outcome, in the Cox PH-TWI model, we included all the two-way interactions of the 7 predictors in the model for each race and sex. We then retained only the interaction terms that had significant p-values for each race and sex.

Models III and IV: Deepsurv and Cox-nnet

Deepsurv is an adaptation of the standard Cox PH [8]. Instead of assuming the linear relationship between covariates and log-hazard, the Deepsurv model can automatically learn the non-linear relationship between risk factors and an individual's risk of failure by its linear (i.e., multi-layer perceptron) and non-linear (activation functions) transformation. Specifically, the log-risk function $\boldsymbol{X}_i^T\boldsymbol{\beta}$ in the Cox equation as shown in Eq. (1) is replaced by the output from neural network $h_W(X)$, where w are weights for neural network (see Figure 1A). The neural network optimizes the log-partial likelihood function similar to the standard Cox model by tuning parameters $\boldsymbol{W}$:

$$l(\boldsymbol{W}) = \sum_{i:\Delta_i=1}\left(h_w(\boldsymbol{X}_i) - log\sum_{j:Y_j\geq Y_i}(\exp\left(h_w(\boldsymbol{X}_j)\right)\right) .$$

The Deepsurv model also adapted modern techniques to improve the training of the network. These techniques include standardizing the input, Scaled Exponential Linear Units (SELU) as the activation function, Adaptive Moment Estimation (Adam) for the gradient descent algorithm, Nesterov momentum, and learning rate scheduling [8].

The architecture of Cox-nnet is slightly different than the Deepsurv model (see Figure 1B). Instead of replacing the log-risk function $\boldsymbol{X}_i^T\boldsymbol{\beta}$ in Eq. (1) using the output of neural network as does in the Deepsurv model, Cox-nnet replaces $\boldsymbol{X}_i^T$ in Eq. (1) using the output from the neural network $\boldsymbol{G}(\boldsymbol{W}\boldsymbol{X}_i + \boldsymbol{b})^T$, where $\boldsymbol{W}$ is the coefficient weight matrix between the input and hidden layers, $\boldsymbol{b}$ is the bias term for each hidden node and $\boldsymbol{G}$ is the activation function. Cox-nnet was proposed to deal with high dimensional features especially in genomic studies. To avoid overfitting, Cox-nnet introduces a ridge regularization term and subsequently, the partial log likelihood in Eq. (2) is extended as the following:

$$l(\boldsymbol{\beta},\boldsymbol{W}) = \sum_{i:\Delta_i=1}\left(\boldsymbol{G}(\boldsymbol{W}\boldsymbol{X}_i + \boldsymbol{b})^T\boldsymbol{\beta} - log\sum_{j:Y_j\geq Y_i}\exp\left(\boldsymbol{G}\left(\boldsymbol{W}\boldsymbol{X}_j + \boldsymbol{b}\right)^T\boldsymbol{\beta}\right)\right) + \lambda(\| \boldsymbol{\beta} \|_2 + \| \boldsymbol{W} \|_2),$$

where $\boldsymbol{\beta}$ is the coefficient vector in the original Cox PH model, $\boldsymbol{W}$ is the weight matrix from neural network. In addition to L_2-regularizer, Cox-nnet also allows drop-out for regularization to avoid overfitting. The model is based on Theano framework, therefore, Cox-nnet can be run on a Graphics Processing Unit or multiple threads.

Model V: Nnet-survival

Nnet-survival is a fully parametric survival model that discretizes survival time. Nnet-survival is proposed to improve two main aspects of the neural network model that are adapted from Cox model: computational speed and the violation of the proportional hazard assumption. Neural network survival models that adapt from Cox model (e.g., Deepsurv, Cox-nnet) use partial likelihood function as the loss function to optimize. The partial likelihood function is calculated based on not only the current individual but also all the individuals that are at risk at the time point. This makes it difficult to use stochastic gradient descent or mini-batch gradient descent, both of which use a small subset of the whole dataset. Therefore, both Deepsurv and Cox-nnet may have slow convergence and cannot be applied to large datasets that runs out of memory [9]. Nnet-survival was proposed to discretize time, which transforms the model into a fully parametric model and avoids the use of partial likelihood as the loss function. In Nnet-survival models, follow-up time is discretized to n intervals. Hazard h_j is defined as the conditional probability of surviving time interval j given the individual is alive at the beginning of interval j. Survival probability at the end of interval j can be then calculated as the following:

$$S_j = \prod_{i=1}^{j}(1 - h_i).$$

The loss function is defined as the following:

$$L = h_j \prod_{i=1}^{j-1} \left(1 - h_{(i)}\right),$$

for individuals who failed at interval j, and

$$L = \prod_{i=1}^{j-1} \left(1 - h_{(i)}\right),$$

for individuals who are censored at the second half of interval $j - 1$ or the first half of interval j. There are two main architectures of Nnet-survival: a flexible version and a proportional hazards version. In the flexible version, output layers have m neurons, where m is the number of intervals and each output neuron represents the survival probability at the specific time interval given an individual is alive at the beginning of the time interval. In the proportional hazard version, the final layer only has a single neuron representing $\boldsymbol{X}_i^T \boldsymbol{\beta}$:

$$h_\beta(\boldsymbol{X_i}) = \boldsymbol{X}_i^T \cdot \boldsymbol{\beta}\,,$$

In our study, the flexible version is used, with its architecture of the flexible version shown in Figure 1C.

