

1 **Deep Neural Survival Networks for Cardiovascular Risk Prediction:**  
2 **The Multi-Ethnic Study of Atherosclerosis (MESA)**

3  
4 Quincy A. Hathaway, PhD<sup>1</sup>, Naveena Yanamala, PhD<sup>1</sup>, Matthew J. Budoff, MD<sup>2</sup>, Partho P.  
5 Sengupta, MD, DM<sup>1</sup>, and Irfan Zeb, MD<sup>1</sup>

6  
7 <sup>1</sup>Heart and Vascular Institute, West Virginia University, Morgantown, WV, USA, <sup>2</sup>Lundquist  
8 Institute, Harbor-University of California, Los Angeles, Torrance, CA, USA

9  
10 *Running Title: Deep Neural Survival Networks: MESA*

11  
12 **Corresponding Author**

13 Partho P. Sengupta, MD, DM

14 PO Box 8003

15 WVU Heart & Vascular Institute,

16 1 Medical Center Drive, Morgantown, WV 26506.

17 Phone: 304-598-4478, Fax: 304-598-4779

18 Email: [Partho.Sengupta@hsc.wvu.edu](mailto:Partho.Sengupta@hsc.wvu.edu)

1 **Abstract**

2 **Background:** There is growing interest in utilizing machine learning techniques for routine  
3 atherosclerotic cardiovascular disease (ASCVD) risk prediction. We investigated whether novel  
4 deep learning survival models can augment ASCVD risk prediction over existing statistical and  
5 machine learning approaches.

6 **Methods:** 6,814 participants from the Multi-Ethnic Study of Atherosclerosis (MESA) were  
7 followed over 16 years to assess incidence of all-cause mortality (mortality) or a composite of  
8 major adverse events (MAE). Features were evaluated within the categories of traditional risk  
9 factors, inflammatory biomarkers, and imaging markers. Data was split into an internal  
10 training/testing (four centers) and external validation (two centers). Both machine learning  
11 (COXPH, RSF, and LSVM) and deep learning (nMTLR and DeepSurv) models were evaluated.

12 **Results:** In comparison to the COXPH model, DeepSurv significantly improved ASCVD risk  
13 prediction for MAE (AUC: 0.82 vs. 0.79,  $P \leq 0.001$ ) and mortality (AUC: 0.86 vs. 0.80,  $P \leq 0.001$ )  
14 with traditional risk factors alone. Implementing non-categorical NRI, we noted a 65% increase  
15 in correct reclassification compared to the COXPH model for both MAE and mortality ( $P \leq 0.05$ ).  
16 Assessing the relative risk of participants, DeepSurv was the only learning algorithm to develop  
17 a significantly improved risk score criteria, which outcompeted COXPH for both MAE (4.07 vs.  
18 2.66,  $P \leq 0.001$ ) and mortality (6.28 vs. 4.67,  $P = 0.014$ ). The addition of inflammatory or imaging  
19 biomarkers to traditional risk factors showed minimal/no significant improvement in model  
20 prediction.

21 **Conclusion:** DeepSurv can leverage simple office-based clinical features alone to accurately  
22 predict ASCVD risk and cardiovascular outcomes, without the need for additional features, such  
23 as inflammatory and imaging biomarkers.

1

2 **Keywords:** machine learning; deep neural survival networks; event prediction; MESA; risk

3 factors; biomarkers

## 1 **1.1 Introduction**

2 Coronary artery disease is the leading cause of morbidity and mortality in the United  
3 States [1]. There are multiple algorithms available such as the Framingham risk score (FRS) [2],  
4 Reynold risk score [3, 4], and Pooled Cohort Equation (PCE) [5] for risk stratification and to  
5 help guide preventive strategies. These algorithms use traditional risk factors to derive an  
6 individual's cardiac risk. Traditional risk factors have commonly either underestimated or  
7 overestimated an individual's risk of coronary disease, thus promoting a search for subclinical  
8 markers. Detection of subclinical calcified atherosclerosis, such as coronary artery calcium  
9 (CAC), has been shown to improve the performance of these risk prediction models [6-8]. While  
10 subclinical markers have provided additive benefit to risk prediction, there remains a limited  
11 understanding regarding the incremental value of these biomarkers to allow for cost-effective  
12 clinical predictions [9].

13 Traditionally, survival models have been used to investigate the relationships between  
14 patients' features and adverse outcomes, such as cardiovascular events, mortality, etc. [10-12].  
15 Standard survival models like the Cox proportional hazards model (COXPH) employ several  
16 linear assumptions that allow it to predict outcomes for an individual patient. While COXPH  
17 utilizes a standard statistical approach, it can be evaluated in the context of training and testing  
18 sets to provide predictive values and function as a machine learning algorithm. Recently, a  
19 modern COXPH model implementing deep neural networks, referred to as DeepSurv, has been  
20 suggested as an alternative method for solving survival analysis problems [13]. DeepSurv has the  
21 capacity to create more complex and abstract feature combinations by utilizing hidden layers.  
22 These hidden layers provide independent nodes of neural network development, allowing the  
23 algorithm to exhaust the potential of features and feature combinations for event prediction.



1 Additionally, DeepSurv is equipped with gradient descent optimization, providing an efficient  
2 means of minimizing the error of the algorithm through continual optimization of its parameters  
3 [13].

4 In this study we illustrate the value of DeepSurv over traditional models, such as  
5 COXPH, for predicting adverse cardiovascular events in the multiethnic study of atherosclerosis  
6 (MESA). Further, we assess the additive value of feature combinations by integrating traditional  
7 risk factors, inflammatory biomarkers, and imaging markers for deriving optimal risk  
8 stratification algorithms. Ultimately, we suggest a framework for the development, and use, of  
9 deep neural network risk scores in clinical medicine.

## 1 **1.2 Methods**

### 2 *1.2.1 Implementation of The Multi-Ethnic Study of Atherosclerosis (MESA)*

3           The MESA is a prospective study examining 6,814 men and women in the United States,  
4 with participants across an age range of 45 – 84 years old. The study was designed to include a  
5 multiethnic population of White/Caucasian (38%), Black/African American (28%), Hispanic  
6 (23%), and Chinese American (11%) participants. In the current study, information from  
7 participants obtained during the first exam (between the years 2000-2002) were used as features  
8 for event prediction using machine learning and deep learning approaches (**Figure 1A**). The  
9 primary outcomes included myocardial infarction, resuscitated cardiac arrest, congestive heart  
10 failure, coronary revascularization, or all-cause mortality as major adverse events (MAE). Time  
11 of the event was classified as death (mortality) or the time at which the first cardiovascular  
12 outcome variable was detected (MAE). Only participants without clinical cardiovascular disease  
13 at baseline were included in the study.

14

### 15 *1.2.2 Feature Selection*

16           To assess the strength of event prediction in the presence of features obtained from  
17 different clinical modalities, we selected features from three major categories, including  
18 traditional risk factors (Baseline Model), inflammatory biomarkers (Biomarkers), and left  
19 ventricular morphology, pericardial fat, and coronary artery calcium (CAC) (Imaging). Features  
20 included in each category were selected based on the following: 1) the feature was  
21 measured/assessed in greater than 70% of the participant cohort, 2) the feature has been shown to  
22 be correlated with clinical or subclinical findings, and 3) features exhibiting multicollinearity  
23 (correlation coefficient  $> 0.80$ ) were removed except for one in each set of correlations

1 **(Supplemental Figure 1)**. The R package corrplot 0.84 was used for multicollinearity  
2 correlations [14].

3 We utilized a linear proportional hazards model, without machine learning, to examine  
4 concordance index on the entire dataset. Imaging features provided higher concordance index  
5 scores (0.77 vs. 0.75 Baseline Model) for MAE, whereas Biomarkers provided higher  
6 concordance index scores (0.79 vs. 0.77 Baseline Model) for predicting mortality (**Supplemental**  
7 **Figure 2A**). The feature combinations suggest that factors shaping the ultrastructure of the left  
8 ventricle and coronary atherosclerosis, are the most influential toward correctly identifying  
9 MAE, while inflammatory biomarkers markers (homocysteine, IL6, plasmin antiplasmin, etc.)  
10 are involved more prominently with understanding mortality, which supports previous findings  
11 [7, 10, 15-17]. The R package pec 2019.11.03 provided concordance indexes, and curves,  
12 through COXPH [18].

13

### 14 *1.2.3 Data Transformation and Data Distribution*

15 All non-binary features were given a normal, standard distribution using the  
16 preprocessing function in scikit-learn 0.23.1 [19]. Missing values were imputed with the median  
17 value of each column through the mlr 2.17.1 package in R 3.6.2 [20]. Data was split into an  
18 internal training/testing (~66%, 4,584 participants) and external validation (~33%, 2,230  
19 participants) set. The internal training/testing and external validation split was determined by the  
20 field center in which the participants were enrolled. In order to derive an external validation set  
21 of ~33%, field centers six (St. Paul, Minnesota) and seven (Chicago, Illinois) were allocated to  
22 the external validation group while field centers three (Forsyth County, North Carolina), four  
23 (New York, New York), five (Baltimore City and Baltimore County, Maryland), and eight (Los

1 Angeles County, California) were selected as the internal training/testing set. Demographic  
 2 information for the internal training/testing and external validation sets are provided (**Table 1**).

3

4 **Table 1: Internal Training/Testing and External Validation Set Demographics**

Features	Internal Training/Testing (Centers 3,4,5,8) N = 4,584 [95% CI]	External Validation (Centers 6,7) N = 2,230 [95% CI]	P-Value
<b>Traditional Risk Factors</b>			
Age (Years)	62.52 [62.23-62.82]	<b>61.39 [60.96-61.82]*</b>	<b>≤0.001</b>
Gender (Male, %)	46.71 [45.26-48.15]	48.07 [46.00-50.15]	0.289
BMI (kg/m <sup>2</sup> )	28.49 [28.33-28.65]	<b>28.02 [27.80-28.25]*</b>	<b>≤0.001</b>
Education Status (0-8 scale)	4.52 [4.46-4.59]	<b>5.20 [5.10-5.30]*</b>	<b>≤0.001</b>
Income Category (1-13 scale)	8.15 [8.05-8.26]	<b>9.15 [9.00-9.30]*</b>	<b>≤0.001</b>
Race/Ethnicity (%)	31.85 [30.50-33.20]	<b>52.11 [50.03-54.18]*</b>	<b>≤0.001</b>
1. White/Caucasian	10.89 [9.98-11.79]	<b>13.68 [12.25-15.10]*</b>	<b>≤0.001</b>
2. Chinese American	34.69 [33.31-36.06]	<b>13.54 [12.12-14.96]*</b>	<b>≤0.001</b>
3. Black, African American	22.58 [21.37-23.79]	20.67 [18.99-22.35]	0.075
4. Hispanic			
Diabetic (%)	14.77 [13.74-15.80]	<b>10.54 [9.26-11.81]*</b>	<b>≤0.001</b>
Kidney Dysfunction (%)	10.12 [9.25-11.00]	<b>8.16 [7.02-9.30]*</b>	<b>0.010</b>
Hypertensive (%)	47.08 [45.63-48.52]	<b>37.00 [34.99-39.00]*</b>	<b>≤0.001</b>
Hyperlipidemic (%)	38.53 [37.12-39.93]	<b>36.01 [34.02-38.00]*</b>	<b>0.044</b>
Metabolic Syndrome (%)	37.46 [36.05-38.86]	<b>33.14 [31.18-35.10]*</b>	<b>≤0.001</b>
Smoking Status (%)	51.64 [50.19-53.08]	<b>48.12 [46.04-50.19]*</b>	<b>0.006</b>
1. Never	35.49 [34.11-36.88]	<b>38.57 [36.54-40.59]*</b>	<b>0.014</b>
2. Former	12.87 [11.90-13.84]	13.32 [11.91-14.73]	0.607
3. Current			
Family History of Heart Disease (%)	42.61 [41.13-44.09]	43.04 [40.93-45.15]	0.744
<b>Inflammatory Biomarkers</b>			
Homocysteine (μmol/L)	9.33 [9.23-9.44]	9.33 [9.16-9.50]	0.586
IL6 (pg/mL)	1.58 [1.54-1.62]	1.52 [1.47-1.57]	0.054
Plasmin Antiplasmin (nM)	4.82 [4.76-4.88]	4.75 [4.65-4.85]	0.237
Fibrinogen Antigen (mg/dL)	348.1 [346.0-350.3]	<b>344.1 [341.1-347.1]*</b>	<b>0.021</b>
C Reactive Protein (mg/L)	3.88 [3.71-4.05]	<b>3.58 [3.33-3.82]*</b>	<b>≤0.001</b>
D Dimer (μg/mL)	0.38 [0.36-0.41]	0.36 [0.33-0.40]	0.447