Statistical analysis

In this study, we used the harmonized, individual-level data from 6 cohorts in the Lifetime Risk Pooling Project, including Atherosclerosis Risk in Communities (ARIC) study, Cardiovascular Health Study (CHS), Framingham Offsprinig study, Coronary Artery Risk Development in Young Adults (CARDIA) study, the Framingham Original study, and the Multi-Ethnic Study of Atherosclerosis (MESA). The first 5 cohort data were used for model development and internal validation, and the MESA data was used for external validation. We included individuals that meet the following criteria: (i) age between 40 to 79; and (ii) free of a previous history of

myocardial infarction, stroke, congestive heart failure, or atrial fibrillation. ASCVD was defined as nonfatal myocardial infarction or coronary heart disease death, or fatal or nonfatal stroke (see [12] for details of selection criteria). All study individuals were free of ASCVD at the beginning of the study and were followed up until the first ASCVD event, loss to follow up, or death, whichever came first. We fit PCE, Cox PH with all two-way interactions (Cox PH-TWI), Nnet-survival, Deepsurv, and Cox-nnet models in white male, white female, black male, and black female participants. For comparison purposes, for all the models, we included the same predictors as used in the PCEs: age at baseline, systolic blood pressure (SBP), diabetes medical history, treatment for hypertension, current smoker, high density cholesterol (HDL-C) and total cholesterol. Individuals who had missing data at baseline were excluded from the study. Individuals who were lost to follow-up were censored.

To obtain high performance neural network survival models, we manually tuned various hyper-parameters including learning rate, number of layers, number of neurons, number of epochs, batch size, momentum, optimizer, learning rate decay, batch normalization, L_2 regularization, and dropout. After selecting the optimal hyper-parameters, we evaluated model performance through internal validation with 10x10 cross validation and external validation with the MESA data. To perform 10x10 cross-validation, we randomly partitioned the pooled cohort data into 10 equal-sized subsamples. Of the 10 subsamples, 9 subsamples were used as training data and the remaining single subsample was retained as the validation data for testing the model. Each of the subsamples is used in turn as the validation data. We repeated this process 10 times, during which 100 models were built. The average C-statistics and calibration plot of the 100 models were used as the final 10x10 cross-validation result. In the calibration plots, the observed and predicted events were shown in deciles [9]. For the external validation, we trained the model in

the whole harmonized dataset (not including MESA cohort), and evaluated the model discrimination and calibration in the external MESA cohort [13].

Nnet-survival, Deepsurv, and Cox-nnet were implemented in python, version 3.7.3. Cox PH model was conducted using the "survival" package in R, version 3.6.0. C-statistics was calculated using the "survC1" package in R, version 3.6.0 [14].

All data were de-identified, and all study protocols and procedures were approved by the Institutional Review Board at Northwestern University with a waiver for informed consent. All methods were performed in accordance with the relevant guidelines and regulations.

Results

Overall, there were 23246 participants, including 8644 white male, 1354 black male, 10719 white female, 2499 black female individuals. The average age at baseline was 57.8 years old (SD = 9.6). Among these individuals, 16.0% developed ASCVD with average follow-up of 10.50 (SD = 3.02) years. The mean SBP value was 127.1 mmHg (SD = 21.0), the mean HDL-C value was 51.6 mg/dL (SD = 16.4), total cholesterol was 217.8 mg/dL (SD = 43.0). For binary predictors, 4.6% individuals had a history of diabetes, 26.0% individuals were current smokers, 31.6 % individuals had treatment for hypertension. The descriptive statistics for each race and sex group were shown in Table 1.

In the MESA external validation dataset, there were 4259 individuals in total. The average age at baseline was 61.6 years old (SD = 9.6). Among the 4259 individuals, 331 (7.77%) developed ASCVD with average follow-up years of 10.97 years old (SD = 2.48). Among these individuals, there were 1194 white male, 799 black male, 1284 white female, and 982 black female. Baseline characteristics of the study sample were shown in Table 1, stratified by sex and race group.

In 10x10 cross validation, in the white male population (see Figure 2), Cox PH-TWI achieved the highest C-statistics (0.7372) among all the models. In the white female population, PCE had the highest C-statistics (0.7973) comparable to Cox PH-TWI (0.7972). In the black male population, Cox PH-TWI had the highest C-statistics. In the black female population, Deepsurv had the highest C-statistics (0.7874) and Nnet-survival had comparable performance (0.7810). The details of C-statistics for each model and race sex group were shown in Supplemental Table 1.

In the external validation dataset, in white male population, PCE had the highest C-statistics (0.7102). In white female population, Deepsurv had the highest C-statistics (0.7293). In black male population, PCE had the highest C-statistics (0.6907). In black female population, Deepsurv (0.7237) and Nnet-survival (0.7243) had the highest and comparable C-statistics.

In terms of calibration in 10x10 cross-validation (see Figure 3), the calibration plot showed that all five models had similar calibration compared to PCE in white male population with Nnet-survival slightly underestimating the survival probability. In white female population, PCE and Nnet-survival had better performance than Deepsurv, Coxnnet, and Cox PH-TWI. In black male population, PCE had better calibration compared to the neural network models and Cox PH-TWI while in the black female population, Nnet-survival and Deepsurv had better calibration than PCE, Cox PH-TWI, and Cox-nnet.

In MESA external validation, calibration plot showed that in the white male population, all five models similarly over-estimated event rate (see Figure 4). In the black male population, Deepsurv and Cox-nnet were closer to the Kaplan Meier estimation compared to Cox PH-TWI, PCE, and Nnet-survival. In the black female population, Deepsurv under-estimated the event rate while all the four other models over-estimated the event rate. In white female population,

Deepsurv seemed to be the closest to the Kaplan Meier prediction. However, in general, all prediction models over-estimated the event compared to Kaplan Meier estimation.

Discussion

In this study, we implemented state-of-the-art neural network survival models in predicting 10-year risk for a first ASCVD event. Our results showed that when using the same predictors as in the PCEs, three survival neural network models and Cox PH-TWI model had similar performance compared to PCEs in both internal and external validation. In the black female population, Deepsurv, Nnet-survival, and Cox PH-TWI outperformed PCE in both internal and external validation. However, in other racial-gender groups, the advantage of neural network models was not obvious. Among different neural network models, the performance of each model was dataset-specific and no obvious trend of which model was better than another was found. However, it is interesting to notice that Cox-nnet and Cox PH-TWI had similar performance across all four race-sex groups in internal and external validation. In terms of calibration, in internal validation dataset, PCEs had good calibration in white male, white female, and black male population. In black female population, Deepsurv outperformed PCEs and other three models. In external validation dataset, all models over-estimated the event rate in all four race-sex groups except for Deepsurv, which overestimated the survival probability in the black female population.

Theoretically, Nnet-survival is faster in training than Deepsurv and Cox-nnet models. Nnet-survival's loss function only relies on individuals in the current minibatch which allows mini-batch gradient descent while both Deepsurv and Cox-nnet require the entire dataset for each

gradient descent update. On the other hand, the discretization of time-to-event in Nnet-survival leads to a less smooth predicted survival curve compared to Deepsurv and Cox-nnet.

In prior studies, Gensheimer et al applied Cox PH, Nnet-survival, Deepsurv, and Cox-nnet in life expectancy prediction using the Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT) dataset [9]. The dataset consisted of 9105 individuals and 39 predictors. The four neural network survival models generated similar C-statistics compared to the Cox PH model, which was consistent with our findings in ASCVD prediction. Both the SUPPORT dataset and our dataset had low dimension number of predictors. Several studies explored other machine learning methods for CVD prediction. Joo et al [15] applied logistic regression, deep neural networks, random forests, and LightGBM to predict CVD as a binary outcome using the Korean National Health Insurance Service–National Health Sample Cohort dataset. The authors found that deep neural network had better performance (C-statistics = 0.7446) compared to the PCE (C-index = 0.7381) in that cohort. However, the ML models used more predictors (hemoglobin level, diastolic blood pressure, presence of proteinuria, serum aspartate aminotransferase, serum alanine aminotransferase, and total cholesterol) compared to the PCE. In another study, Dimopoulos et al implemented KNN, random forest, and decision tree to predict CVD compared to the HellenicSCORE, a Cox regression based model [16]. Their results showed that ML models have comparable performance compared to the HellenicSCORE [17] using 5 and 13 same predictors respectively but were not able to outperform the baseline model.

Similar to other machine learning models, neural network models often show advantage in modeling non-linear complex relationships between predictors and outcome, which is likely to be the case for survival neural network models. In a clinical setting for CVD prediction, a simple

model such as PCE using routinely collected data elements is often desired for a clinician to calculate an individual's future risk easily and with acceptable predictive accuracy for a certain disease. Even though survival neural network models showed comparable performance to Cox PH model in our study, we expect it to be a powerful tool in CVD prediction when more abundant data and more complex repeated measures data become available (e.g. electronic health records data) in the future.

Limitations

Our study has several limitations. First, the cohorts we used from the Lifetime Risk Pooling Project were the same cohorts used in the derivation of the PCEs. This may have led to some optimism in the performance of the PCEs. Second, the participants of our external validation cohort, MESA, were perhaps healthier than the general population. More importantly, they received intensive screening for subclinical CVD, which influenced health behaviors and preventive interventions including use of effective drug therapies; this may result in the lower event rate in MESA participants than what would have been predicted because of the use of effective preventive therapies selectively in higher-risk individuals.

Conclusion

Neural network survival models can achieve comparable discrimination if not superior performance compared to the PCEs in 10-year time-to-ASCVD prediction in the white female, white male, black female, and black male population in our dataset. In future studies, high dimensional features or longitudinal data should be considered to fully explore the benefits of neural network survival models.