Factor VIII (%)	100.8 [99.65-101.9]	<b>95.78 [94.31-97.24]*</b>	<b>≤0.001</b>
<b>Left Ventricular Structure and Function</b>			
LV Ejection Fraction < 50 (%)	1.37 [1.04-1.71]	0.99 [0.58-1.40]	0.176
Left Ventricular Area (mm <sup>2</sup> )	4034 [4013-4055]	<b>3985 [3955-4015]*</b>	<b>0.013</b>
Pericardial Fat (mm <sup>2</sup> )	78.18 [76.99-79.36]	<b>82.47 [80.64-84.31]*</b>	<b>≤0.001</b>
Total Agatston Calcium Score (Phantom Adjusted)	145.7 [133.4-158.0]	148.1 [131.1-165.1]	0.924
<b>Outcome Variables</b>			
MAE (%)	28.40 [27.10-29.71]	26.23 [24.41-28.06]	0.060
Mortality (%)	22.19 [20.98-23.39]	<b>18.70 [17.08-20.33]*</b>	<b>≤0.001</b>

1  
2 **Table 1:** Demographics and feature distribution of internal training/testing and external  
3 validation datasets. Data are statistically significant with a  $P \leq 0.05$  when denoted by a “\*”. Data  
4 are represented as the mean and the 95% confidence interval. MAE = major adverse events,  
5 including all-cause mortality and cardiovascular outcomes.

6  
7 *1.2.4 Machine and Deep Learning Models and Evaluation*

8 The framework for implementing and evaluating the machine and deep learning models  
9 is provided in the *Supplemental Materials and Methods*. Event prediction and feature importance  
10 were evaluated in three machine learning models (Cox Proportional Hazards Regression Model  
11 (COXPH), Random Survival Forest (RSF), Linear Support Vector Machines (LSVM)) and two  
12 deep learning models (Neural Multi-Task Logistic Regression (nMTLR) and Non-Linear Cox  
13 Proportional Hazards Deep Neural Network (DeepSurv)). Computations were performed in  
14 Python 3.7, all code is provided: <https://github.com/qahathaway/MESA>.

15  
16 *1.2.5 Statistics*

1           GraphPad Prism 8.4.3 and RStudio 3.6.2 were used for statistical analyses. For assessing  
2 two continuous variables, a two-tailed, unpaired Student's t-test was employed when normally  
3 distributed. Normality was tested using the Shapiro-Wilk test. If non-Gaussian distribution, two  
4 continuous variables were assessed using the Mann Whitney test. For analyses of more than one  
5 group of continuous variables, a one-way analysis of variance (ANOVA) was used. Confidence  
6 intervals were calculated for ROC curves through the combined Wilson/Brown method. Relative  
7 risk (RR) was calculated through the Koopman asymptotic score. Significance of RR, positive  
8 predictive value (PPV), negative predictive value (NPV), and AUCs were determined through  
9 log transformation of mean and confidence intervals using a z-score distribution [21]. All data  
10 was considered statistically significant if the *P*-value was less than or equal to 0.05 ( $P \leq 0.05$ ).  
11 Data are reported as the mean with the 95% confidence interval (CI).

## 1 **1.3 Results**

### 2 *1.3.1 Characteristics of the Internal Training/Testing and External Validation Sets*

3       The data was split into an internal training/testing (~66%) and external validation (~33%)  
4 set, which was determined by the field center/county in which the participants were assessed  
5 (**Table 1**). Both the external validation and internal training/testing sets were of similar age,  
6 though statistically different (61.39 vs. 62.52 years old,  $P \leq 0.001$ ), with no differences in gender  
7 (48.07% vs. 46.71% male,  $P = 0.289$ ). The external validation set compared to the internal  
8 training/testing set contained a higher percentage of White/Caucasians (52.11% vs. 31.85%,  
9  $P \leq 0.001$ ) and Chinese Americans (13.68% vs. 10.89%,  $P \leq 0.001$ ), with a lower percentage of  
10 Black/African Americans (13.54% vs. 34.69%,  $P \leq 0.001$ ). The external validation set exhibited a  
11 lower risk of experiencing mortality (18.70% vs. 22.19%,  $P \leq 0.001$ ) but similar risk of MAE  
12 (26.23% vs. 28.40%,  $P = 0.060$ ) compared to the internal training/testing set, respectively. Of  
13 note, coronary artery calcium (CAC) was not significantly different in the external validation  
14 group (148.1 Phantom Adjusted) compared to the internal training/testing (145.7 Phantom  
15 Adjusted) ( $P = 0.924$ ).

16

### 17 *1.3.2 Evaluation Metrics of Model Performance and Prediction Accuracy*

18       The cumulative/dynamic AUC provides a means for plotting the ROC curve at any time  
19 throughout the survival analysis to assess predictions on which participants have, and do not  
20 have, the outcome at that time. With traditional risk factors, only linear Support Vector Machines  
21 (ISVM) and the Non-Linear Cox Proportional Hazards Deep Neural Network (DeepSurv) had  
22 significantly higher AUCs for both MAE and mortality in the Baseline Model and for All  
23 Features, compared to COXPH (**Table 2**). DeepSurv had the highest performance criteria across

1 all feature combinations and endpoints. Compared to COXPH, DeepSurv provided significant  
 2 improvement with traditional risk factors for MAE (0.82 vs. 0.79,  $P \leq 0.001$ ) and mortality (0.86  
 3 vs. 0.80,  $P \leq 0.001$ ) as well as improved prediction with all features for MAE (0.85 vs. 0.81,  
 4  $P \leq 0.001$ ) and mortality (0.87 vs. 0.84,  $P \leq 0.001$ ). DeepSurv further outperforms COXPH in  
 5 defining outcomes in racial/ethnic groups, indicated by prediction of MAE for White/Caucasians  
 6 (0.82 vs. 0.80,  $P = 0.005$ ), Chinese Americans (0.82 vs. 0.77,  $P = 0.001$ ), Black/African Americans  
 7 (0.77 vs. 0.74,  $P \leq 0.001$ ), and Hispanics (0.81 vs. 0.78,  $P \leq 0.001$ ) and mortality for  
 8 White/Caucasians (0.85 vs. 0.83,  $P = 0.034$ ), Chinese Americans (0.84 vs. 0.70,  $P \leq 0.001$ ),  
 9 Black/African Americans (0.81 vs. 0.75,  $P \leq 0.001$ ), and Hispanics (0.88 vs. 0.81,  $P \leq 0.001$ ), using  
 10 traditional risk factors alone (**Supplemental Table 1**).

11

12 **Table 2: Cumulative/Dynamic Time-Dependent AUC**

MAE				
Features	Baseline Model	Baseline + Biomarkers	Baseline + Imaging	All Features
COXPH [95% CI]	0.79 [0.78-0.80]	0.78 [0.77-0.79]	0.79 [0.79-0.80]	0.81 [0.80-0.82]
nMTLR [95% CI]	0.78 [0.77-0.80] ( $P = 1.000$ )	0.79 [0.78-0.79]	0.81 [0.80-0.82]	0.80 [0.80-0.81] ( $P = 0.4077$ )
RSF [95% CI]	0.80 [0.79-0.82] ( $P = 0.314$ )	0.80 [0.79-0.82]	0.83 [0.82-0.84]	<b>0.83 [0.82-0.84]*</b> ( $P = 0.040$ )
ISVM [95% CI]	<b>0.81 [0.80-0.82]*</b> ( $P = 0.014$ )	0.81 [0.80-0.82]	0.83 [0.82-0.84]	<b>0.83 [0.82-0.84]*</b> ( $P = 0.042$ )
DeepSurv [95% CI]	<b>0.82 [0.81-0.83]*</b> ( $P \leq 0.001$ )	0.82 [0.81-0.83]	0.84 [0.83-0.85]	<b>0.85 [0.84-0.86]*</b> ( $P \leq 0.001$ )
Mortality				
COXPH [95% CI]	0.80 [0.79-0.81]	0.81 [0.80-0.83]	0.83 [0.82-0.83]	0.84 [0.83-0.85]
nMTLR [95% CI]	0.83 [0.82-0.84] ( $P = 0.003$ )	0.83 [0.82-0.84]	0.84 [0.83-0.85]	0.84 [0.83-0.85] ( $P = 0.858$ )
RSF [95% CI]	<b>0.84 [0.83-0.85]*</b>	0.84 [0.83-0.85]	0.85 [0.84-0.86]	0.85 [0.84-0.86]



	( $P \leq 0.001$ )			( $P = 0.250$ )
<b>ISVM [95% CI]</b>	<b>0.85 [0.84-0.86]*</b> ( $P \leq 0.001$ )	0.86 [0.84-0.87]	0.86 [0.85-0.87]	<b>0.86 [0.85-0.87]*</b> ( $P = 0.010$ )
<b>DeepSurv [95% CI]</b>	<b>0.86 [0.85-0.87]*</b> ( $P \leq 0.001$ )	0.86 [0.85-0.87]	0.86 [0.85-0.87]	<b>0.87 [0.86-0.88]*</b> ( $P \leq 0.001$ )

1  
2 **Table 2:** Cumulative/dynamic time-dependent AUCs for the COXPH, RSF, ISVM, nMTLR, and  
3 DeepSurv models. Data are statistically significant with a  $P \leq 0.05$  when denoted by a “\*”. Data  
4 are represented as the mean and the 95% confidence interval. MAE = major adverse events,  
5 including all-cause mortality and cardiovascular outcomes, Baseline Model = traditional risk  
6 factors, Biomarkers = inflammatory biomarkers, Imaging = left ventricular structure and  
7 function, All Features = traditional risk factors, inflammatory biomarkers, and left ventricular  
8 structure and function, COXPH = Cox Proportional Hazards Regression Model, RSF = Random  
9 Survival Forest, ISVM = Linear Support Vector Machines, nMTLR = Neural Multi-Task  
10 Logistic Regression, DeepSurv = Non-Linear Cox Proportional Hazards Deep Neural Network.

11  
12 The use of concordant pairs in survival analyses provides a metric to determine the point  
13 along the survival curve at which a subject is predicted to experience an outcome. Examining the  
14 concordance index for traditional risk factors, only DeepSurv performed significantly better in  
15 predicting concordant pairs (**Supplemental Table 2**) and concordant pairs adjusted for the  
16 inverse probability of censoring weights (IPCW) (**Supplemental Table 3**) for both MAE and  
17 mortality. DeepSurv provided the highest concordance score for MAE (0.78) and mortality  
18 (0.83) for traditional risk factors compared to COXPH (0.74 and 0.78, respectively). The  
19 integrated brier score (**Supplemental Table 4**) indicates the differences between predicted and

1 observed probabilities, which was generally similar between machine and deep learning models  
2 for MAE and mortality.

3

### 4 *1.3.3 Feature Importance*

5 To determine which features contributed to the composition of each of the machine and  
6 deep learning models, permutation importance (COXPH, lSVM, nMTLR, DeepSurv) and mean  
7 decrease Gini (RSF) were employed (**Figure 2**). When examining all of the features, participant  
8 age was one of the top five predictors for predicting MAE and mortality. CAC was one of the top  
9 five predictors in all models predicting MAE, while biomarkers of inflammation were more  
10 important for predicting mortality (**Figure 2**). To further understand how individual features  
11 contributed to the construction of the machine and deep learning models, the best and worst  
12 predictive features were assessed.

13

14 The top five performing features across all models with non-binary values were selected  
15 (**Figure 2**) with each model being trained on the internal training/testing set and fit with the  
16 external validation set. The cumulative/dynamic AUC for MAE (**Supplemental Figure 4**) and  
17 mortality (**Supplemental Figure 5**) for the top five performing features revealed minimal  
18 changes across all models. This suggests that the utility of these features are being completely  
19 exhausted by each of the machine and deep learning algorithms, with little, to no, room for  
20 improvement. To understand if features which comprise the worst predictors for each model  
21 performed similarly to the top predictors, the cumulative/dynamic AUCs were repeated for MAE  
22 (**Supplemental Figure 6**) and mortality (**Supplemental Figure 7**). In this case, there were  
23 marked differences between the ROC curves for mortality and MAE, with the AUCs of COXPH

1 (0.46 and 0.44), RSF (0.54 and 0.55), ISVM (0.55 and 0.55), nMTLR (0.55 and 0.55), and  
2 DeepSurv (0.56 and 0.56) revealing diverse handling of the predictors, respectively.

3

#### 4 *1.3.4 Predicted Risks for Clinical Decision Making*

5 Machine and deep learning models were used in event prediction and risk stratification of  
6 patients by generating a risk score from the predicted risks of each model. Non-categorical net  
7 reclassification improvement (NRI) applied to the risk scores of traditional risk factors highlights  
8 the improved reclassification of nMTLR, RSF, ISVM, and DeepSurv compared to COXPH for  
9 MAE and mortality (**Table 3**). DeepSurv had the highest percent reclassification and correctly  
10 classified cases for MAE (65% and 67%,  $P \leq 0.05$ ) and mortality (65% and 72%,  $P \leq 0.05$ ). While  
11 ISVM had the highest percent correctly classified controls (69% and 69%,  $P \leq 0.05$ ) for MAE and  
12 mortality, respectively. Similar reclassification was seen when applied to all features  
13 (**Supplemental Table 5**). To determine why DeepSurv can better identify patients who  
14 succumbed to MAE or mortality, we performed a categorical NRI using cutoffs derived from the  
15 Pooled Cohort Equation (PCE) estimated 10-year risk for atherosclerotic cardiovascular disease  
16 (ASCVD); this allowed for defining estimated risk score cutoffs of 5% (0.27, low risk), 7.5%  
17 (0.36, intermediate risk), and 20% (0.78, high risk). DeepSurv provided the most significant  
18 reclassification between the low risk and intermediate risk categories for MAE and mortality  
19 (**Supplemental Table 6**), suggesting that DeepSurv can better identify individuals with  
20 outcomes in this “borderline” category than can COXPH.

21

#### 22 **Table 3: Reclassification – Baseline Model**

Non-Categorical NRI (%) [95% CI]	Cases – Increased Risk Category (%) [95% CI]	Controls – Decreased Risk Category (%) [95% CI]
-------------------------------------	---	--

MAE			
Risk Scores	Baseline Model [95% CI]		
COXPH	0.00	50.00	50.00
nMTLR	<b>30.78 [19.77-40.83]*</b> ( $P \leq 0.05$ )	58.93 [54.70-62.94]	56.46 [53.60-58.68]
RSF	<b>44.34 [31.63-57.22]*</b> ( $P \leq 0.05$ )	57.06 [51.20-59.43]	65.11 [62.80-68.43]
ISVM	<b>62.64 [51.71-77.33]*</b> ( $P \leq 0.05$ )	55.17 [52.11-59.72]	69.15 [65.10-72.93]
DeepSurv	<b>64.97 [58.11-71.43]*</b> ( $P \leq 0.05$ )	67.18 [62.97-70.38]	65.31 [63.02-68.44]
Mortality			
Risk Scores	Baseline Model [95% CI]		
COXPH	0.00	50.00	50.00
nMTLR	<b>20.73 [13.35-31.15]*</b> ( $P \leq 0.05$ )	64.67 [61.43-69.00]	45.69 [43.18-48.45]
RSF	<b>35.09 [24.63-44.53]*</b> ( $P \leq 0.05$ )	62.22 [56.54-66.05]	55.33 [53.16-57.35]
ISVM	<b>63.05 [30.74-83.97]*</b> ( $P \leq 0.05$ )	62.47 [46.93-71.34]	69.05 [66.34-73.11]
DeepSurv	<b>65.31 [52.94-72.80]*</b> ( $P \leq 0.05$ )	72.21 [67.06-76.50]	60.44 [58.20-62.81]

1  
2 **Table 3:** Non-categorical net reclassification improvement (NRI) and the percentage of  
3 participants moving up and down risk categories for the COXPH, RSF, ISVM, nMTLR, and  
4 DeepSurv models. Analyses performed on the Baseline Model. Data are statistically significant  
5 with a  $P \leq 0.05$  when denoted by a “\*”. Data are represented as the mean and the 95%  
6 confidence interval. MAE = major adverse events, including all-cause mortality and  
7 cardiovascular outcomes, Cases = participant with an event, Controls = participants without an  
8 event, Baseline Model = traditional risk factors, COXPH = Cox Proportional Hazards Regression  
9 Model, RSF = Random Survival Forest, ISVM = Linear Support Vector Machines, nMTLR =

1 Neural Multi-Task Logistic Regression, DeepSurv = Non-Linear Cox Proportional Hazards Deep  
 2 Neural Network.

3  
 4 The PCE ASCVD risk score classifies individuals with a score greater than 20% as being  
 5 at high-risk for an event in the subsequent 10 years [22]. Using this criterion, 20% was set as a  
 6 cutoff for determining relative risk (RR) of patients for mortality and MAE events in the  
 7 machine and deep learning model risk scores. Using traditional risk factors alone, DeepSurv  
 8 provided the highest increasing in RR and positive predictive value (PPV) for MAE (4.07 vs.  
 9 2.66,  $P \leq 0.001$ , and 66% vs. 51%,  $P \leq 0.001$ ) and mortality (6.28 vs. 4.67,  $P = 0.014$ , and 57% vs.  
 10 50%,  $P = 0.027$ ), compared to COXPH (Table 4). Similar RR, PPV, and negative predictive  
 11 value (NPV) were seen when applied to all features (Supplemental Table 7).

12  
 13 **Table 4: Predicted Risk – Baseline Model**

	Relative Risk [95% CI]	Positive Predictive Value (%) [95% CI]	Negative Predictive Value (%) [95% CI]
<b>MAE</b>			
<b>Risk Scores</b>	<b>Baseline Model [95% CI]</b>		
<b>COXPH</b>	2.66 [2.33-3.02]	51.48 [46.98-55.95]	80.63 [78.71-82.41]
<b>nMTLR</b>	<b>3.28 [2.89-3.72]* (<math>P=0.025</math>)</b>	<b>58.32 [53.77-62.72]* (<math>P=0.036</math>)</b>	82.22 [80.37-83.83] ( $P=0.220$ )
<b>RSF</b>	<b>3.57 [3.14-4.02]* (<math>P \leq 0.001</math>)</b>	<b>60.86 [56.35-65.19]* (<math>P=0.004</math>)</b>	82.94 [81.11-84.62] ( $P=0.076$ )
<b>ISVM</b>	<b>3.91 [3.42-4.47]* (<math>P \leq 0.001</math>)</b>	<b>58.67 [54.59-62.64]* (<math>P=0.021</math>)</b>	<b>84.98 [83.18-86.62]* (<math>P \leq 0.001</math>)</b>
<b>DeepSurv</b>	<b>4.07 [3.59-4.61]* (<math>P \leq 0.001</math>)</b>	<b>65.71 [61.22-69.94]* (<math>P \leq 0.001</math>)</b>	<b>83.85 [82.07-85.49]* (<math>P=0.013</math>)</b>
<b>Mortality</b>			
<b>Risk Scores</b>	<b>Baseline Model [95% CI]</b>		

<b>COXPH</b>	4.67 [3.96-5.49]	49.67 [45.09-54.26]	89.36 [87.84-90.71]
<b>nMTLR</b>	5.02 [4.26-5.92] ( <i>P</i> =0.541)	50.87 [46.32-55.40] ( <i>P</i> =0.716)	89.87 [88.38-91.19] ( <i>P</i> =0.619)
<b>RSF</b>	5.90 [4.98-6.97] ( <i>P</i> =0.051)	54.59 [50.01-59.09] ( <i>P</i> =0.137)	90.74 [89.30-92.00] ( <i>P</i> =0.170)
<b>ISVM</b>	5.87 [4.85-7.10] ( <i>P</i> =0.074)	48.38 [44.35-52.42] ( <i>P</i> =1.000)	<b>92.03 [90.62-93.24]* (<i>P</i>=0.007)</b>
<b>DeepSurv</b>	<b>6.28 [5.30-7.44]* (<i>P</i>=0.014)</b>	<b>57.02 [51.94-61.00]* (<i>P</i>=0.027)</b>	91.08 [89.67-92.32] ( <i>P</i> =0.085)

1

2 **Table 4:** Relative risk (RR), positive predictive value (PPV), and negative predictive value  
3 (NPV) for the COXPH, RSF, ISVM, nMTLR, and DeepSurv models. Analyses performed on the  
4 Baseline Model. Data are statistically significant with a  $P \leq 0.05$  when denoted by a “\*”. Data  
5 are represented as the mean and the 95% confidence interval. MAE = major adverse events,  
6 including all-cause mortality and cardiovascular outcomes, Baseline Model = traditional risk  
7 factors, COXPH = Cox Proportional Hazards Regression Model, RSF = Random Survival  
8 Forest, ISVM = Linear Support Vector Machines, nMTLR = Neural Multi-Task Logistic  
9 Regression, DeepSurv = Non-Linear Cox Proportional Hazards Deep Neural Network.