Tables

Table 1. Baseline characteristics for each race and sex group in training/internal validation dataset and external validation dataset.

	Overall	White male	White female	Black male	Black female
Training/internal validation dataset					
N	23246	8644	1354	10719	2499
Age (year), mean (SD)	57.8 (9.6)	58.0 (9.6)	57.3 (9.5)	58.0 (9.7)	56.4 (9.3)
SBP (mm Hg), mean (SD)	127.1 (21.0)	127.2 (19.5)	131.8 (21.5)	125.5 (21.6)	130.8 (22.7)
HDL-C (mg/dL), mean (SD)	51.6 (16.4)	43.8 (12.7)	49.7 (16.0)	56.8 (16.5)	57.5 (16.4)
TOTCHL (mg/dL), mean (SD)	217.8 (43.0)	212.1 (39.9)	208.6 (44.4)	223.9 (43.8)	216.0 (45.4)
HXDIAB, n (%)	1069 (4.6)	295 (3.4)	175 (12.9)	294 (2.7)	305 (12.2)
Smoker, n (%)	6038 (26.0)	2294 (26.5)	441 (32.6)	2723 (25.4)	577 (23.1)
RXHYP, n (%)	7340 (31.6)	2226 (25.8)	707 (52.2)	2951 (27.5)	1442 (57.7)
MESA external validation dataset					
N	4259	1194	799	1284	982
Age (year), mean (SD)	61.6 (9.6)	61.9 (9.6)	61.5 (9.6)	61.5 (9.6)	61.3 (9.4)
SBP (mmHg), mean (SD)	126.3 (21.0)	123.7 (18.3)	130.0 (19.2)	121.8 (21.4)	132.4 (22.8)
HDL-C (mm/dL), mean (SD)	52.2 (15.5)	45.2 (12.1)	46.5 (12.5)	58.8 (15.8)	56.9 (15.6)

TOTCHL (mm/dL), mean (SD)	193.3 (35.7)	189.2 (34.4)	182.0 (34.6)	202.3 (34.4)	195.7 (36.5)
HXDIAB, n (%)	354 (8.3)	57 (4.8)	117 (14.6)	51 (4.0)	129 (13.1)
Smoker, n (%)	628 (14.7)	137 (11.5)	166 (20.8)	160 (12.5)	165 (16.8)
RXHYP, n (%)	1676 (39.4)	389 (32.6)	370 (46.3)	402 (31.3)	515 (52.4)

Abbreviations: SBP, systolic blood pressure; HDL-C, high density cholesterol; TOTCHL, total cholesterol; HXDIAB, history of diabetes; RXHYP, history of hypertension.

Figures

Figure 1. Frameworks for neural network survival models. The frameworks of Deepsurv, Cox-nnet, and Nnet-survival are shown in A, B, and C, respectively. In Figure 1A, the Deepsurv model outputs $h_w(\boldsymbol{X}_i)$ which is used to replace the log risk $\boldsymbol{X}_i^T\boldsymbol{\beta}$ in the Cox model. In Figure 1B, the Cox-nnet model outputs $\boldsymbol{\theta}_i$ which replaces the linear predictors $\boldsymbol{X_i}$ in the cox model. In Figure 1C, the output layers generate h_j^i which is the hazard for individual i at time j.

Figure 2. C-statistics for PCEs, Nnet-survival, Deepsurv, Cox-nnet, and Cox PH-TWI in 10x10 cross-validation and MESA external validation. The 'x' markers represent C-statistics in 10x10 cross-validation, the 'o' markers represent C-statistics in MESA external validation.

Figure 3. Kaplan-Meier Observed Event Rate and Predicted Event Rate for the ASCVD Outcome in the 10x10 cross-validation. For each model, we divided participants into 10 group (decile) based on their sorted predicted event probability. Then, for each decile, mean observed event rate (Kaplan-Meier method) was plotted against mean predicted event rate. In a perfectly calibrated model, the predicted event rate would be the same as the observed event rate in each decile. This means that all points would be clustered around the blue identity line.

Figure 4. Kaplan-Meier Observed Event Rate and Predicted Event Rate for the ASCVD Outcome in the MESA Cohort. For each model, we divided participants into 10 group (decile)

based on their sorted predicted event rate. Then, for each decile, mean observed event rate (Kaplan-Meier method) was plotted against mean predicted event rate. In a perfectly calibrated model, the predicted event rate would be the same as the observed event rate in each decile. This means that all points would be clustered around the dotted identity line.

List of Abbreviations:

PCEs: Pooled Cohort Equations

ASCVD: atherosclerotic cardiovascular disease

Cox PH: Cox Proportional Hazards

American College of Cardiology: ACC

AHA: American Heart Association

ARIC: Atherosclerosis Risk in Communities study

CHS: Cardiovascular Health Study

CARDIA: Coronary Artery Risk Development in Young Adults

SBP: systolic blood pressure

Declarations

Ethics approval and consent to participate

All data were de-identified, and all study protocols and procedures were approved by the Institutional Review Board at Northwestern University with a waiver for informed consent.