## 1 **1.4 Discussion**

2           The progression from traditional statistics to machine learning to deep neural networks  
3 continue a trend of identifying ever more complex networks of features and maximizing the  
4 value that each feature has in contributing to building of a model. The current study proposes a  
5 role for deep neural networks to augment current survival analysis pipelines. DeepSurv, which is  
6 a COXPH model that builds on the nonlinear aspect of deep neural networks [13], provides a  
7 more robust time-to-event prediction of both mortality and a composite of cardiovascular  
8 outcomes. Moreover, this study is one of the first to provide an in-depth examination of the  
9 additive value of cardiovascular biomarkers in machine and deep learning survival networks.  
10 While we see incremental improvements with the addition of features, the DeepSurv algorithm  
11 suggests that traditional risk factors alone are sufficient to provide significant improvement over  
12 routine statistical approaches, such as COXPH, and that addition of other features outlined in this  
13 study may not be necessary.

14           Because this is a prospective, survival network spanning 16-years, age discriminates well  
15 across both MAE and mortality, likely limiting the impact of other features, such as CAC.  
16 Nevertheless, we continue to confirm small, yet significant improvement in model performance  
17 with the addition of coronary artery calcium (CAC) confirming previous observations regarding  
18 improvements of AUC [23] and NRI [8, 12, 24] with CAC. Examining the MESA cohort,  
19 Folsom et al. revealed an increase in the AUC when transitioning from traditional risk factors  
20 alone to risk factors including CAC for cardiovascular disease (0.772 vs. 0.811) and coronary  
21 heart disease (0.771 vs. 0.824) [23]. In the current study, prediction of MAE for traditional risk  
22 factors alone (0.82) and those including CAC (0.85) show a similar progression in improvement.  
23 Though, overall, the deep learning algorithms presented in this study are capable of achieving

1 much higher levels of prediction compared to standard statistical approaches. While both the  
2 current study and previous research have shown improvements with the addition of CAC, the  
3 data presented here provides one of the first examples of how features can be used progressively  
4 in machine learning driven survival networks, with the findings indicating modest, at best,  
5 improvements with the addition of features beyond traditional risk factors. This concept holds  
6 true for not only DeepSurv, but also the other machine and deep learning models tested.

7         While each of the models may perform comparably with the features that most heavily  
8 influence the construction of the model, such as participant age and CAC, the machine and deep  
9 learning algorithms differ in the utility they can derive from other less “important” features.  
10 DeepSurv outperformed the other models across all features, but DeepSurv’s ability to rely on  
11 the traditional risk factors within the feature sets may be its most important attribute. DeepSurv,  
12 as hyperparameter optimized in this study, can provide a robust risk score criteria on traditional  
13 risk factors alone, significantly improving the classification of true positives. Both NRI and RR  
14 highlight the ability of DeepSurv to augment ASCVD prediction and provide a generalizable  
15 criterion for risk stratification of patients. This suggests that while all the models show  
16 incremental improvement with the addition of features, DeepSurv has the capacity to establish a  
17 higher baseline AUC, with a better predictive algorithm for traditional risk factors.

18         In examining the MESA data, this is the first study to examine both support vector  
19 machines (ISVM) as well as deep learning algorithms (nMTLR and DeepSurv) in cardiovascular  
20 event prediction for survival analysis. The ISVM provided the second highest prediction scores  
21 for both classification of mortality and MAE, acting in a similar fashion to neural networks.  
22 SVM functions by employing hyperplanes and n-dimensional spaces, similar to the hidden layers  
23 in neural networks [25]. One of the major differences is that SVM work in a linear fashion,



1 whereas neural networks are non-linear algorithms. The strengths of both the ISVM and  
2 DeepSurv models presented here likely come from the multi-dimensional structuring of these  
3 models, improving their predictive capacity. It is important to note that while DeepSurv provided  
4 the best prediction in this study, further hyperparameter optimization of DeepSurv or parameter  
5 tuning of the other machine learning models, including ISVM, could increase their predictive  
6 power.

7         The current study is limited by a lack of true optimization for all of the machine and deep  
8 learning models. DeepSurv was the only algorithm optimized for the current studies, whereas the  
9 other machine and deep learning algorithms were used with the “out-of-the-box” functions. This  
10 warrants a disclaimer that improvement in prediction of these algorithms can likely be made,  
11 even in the case of DeepSurv, which could undergo further hyperparameter optimization.  
12 Though machine and deep learning provide progressive leaps for clinical medicine, unlike  
13 traditional statistical methods, parameter optimization for each dataset is needed to unlock the  
14 algorithm’s complete potential.

## 1 **1.5 Conclusions**

2           In the MESA cohort, we provide an examination of machine and deep learning  
3 approaches applied to survival analysis. We highlight the ability of deep neural networks to  
4 modestly increase the predictive capacity of our endpoints, including cardiovascular outcomes  
5 and death. DeepSurv has the ability to gain additional information from traditional risk factors  
6 within the model and leverage its ability to form new, non-linear connections, providing unique  
7 associations within the dataset. Ultimately, these machine and deep learning models can be  
8 implemented as risk scores, creating generalizable models that can inform clinical prognoses.

## 1 **1.6 Acknowledgements**

2 MESA and the MESA SHARe project are conducted and supported by the National  
3 Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support  
4 for MESA is provided by contracts N01-HC95159, N01-HC-95160, N01-HC-95161, N01-HC-  
5 95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC95166, N01-HC-95167, N01-  
6 HC-95168, N01-HC-95169 and CTSA UL1-RR-024156. This manuscript was not prepared in  
7 collaboration with MESA investigators and does not necessarily reflect the opinions or views of  
8 MESA, or the NHLBI.

## 10 **1.7 Disclosures**

11 Partho P. Sengupta is a consultant to Heart Sciences, Ultromics, and Kencor Health. The  
12 other authors have nothing to disclose.

## 14 **1.8 Sources of Funding**

15 This work is supported in part by funds from the National Science Foundation (NSF:  
16 #1920920). Including funding from the American Heart Association (AHA-17PRE33660333, to  
17 Q.A.H.)

## 19 **1.9 Author Contributions**

20 Conceptualization (QAH, NY, MJB, PPS, IZ); Data curation (QAH, NY, PPS, IZ);  
21 Formal analysis (QAH, NY); Funding acquisition (QAH, PPS); Investigation (QAH, NY, MJB,  
22 PPS, IZ); Methodology (QAH, NY, MJB, PPS, IZ); Project administration (PPS, IZ); Resources  
23 (PPS); Software (QAH, NY, PPS); Supervision (PPS, IZ); Validation (QAH, NY, MJB, PPS,

- 1 IZ); Visualization (QAH, NY); Roles/Writing - original draft (QAH, NY, IZ); Writing - review
- 2 & editing (MJB, PPS).

## 1 1.10 References

- 2 [1] Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, et al. Heart  
3 Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association.  
4 *Circulation*. 2020;141:e139-e596.
- 5 [2] Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction  
6 of coronary heart disease using risk factor categories. *Circulation*. 1998;97:1837-47.
- 7 [3] Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved  
8 algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score.  
9 *JAMA*. 2007;297:611-9.
- 10 [4] Ridker PM, Paynter NP, Rifai N, Gaziano JM, Cook NR. C-reactive protein and parental  
11 history improve global cardiovascular risk prediction: the Reynolds Risk Score for men.  
12 *Circulation*. 2008;118:2243-51, 4p following 51.
- 13 [5] Goff DC, Jr., Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Sr., Gibbons R, et al.  
14 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American  
15 College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am*  
16 *Coll Cardiol*. 2014;63:2935-59.
- 17 [6] Detrano R, Guerci AD, Carr JJ, Bild DE, Burke G, Folsom AR, et al. Coronary calcium as a  
18 predictor of coronary events in four racial or ethnic groups. *N Engl J Med*. 2008;358:1336-45.
- 19 [7] Yeboah J, Carr JJ, Terry JG, Ding J, Zeb I, Liu S, et al. Computed tomography-derived  
20 cardiovascular risk markers, incident cardiovascular events, and all-cause mortality in  
21 nondiabetics: the Multi-Ethnic Study of Atherosclerosis. *Eur J Prev Cardiol*. 2014;21:1233-41.
- 22 [8] Yeboah J, McClelland RL, Polonsky TS, Burke GL, Sibley CT, O'Leary D, et al. Comparison  
23 of novel risk markers for improvement in cardiovascular risk assessment in intermediate-risk  
24 individuals. *JAMA*. 2012;308:788-95.
- 25 [9] Wang TJ. Assessing the role of circulating, genetic, and imaging biomarkers in  
26 cardiovascular risk prediction. *Circulation*. 2011;123:551-65.
- 27 [10] Budoff MJ, Shaw LJ, Liu ST, Weinstein SR, Mosler TP, Tseng PH, et al. Long-term  
28 prognosis associated with coronary calcification: observations from a registry of 25,253 patients.  
29 *J Am Coll Cardiol*. 2007;49:1860-70.
- 30 [11] Giolo SR, Krieger JE, Mansur AJ, Pereira AC. Survival analysis of patients with heart  
31 failure: implications of time-varying regression effects in modeling mortality. *PLoS One*.  
32 2012;7:e37392.
- 33 [12] Polonsky TS, McClelland RL, Jorgensen NW, Bild DE, Burke GL, Guerci AD, et al.  
34 Coronary artery calcium score and risk classification for coronary heart disease prediction.  
35 *JAMA*. 2010;303:1610-6.
- 36 [13] Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized  
37 treatment recommender system using a Cox proportional hazards deep neural network. *BMC*  
38 *Med Res Methodol*. 2018;18:24.
- 39 [14] Wei T, Simko V. R package "corrplot": Visualization of a Correlation Matrix. 2017.
- 40 [15] Raggi P, Gongora MC, Gopal A, Callister TQ, Budoff M, Shaw LJ. Coronary artery  
41 calcium to predict all-cause mortality in elderly men and women. *J Am Coll Cardiol*.  
42 2008;52:17-23.
- 43 [16] Yeboah J, Young R, McClelland RL, Delaney JC, Polonsky TS, Dawood FZ, et al. Utility  
44 of Nontraditional Risk Markers in Atherosclerotic Cardiovascular Disease Risk Assessment. *J*  
45 *Am Coll Cardiol*. 2016;67:139-47.

- 1 [17] Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, et al.  
2 Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of  
3 Atherosclerosis. *Circ Res*. 2017;121:1092-101.
- 4 [18] Mogensen UB, Ishwaran H, Gerds TA. Evaluating Random Forests for Survival Analysis  
5 Using Prediction Error Curves. *J Stat Softw*. 2012;50:1-23.
- 6 [19] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn:  
7 Machine learning in Python. *J Mach Learn Res*. 2011;12:2825-30.
- 8 [20] Bischl B, Lang M, Kotthoff L, Schiffner J, Richter J, Studerus E, et al. mlr: Machine  
9 Learning in R. *J Mach Learn Res*. 2016;17:1-5.
- 10 [21] Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ*.  
11 2003;326:219.
- 12 [22] Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, et al. 2019  
13 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: Executive  
14 Summary: A Report of the American College of Cardiology/American Heart Association Task  
15 Force on Clinical Practice Guidelines. *J Am Coll Cardiol*. 2019;74:1376-414.
- 16 [23] Folsom AR, Kronmal RA, Detrano RC, O'Leary DH, Bild DE, Bluemke DA, et al.  
17 Coronary artery calcification compared with carotid intima-media thickness in the prediction of  
18 cardiovascular disease incidence: the Multi-Ethnic Study of Atherosclerosis (MESA). *Arch*  
19 *Intern Med*. 2008;168:1333-9.
- 20 [24] Elias-Smale SE, Proenca RV, Koller MT, Kavousi M, van Rooij FJ, Hunink MG, et al.  
21 Coronary calcium score improves classification of coronary heart disease risk in the elderly: the  
22 Rotterdam study. *J Am Coll Cardiol*. 2010;56:1407-14.
- 23 [25] Gudadhe M, Wankhade K, Dongre S. Decision support system for heart disease based on  
24 support vector machine and Artificial Neural Network. *International Conference on Computer*  
25 *and Communication Technology (ICCCT)2010*. p. 741-5.

26  
27

1 **Figure Legends:**

2 **Figure 1:** Methodological and data considerations for the Multi-Ethnic Study of Atherosclerosis  
3 (MESA). **(A)** Schematic detailing the flow of information through the survival analysis pipeline.  
4 Briefly, features from 6,814 participants in the MESA cohort were selected from three major  
5 categories: traditional risk factors (Baseline Model), inflammatory biomarkers (Biomarkers), and  
6 left ventricular structure/function (Imaging). Features exhibiting multicollinearity were removed,  
7 data was standardized and imputed. Classification of endpoints (mortality and MAE) were  
8 trained (internal training/testing set) and tested (external validation) using the machine learning  
9 models COXPH, RSF, and ISVM and deep learning models nMTLR and DeepSurv. Outcomes  
10 were evaluated through the cumulative/dynamic AUCs (Time-Dependent AUC), net  
11 reclassification improvement (NRI), relative risk (RR), concordance index (CI), concordance  
12 index with inverse probability of censoring weights (IPCW), and integrated brier score (IBS).  
13 **(B)** Data distribution along the survival timeline. Censored participants for mortality (78.7%) and  
14 MAE (72.3%) are displayed in orange and those with outcome/death in blue. MAE = major  
15 adverse events, including all-cause mortality and cardiovascular outcomes, Time in Days = time  
16 in days to the predicted event, COXPH = Cox Proportional Hazards Regression Model, RSF =  
17 Random Survival Forest, ISVM = Linear Support Vector Machines, nMTLR = Neural Multi-  
18 Task Logistic Regression, DeepSurv = Non-Linear Cox Proportional Hazards Deep Neural  
19 Network.

20  
21 **Figure 2:** Feature importance for the COXPH, RSF, ISVM, nMTLR, and DeepSurv models.  
22 Permutation importance was applied to COXPH, ISVM, nMTLR, and DeepSurv, while mean  
23 decrease Gini (RSF) was implemented for RSF. The top five (orange) and worst five (blue)

1 predictive features are shown for each model. MAE = major adverse events, including all-cause  
2 mortality and cardiovascular outcomes, COXPH = Cox Proportional Hazards Regression Model,  
3 RSF = Random Survival Forest, LSVM = Linear Support Vector Machines, nMTLR = Neural  
4 Multi-Task Logistic Regression, DeepSurv = Non-Linear Cox Proportional Hazards Deep Neural  
5 Network.



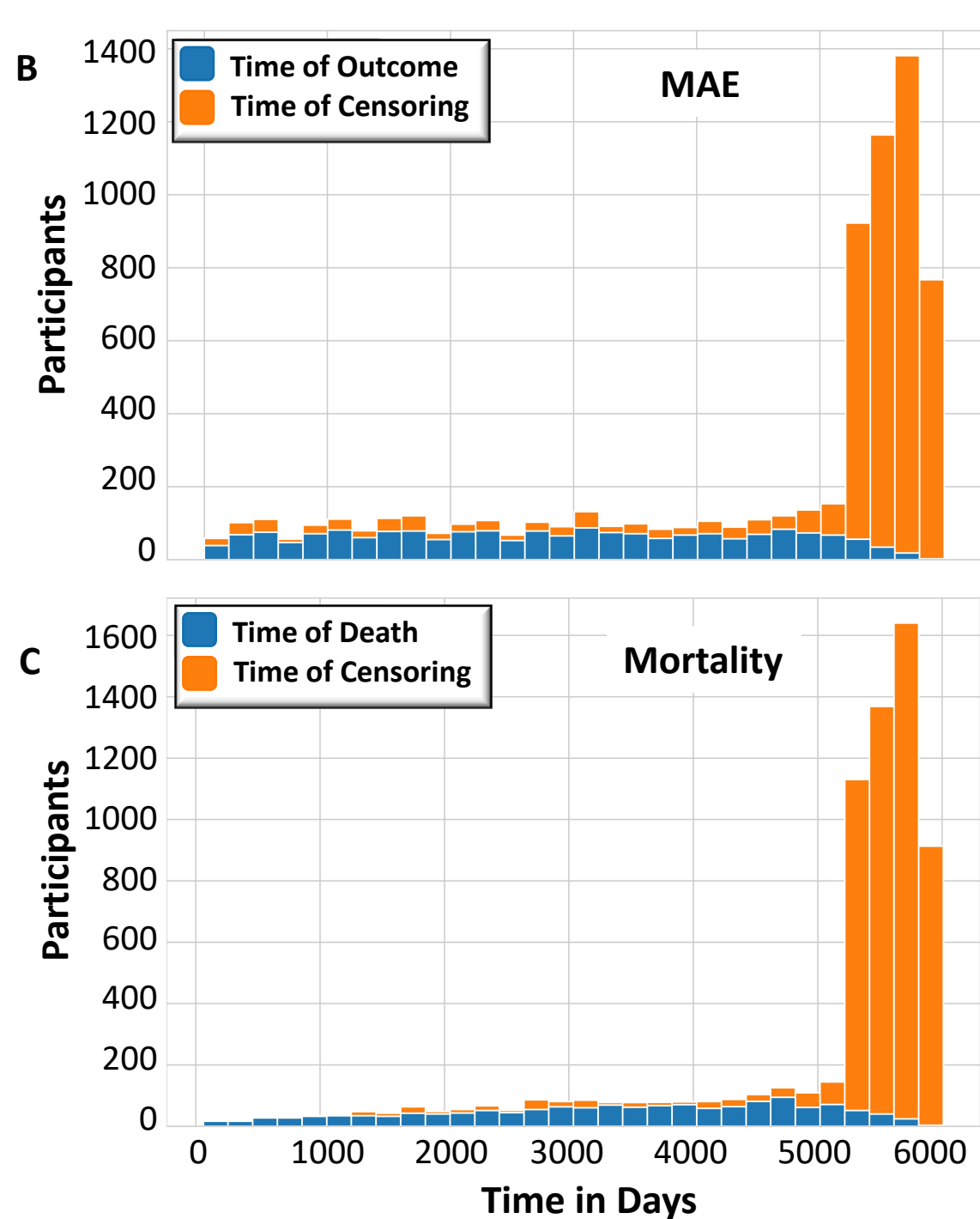
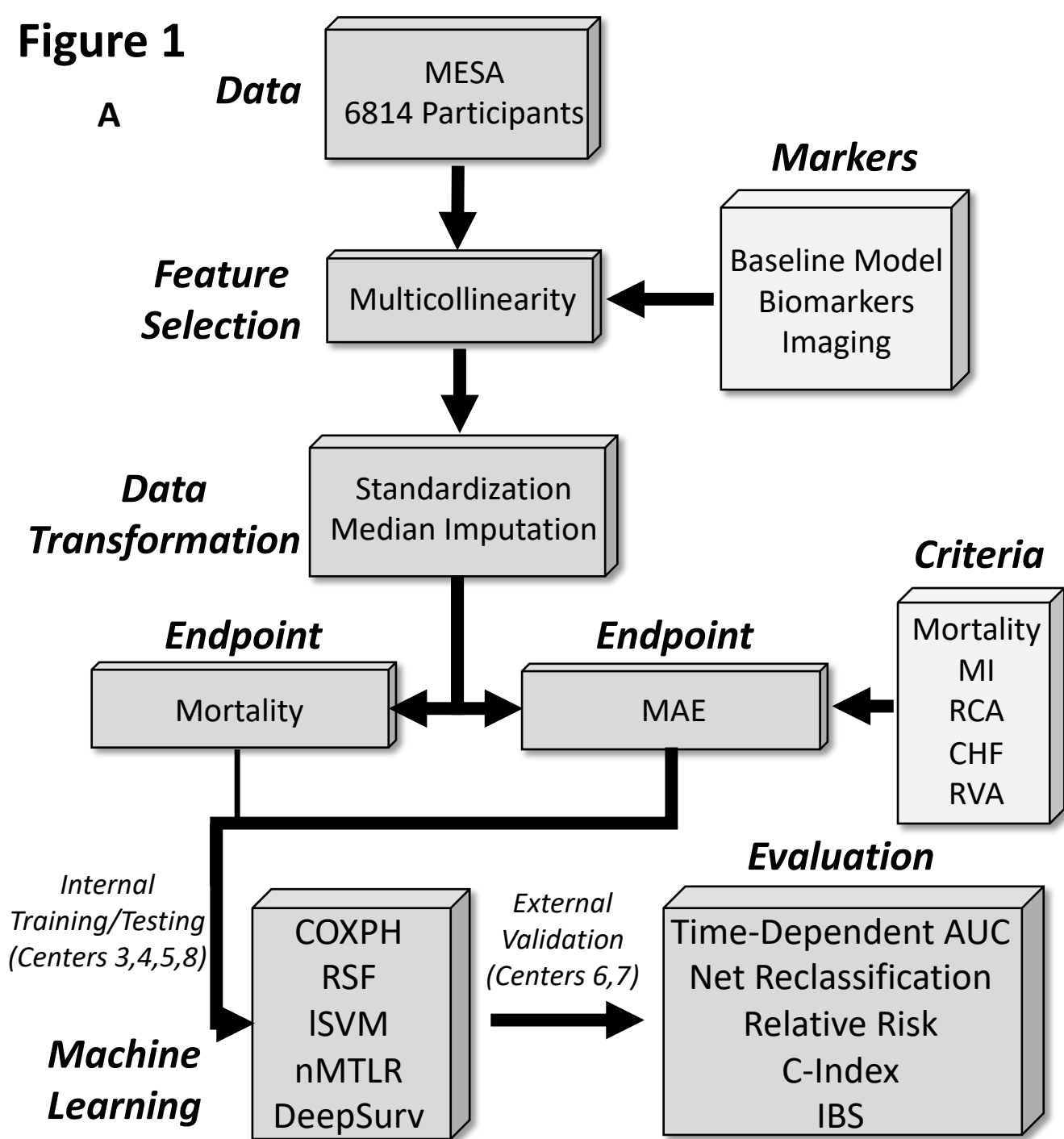
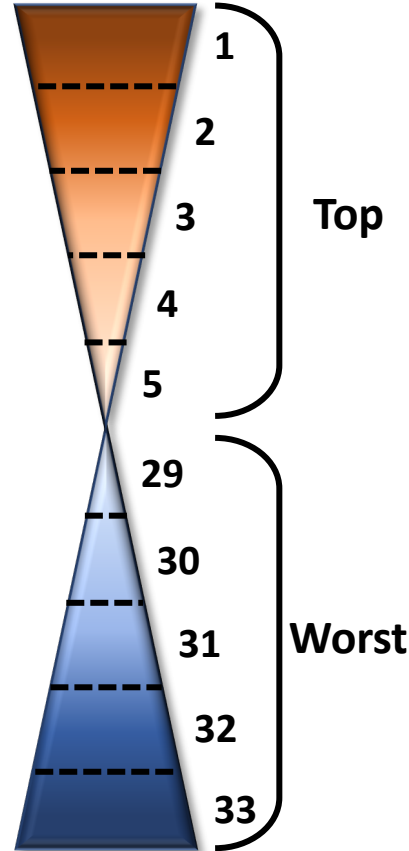
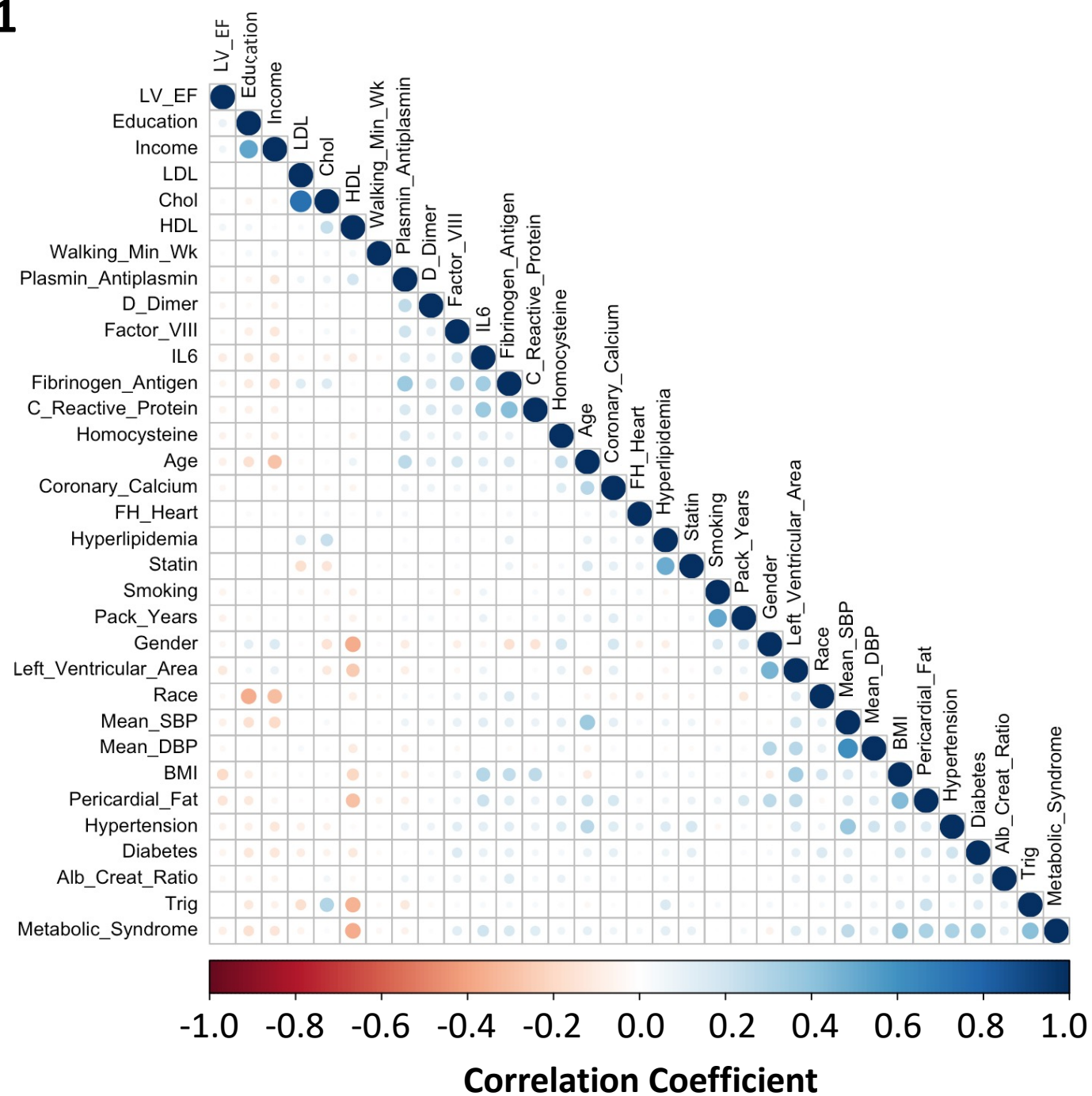
**Figure 1**