Consent for publication

N/A

Availability of data and materials

N/A

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by NIH R01: R01HL136942 and the NIH Intramural Research Program, National Library of Medicine.

Authors' contributions

YD led the study, performed all data analyses, and wrote the manuscript. LZ designed and supervised the project. LL and HJ provided statistical expertise on data cleaning and model evaluation. YP provided deep learning expertise on tuning the deep learning models. HN made substantial contribution to the data acquisition. All the other authors read, edited, and approved the final manuscript.

Acknowledgements

NA

References

1. Cox DR. Analysis of survival data. CRC Press; 1984.

2. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.; 2012. https://papers.nips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html. Accessed 17 Aug 2021.

3. Zeng Z, Deng Y, Li X, Naumann T, Luo Y. Natural Language Processing for EHR-Based Computational Phenotyping. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2019;16:139–53.

4. Zhao Y, Hong Q, Zhang X, Deng Y, Wang Y, Petzold L. BERTSurv: BERT-Based Survival Models for Predicting Outcomes of Trauma Patients. arXiv:210310928 [cs]. 2021. http://arxiv.org/abs/2103.10928. Accessed 17 Aug 2021.

5. 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. Curran Associates, Inc.; 2017. https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html. Accessed 17 Aug 2021.

6. Faraggi D, Simon R. A neural network model for survival data. Statistics in Medicine. 1995;14:73–82.

7. Ching T, Zhu X, Garmire LX. Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. PLOS Computational Biology. 2018;14:e1006076.

8. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Medical Research Methodology. 2018;18:24.

9. Gensheimer MF, Narasimhan B. A Scalable Discrete-Time Survival Model for Neural Networks. PeerJ. 2019;7:e6257.

10. Cardiovascular diseases. https://www.who.int/westernpacific/health-topics/cardiovascular-diseases. Accessed 17 Aug 2021.

11. D'Agostino RB Sr, Grundy S, Sullivan LM, Wilson P, for the CHD Risk Prediction Group. Validation of the Framingham Coronary Heart Disease Prediction Scores: Results of a Multiple Ethnic Groups Investigation. JAMA. 2001;286:180–7.

12. Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, et al. 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk. Circulation. 2014;129 25_suppl_2:S49–73.

13. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for Evaluating Overall Adequacy of Risk Prediction Procedures with Censored Survival Data. Stat Med. 2011;30:1105–17.

14. survC1-package: C-Statistics for Risk Prediction Models with Censored... in survC1: C-Statistics for Risk Prediction Models with Censored Survival Data. https://rdrr.io/cran/survC1/man/survC1-package.html. Accessed 17 Aug 2021.

15. Joo G, Song Y, Im H, Park J. Clinical Implication of Machine Learning in Predicting the Occurrence of Cardiovascular Disease Using Big Data (Nationwide Cohort Data in Korea). IEEE Access. 2020;8:157643–53.

16. Dimopoulos AC, Nikolaidou M, Caballero FF, Engchuan W, Sanchez-Niubo A, Arndt H, et al. Machine learning methodologies versus cardiovascular risk scores, in predicting disease risk. BMC Medical Research Methodology. 2018;18:179.

17. Panagiotakos DB, Fitzgerald AP, Pitsavos C, Pipilis A, Graham I, Stefanadis C. Statistical modelling of 10-year fatal cardiovascular disease risk in Greece: the HellenicSCORE (a calibration of the ESC SCORE project). Hellenic J Cardiol. 2007;48:55–63.