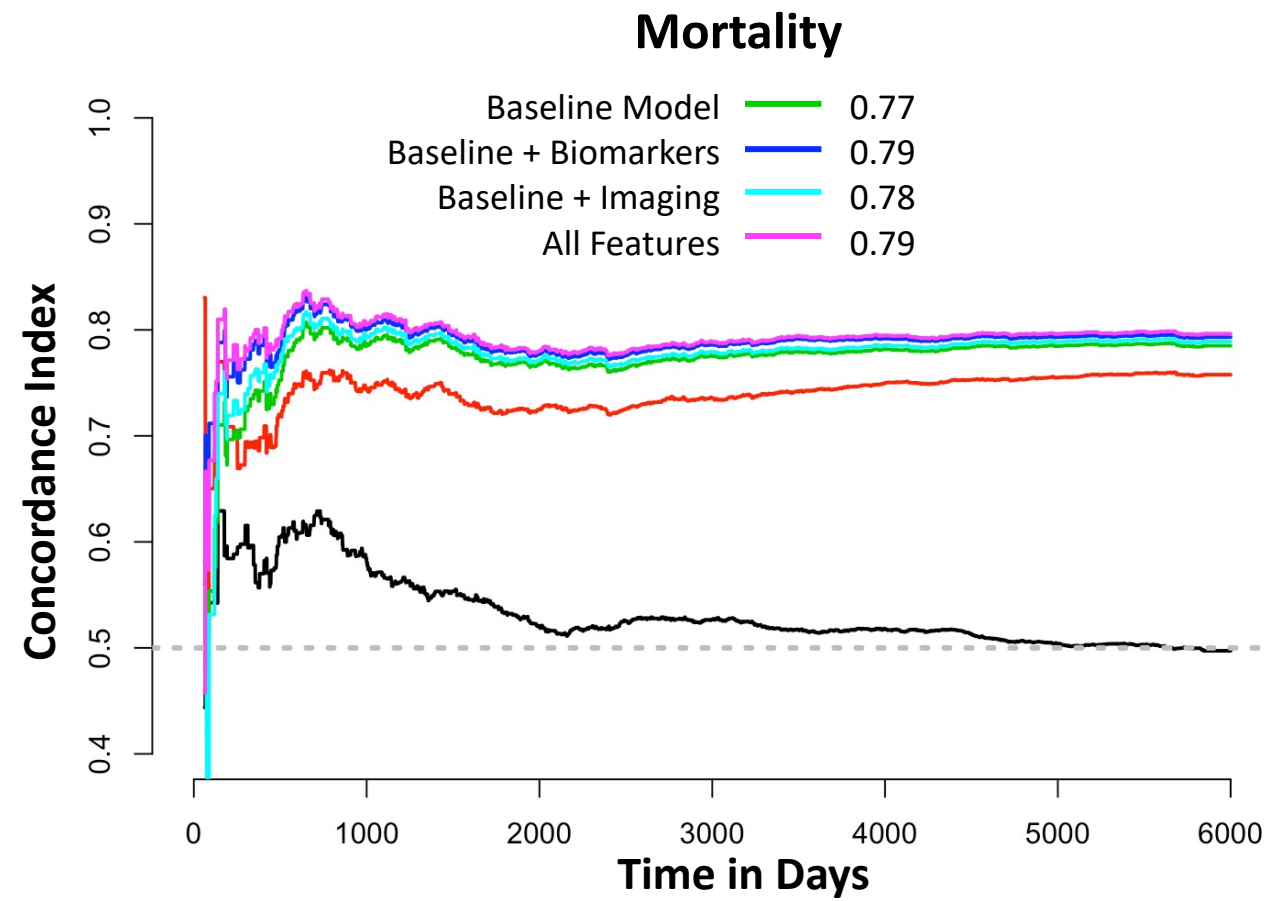
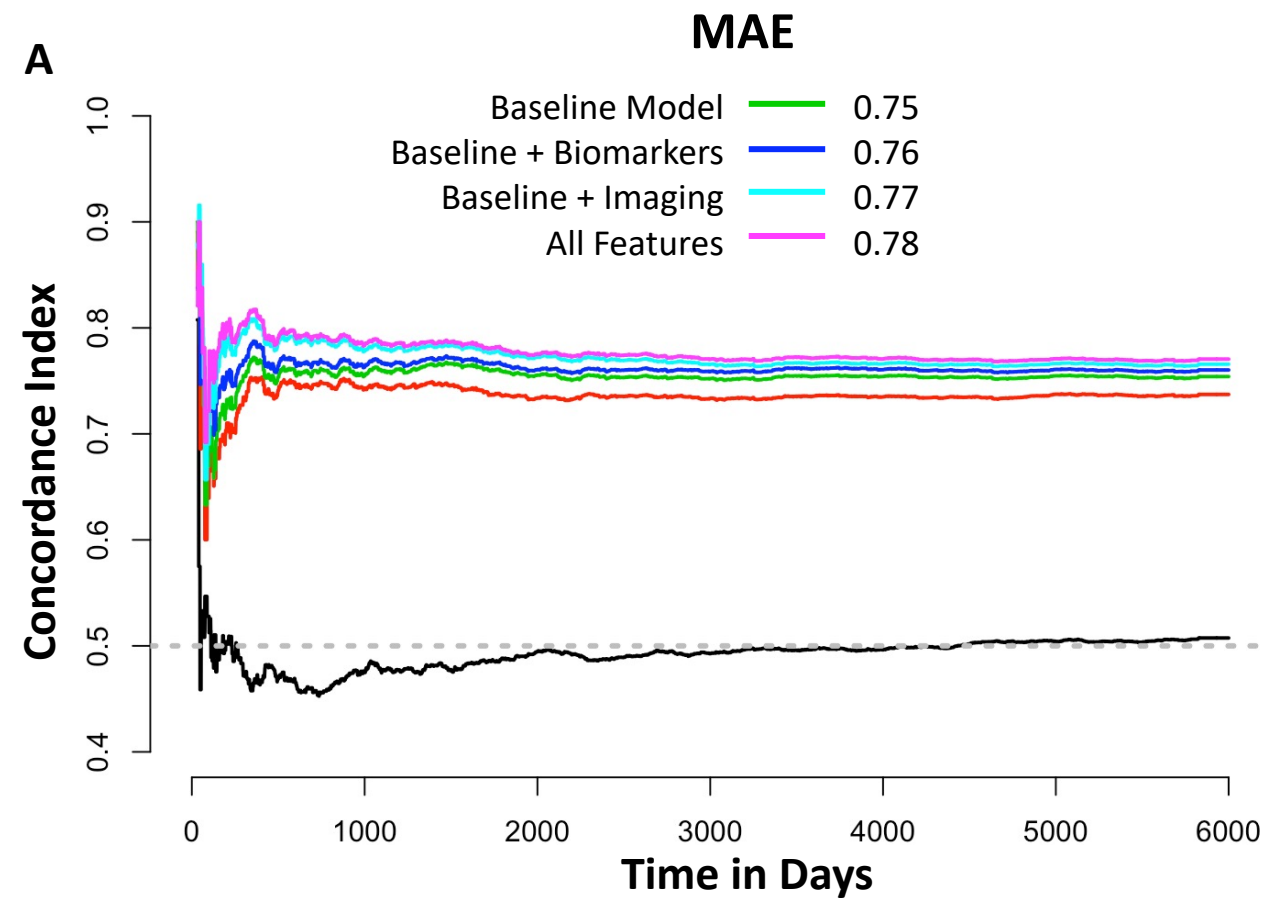
Figure 2

Feature	COXPH - MAE	nMTRL - MAE	RSF - MAE	ISVM - MAE	DeepSurv - MAE	COXPH - Mortality	nMTRL - Mortality	RSF - Mortality	ISVM - Mortality	DeepSurv - Mortality
Age	0.0330	0.0274	26.3293	0.1954	0.0373	0.0478	0.0302	27.7874435	0.235	0.0443
Gender	-0.0002	0.0027	9.4771	0.0666	0.0048	0.0005	0.0031	6.55639006	0.0507	0.0054
BMI	-0.0006	0.0028	1.6905	-0.0177	0.0020	-0.0021	0.001	3.56222953	-0.0116	0.0032
Education Status	0.0003	-0.0048	1.5214	0.0211	0.0018	-0.0013	-0.0042	2.08086395	0.0122	0.0012
Income Category	0.0079	-0.0004	9.3495	-0.0659	0.0004	0.0074	-0.0037	11.9580847	-0.0649	0.0065
Race/Ethnicity	0.0011	-0.0025	1.0607	-0.0340	0.0031	-0.0014	-0.0031	2.98914204	-0.0156	-0.0005
Diabetes	-0.0004	-0.0008	7.3612	0.0420	-0.0005	0.0004	-0.0036	5.317218	0.033	0.0011
Albumin/Creatinine Ratio	0.0017	0.0032	19.9846	0.0234	0.0021	0.0031	0.0014	19.6553556	0.0385	0.0034
Hypertension	-0.0007	-0.0078	9.2653	0.0264	-0.0048	-0.0039	-0.0013	7.67367689	0.0173	0.0011
Mean Diastolic Blood Pressure	0.0013	0.0053	4.5530	-0.0157	0.0023	-0.0013	0.0015	2.09412041	-0.0185	0.0062
Mean Systolic Blood Pressure	-0.0014	-0.0034	15.8568	0.0586	-0.0010	-0.0002	-0.0023	12.9244876	0.0425	0.0042
Hyperlipidemia	0.0008	-0.0008	-0.1801	-0.0150	0.0015	0.0009	0.0055	1.86370858	-0.0491	0.0035
HDL	0.0004	-0.0026	5.6802	0.0074	-0.0004	-0.0045	0.0032	3.98910591	-0.0137	0.0022
LDL	-0.0003	0.0002	1.2425	-0.0099	0.0003	0.0004	-0.0024	1.88021261	-0.0366	0.0023
Cholesterol	0.0017	-0.0024	0.3609	0.0094	-0.0002	-0.0002	-0.0002	4.56124513	0.018	0.0031
Triglycerides	-0.0028	-0.0020	2.3332	0.0177	-0.0011	-0.0004	0.0017	2.0452883	0.0096	0.0014
Statin Use	-0.0006	-0.0003	0.9573	-0.0052	-0.0015	-0.0005	-0.0009	1.88669502	0.0009	0.002
Smoking	0.0031	0.0021	6.2579	0.0689	0.0021	0.0032	0.0001	6.54691247	0.0645	0.0086
Smoking – Pack Years	-0.0003	0.0030	14.3605	0.0219	0.0057	-0.0024	0.0002	15.8650873	0.0332	0.0024
Walking Minutes a Week	0.0022	0.0005	2.7788	-0.0060	-0.0001	-0.0008	-0.0023	3.90569267	0.0014	0.0021
Metabolic Syndrome	0.0015	-0.0032	6.4625	0.0338	-0.0034	-0.0014	-0.0015	4.35484533	0.0173	0.0021
Family History of Heart Disease	0.0004	-0.0050	-1.0019	0.0221	0.0000	-0.0019	-0.0017	-0.1404428	-0.0049	0.0002
Homocysteine	0.0002	0.0025	17.2868	0.0308	0.0032	0.0002	0.0004	18.9550095	0.0412	0.0062
Interleukin-6 (IL6)	-0.0023	-0.0004	12.9540	0.0360	0.0028	0.0007	0.0002	14.6470494	0.0549	0.0017
Plasmin Antiplasmin	0.0082	-0.0013	11.4076	0.0267	0.0008	-0.0002	0.0023	13.1955998	0.0327	0.0052
Fibrinogen Antigen	-0.0001	-0.0047	6.7405	0.0064	-0.0009	0	-0.0032	5.00122664	0.0004	0.0026
C Reactive Protein	-0.0035	0.0015	2.3863	0.0102	0.0013	-0.001	0.0013	3.73310542	0.0086	-0.0003
D Dimer	0.0034	0.0008	13.5513	0.0191	-0.0013	0.0032	0.0025	17.5742785	0.0307	-0.0012
Factor VIII	0.0001	-0.0016	7.0988	0.0216	0.0001	0.0017	-0.0044	7.24939309	0.0116	0.0001
Left Ventricular Ejection Fraction	-0.0003	-0.0058	6.8318	-0.0271	-0.0017	-0.002	-0.003	4.14990122	-0.0098	0.0024
Left Ventricular Area	0.0051	-0.0007	10.9476	0.0478	0.0006	-0.001	-0.0002	2.66629332	0.0132	0.0035
Pericardial Fat	0.0013	-0.0020	13.8316	0.0114	-0.0014	0.0002	0.0013	10.3053957	0.0048	0.0014
Coronary Artery Calcium	0.0216	0.0196	28.5456	0.0657	0.0223	0.002	-0.0005	21.2209562	0.0339	0.006



# Supplemental Figure 1





**Supplemental Figure 2**

# Supplemental Figure 3

A

## COXPH

```
fit(X, T, E, init_method='glorot_normal', lr = 1e-2, max_iter = 100, l2_reg = 1e-4, alpha = 0.95, tol = 1e-3, epsilon=1e-9, verbose = True, display_loss=True)
```

## RSF

```
rsf = RandomSurvivalForestModel(num_trees=500)  
fit(X, T, E, max_features = 'sqrt', max_depth = 5, min_node_size = 20, num_threads = -1, weights = None, sample_size_pct = 0.63, importance_mode = 'normalized_permutation', seed = None, save_memory=False)
```

## ISVM

```
fit(X, T, E, with_bias = True, init_method='he_uniform', lr = 0.5, max_iter = 100, l2_reg = 1e-3, tol = 1e-3, verbose = True)
```

## nMTLR

```
fit(X, T, E, init_method = 'orthogonal', optimizer = 'rprop', lr = 1e-3, num_epochs = 500, dropout = 0.2, l2_reg=1e-2, l2_smooth=1e-2, batch_normalization=False, bn_and_dropout=False, verbose=True, extra_pct_time = 0.1, is_min_time_zero=True)
```

## DeepSurv

```
fit(X, T, E, init_method = 'glorot_uniform', optimizer = 'adam', lr = 1e-4, num_epochs = 1000, dropout = 0.2, l2_smooth=1e-2, batch_normalization=False, bn_and_dropout=False, display_loss = True, verbose=True, extra_pct_time = 0.1)
```