Figures

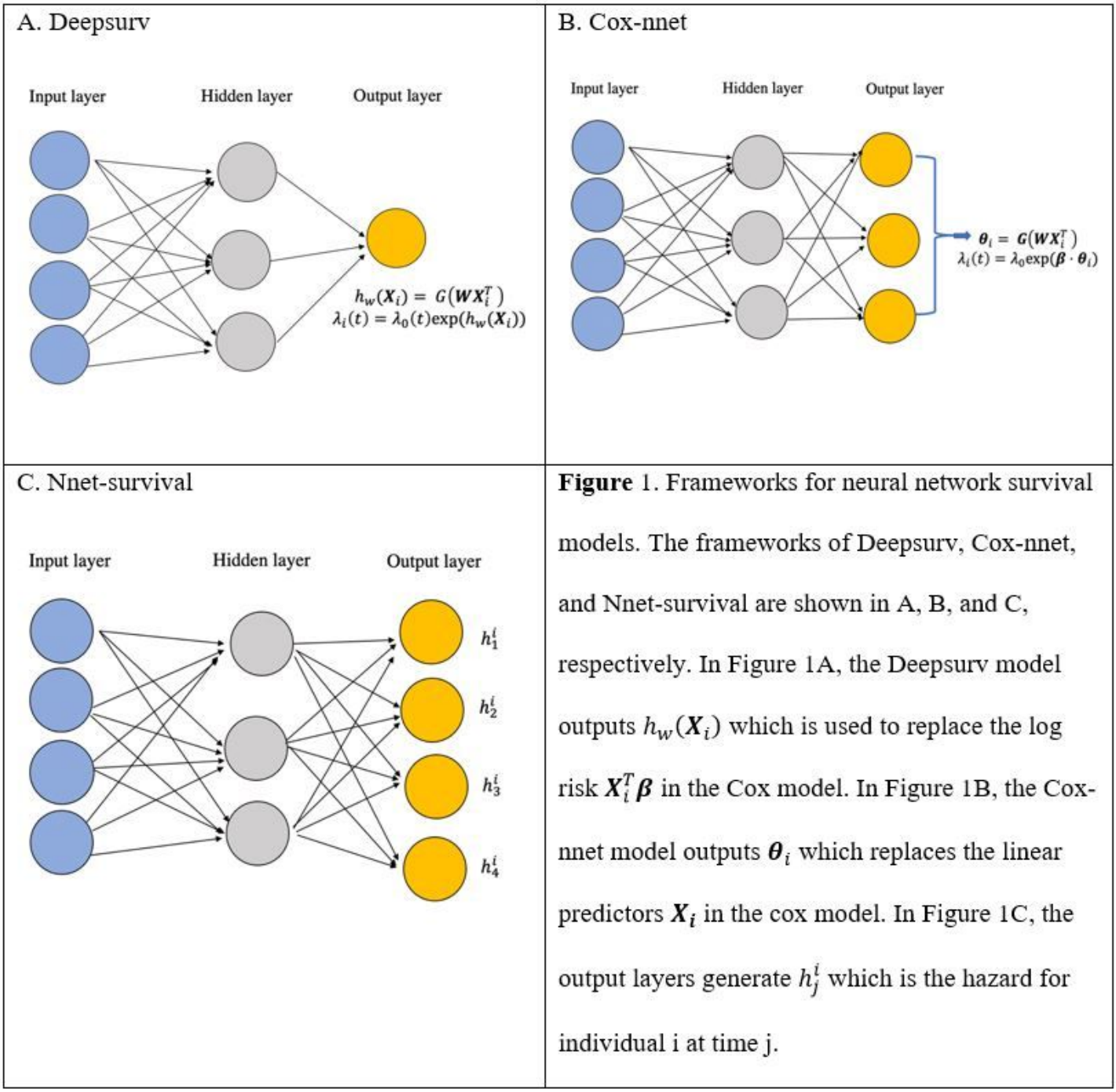

Figure 1. Frameworks for neural network survival models. The frameworks of Deepsurv, Cox-nnet, and Nnet-survival are shown in A, B, and C, respectively. In Figure 1A, the Deepsurv model outputs $h_w(\boldsymbol{X}_i)$ which is used to replace the log risk $\boldsymbol{X}_i^T\boldsymbol{\beta}$ in the Cox model. In Figure 1B, the Cox-nnet model outputs $\boldsymbol{\theta}_i$ which replaces the linear predictors $\boldsymbol{X}_i$ in the cox model. In Figure 1C, the output layers generate h_j^i which is the hazard for individual i at time j.

Figure 1

Frameworks for neural network survival models. The frameworks of Deepsurv, Cox-nnet, and Nnet-survival are shown in A, B, and C, respectively. In Figure 1A, the Deepsurv model outputs h_w (X_i) which is used to replace the log risk X_i^T β in the Cox model. In Figure 1B, the Cox-nnet model outputs θ_i which replaces the linear predictors X_i in the cox model. In Figure 1C, the output layers generate h_j^i which is the hazard for individual i at time j.

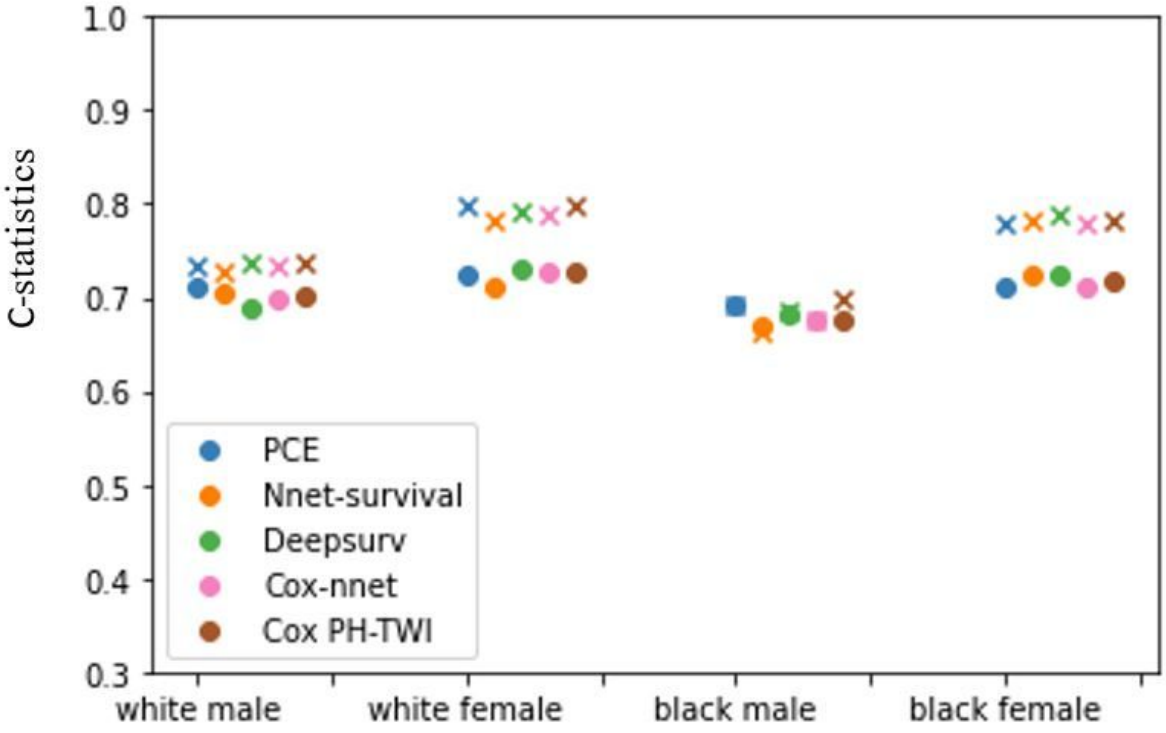

Figure 2

C-statistics for PCEs, Nnet-survival, Deepsurv, Cox-nnet, and Cox PH-TWI in 10x10 cross-validation and MESA external validation. The ‘x’ markers represent C-statistics in 10x10 cross-validation, the ‘o’ markers represent C-statistics in MESA external validation.

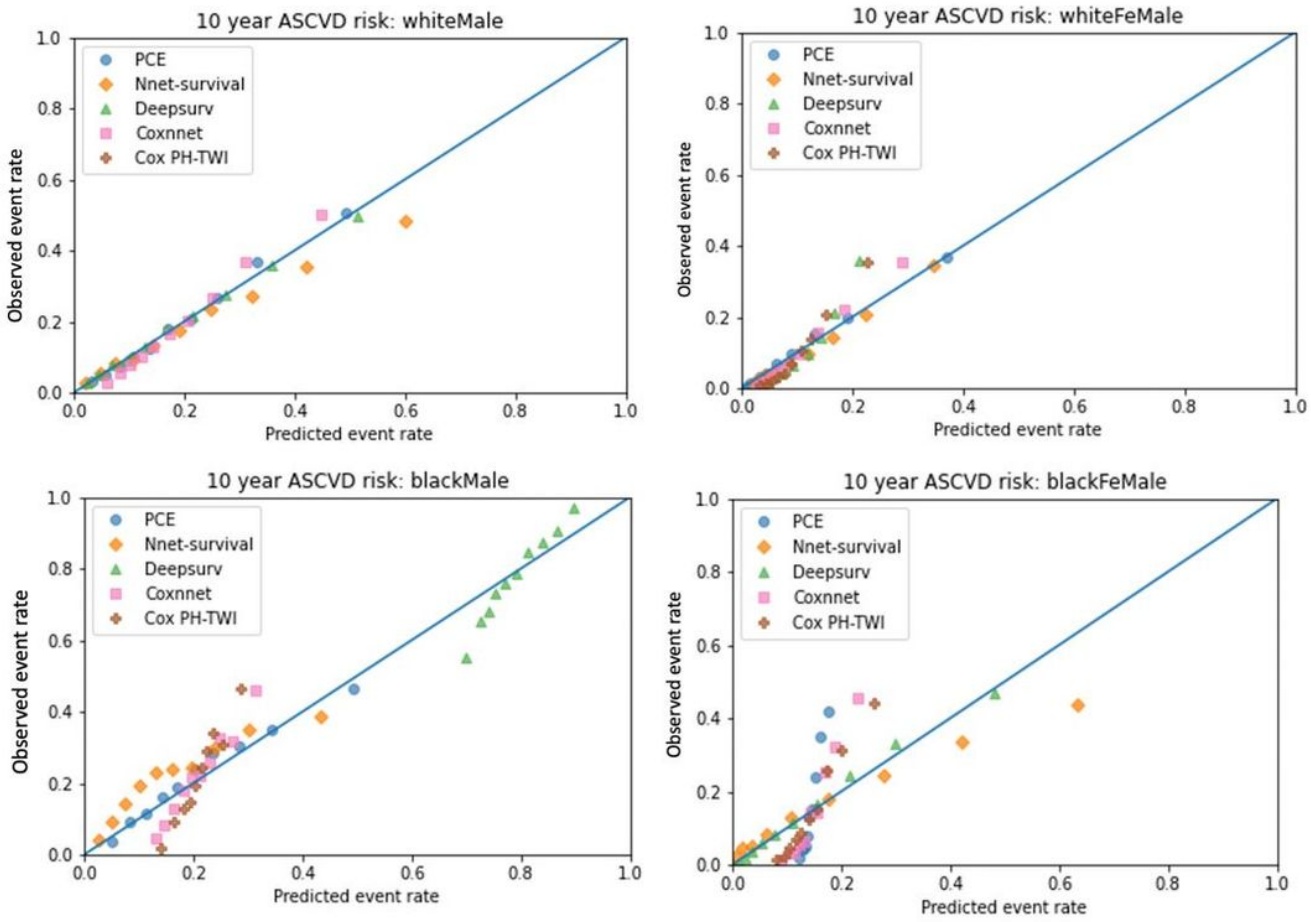

Figure 3

Kaplan-Meier Observed Event Rate and Predicted Event Rate for the ASCVD Outcome in the 10x10 cross-validation. For each model, we divided participants into 10 group (decile) based on their sorted predicted event probability. Then, for each decile, mean observed event rate (Kaplan-Meier method) was plotted against mean predicted event rate. In a perfectly calibrated model, the predicted event rate would be the same as the observed event rate in each decile. This means that all points would be clustered around the blue identity line.

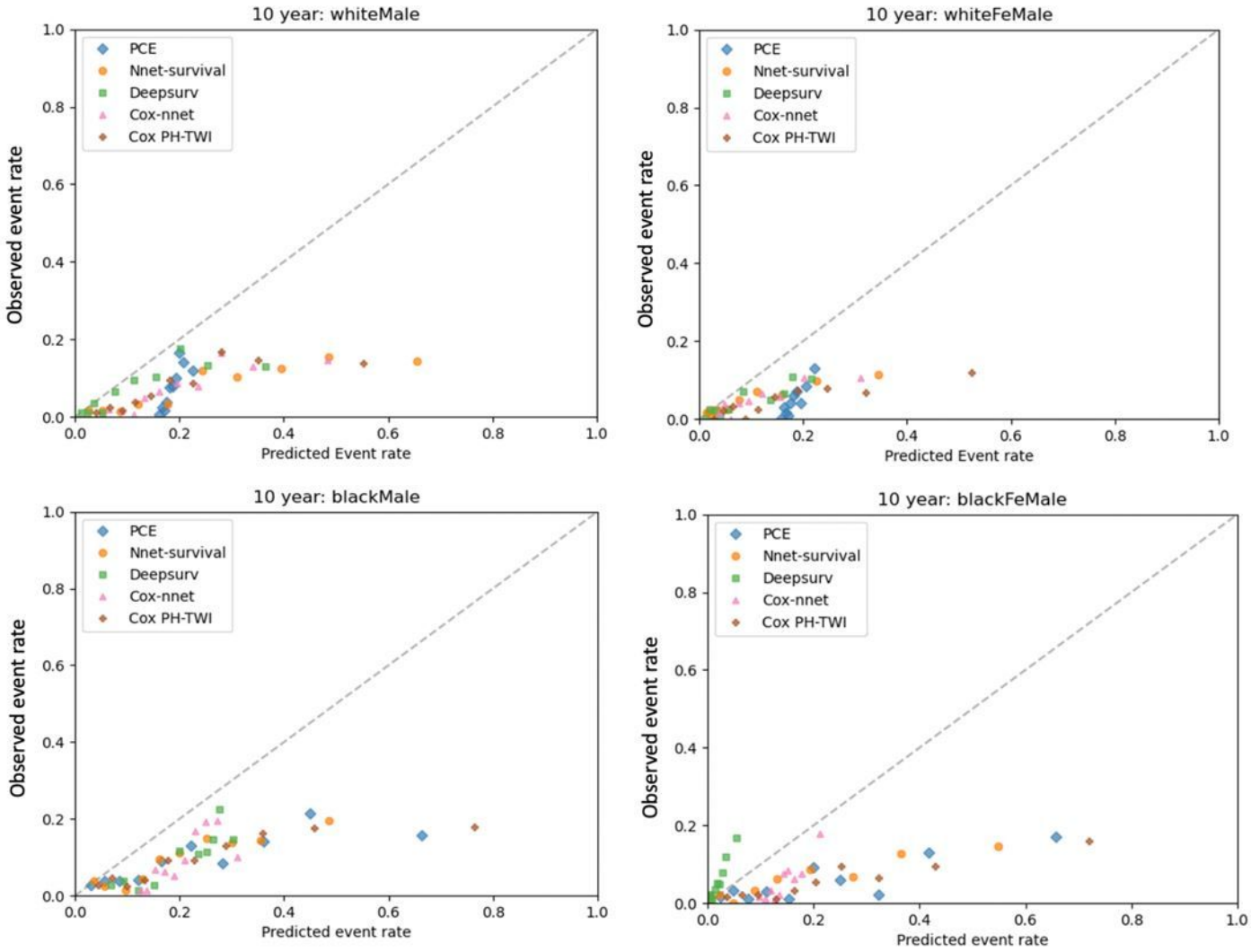

Figure 4

Kaplan-Meier Observed Event Rate and Predicted Event Rate for the ASCVD Outcome in the MESA Cohort. For each model, we divided participants into 10 group (decile) based on their sorted predicted event rate. Then, for each decile, mean observed event rate (Kaplan-Meier method) was plotted against mean predicted event rate. In a perfectly calibrated model, the predicted event rate would be the same as the observed event rate in each decile. This means that all points would be clustered around the dotted identity line.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaltable1006.docx