B

## nMTLR

*Hidden Layers*

ReLU

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases}$$

C

## DeepSurv

*Hidden Layers*

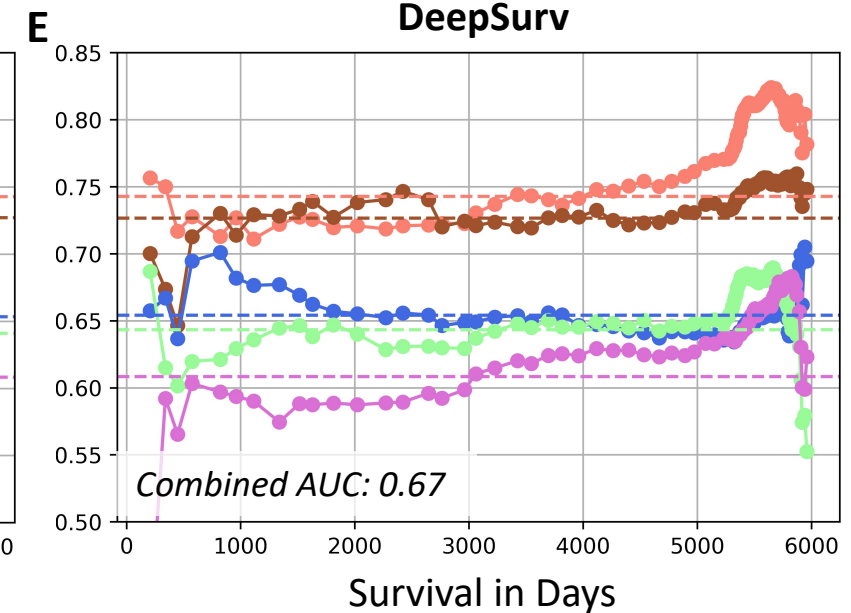
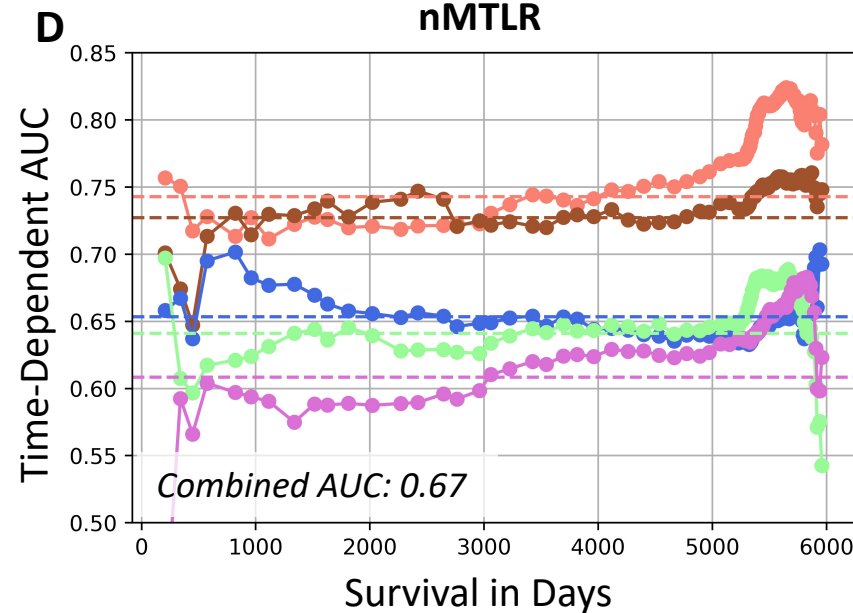
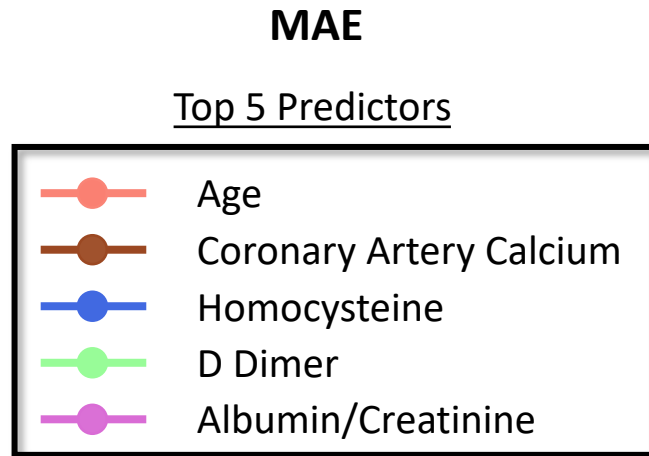
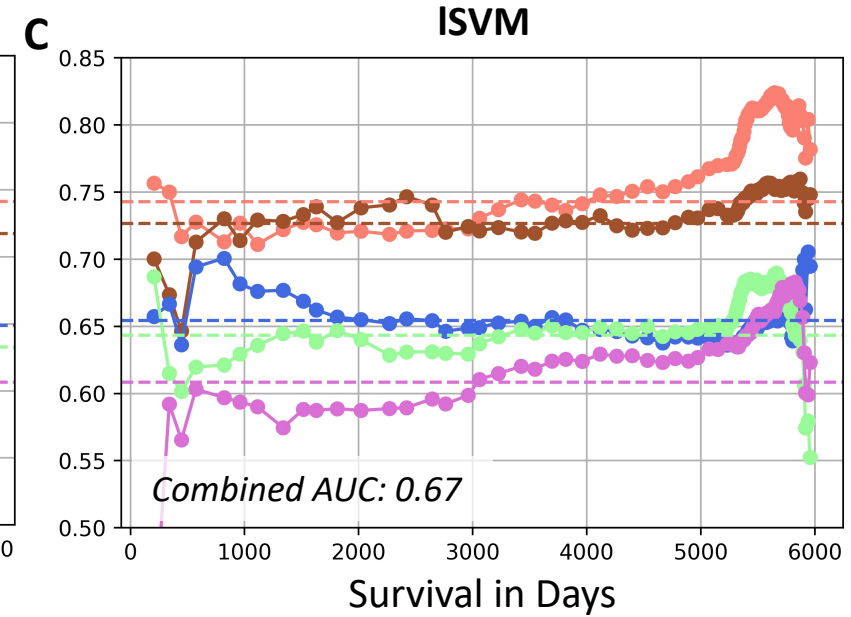
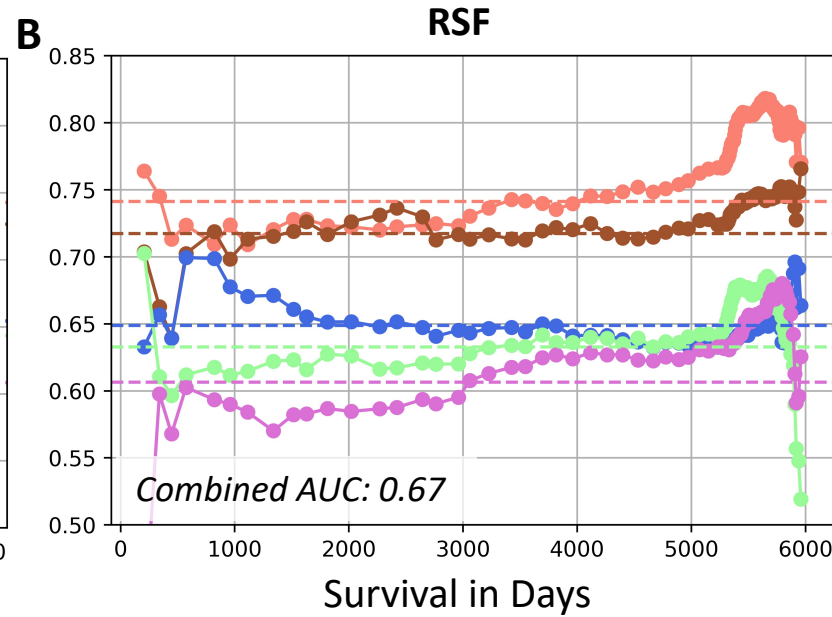
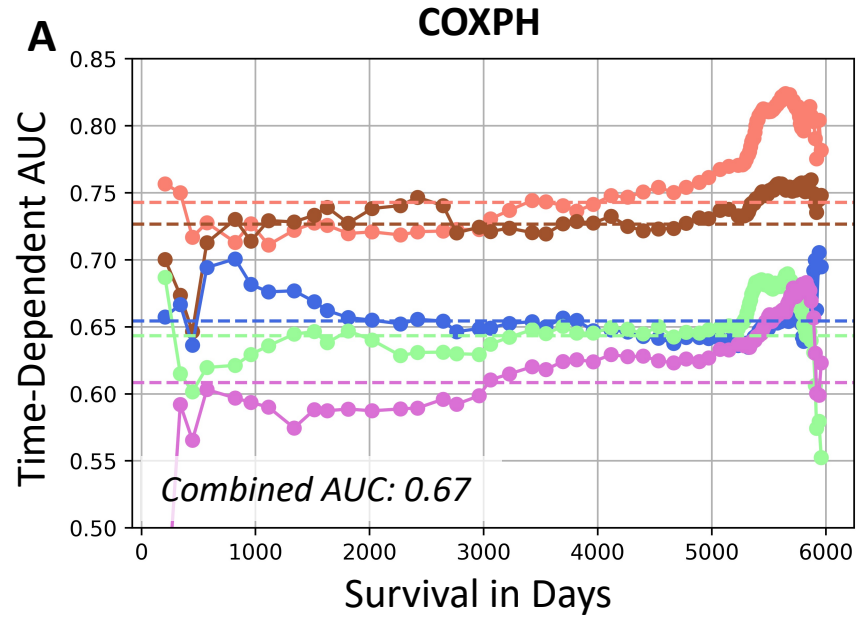
Tanh

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Hardtanh

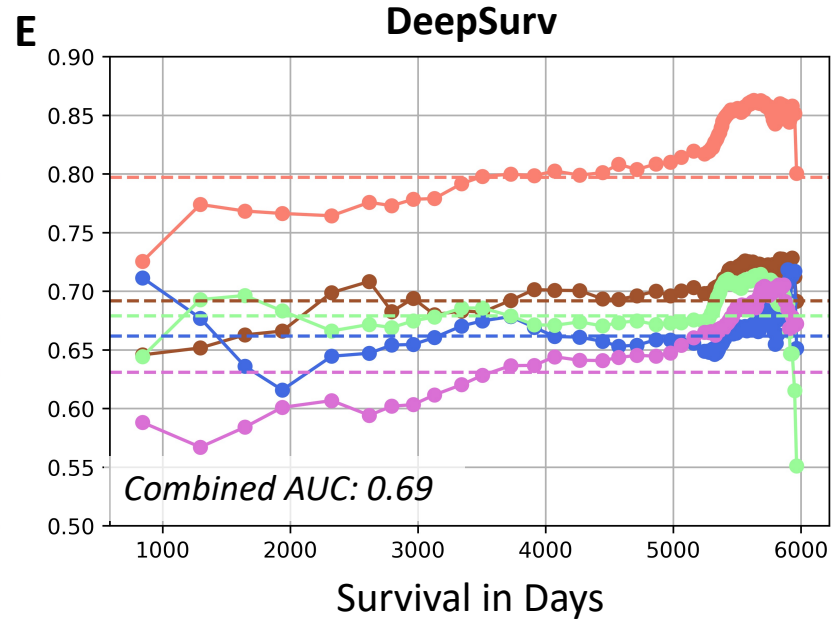
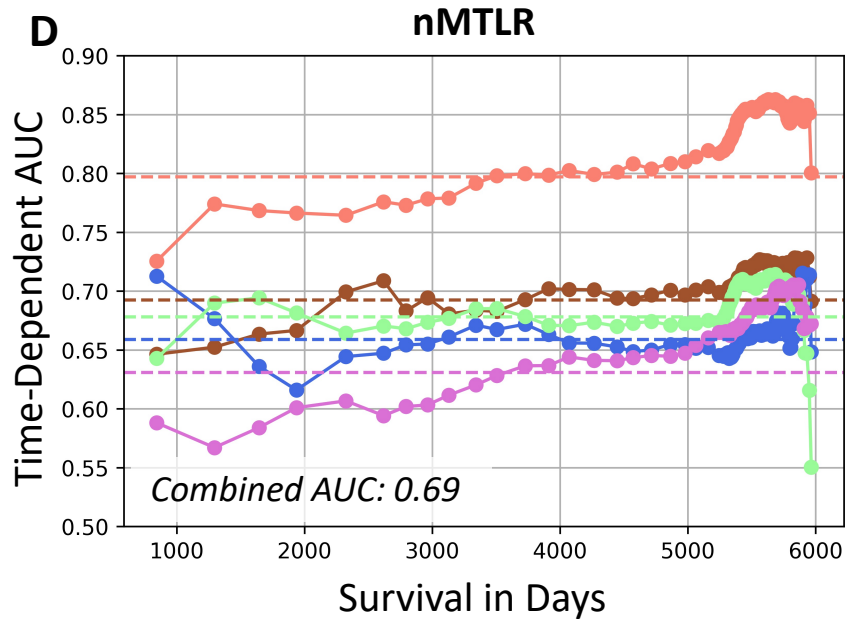
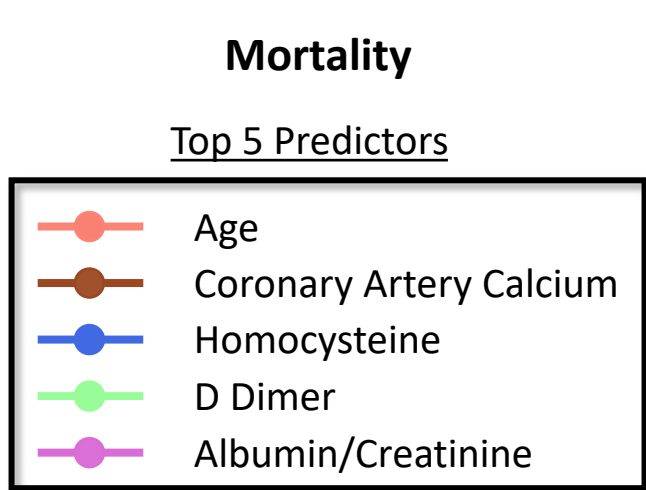
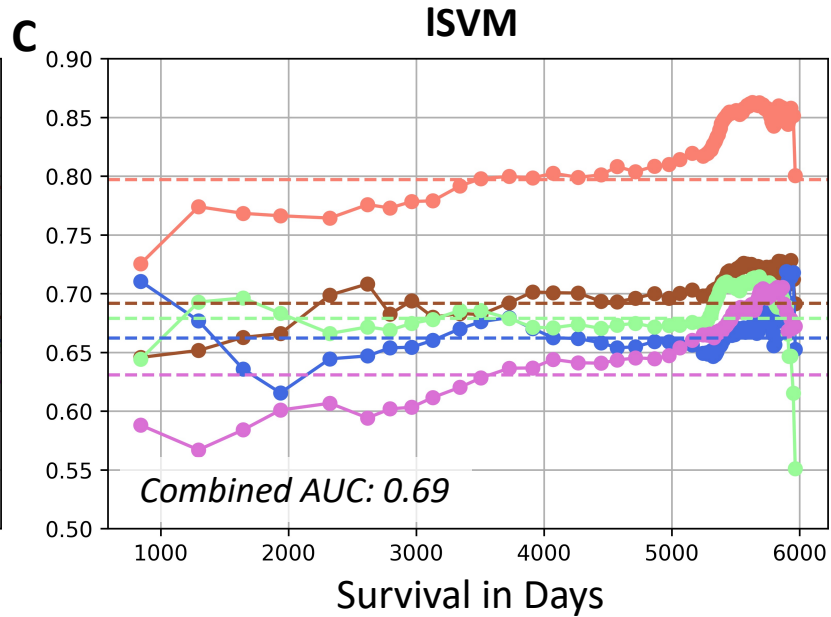
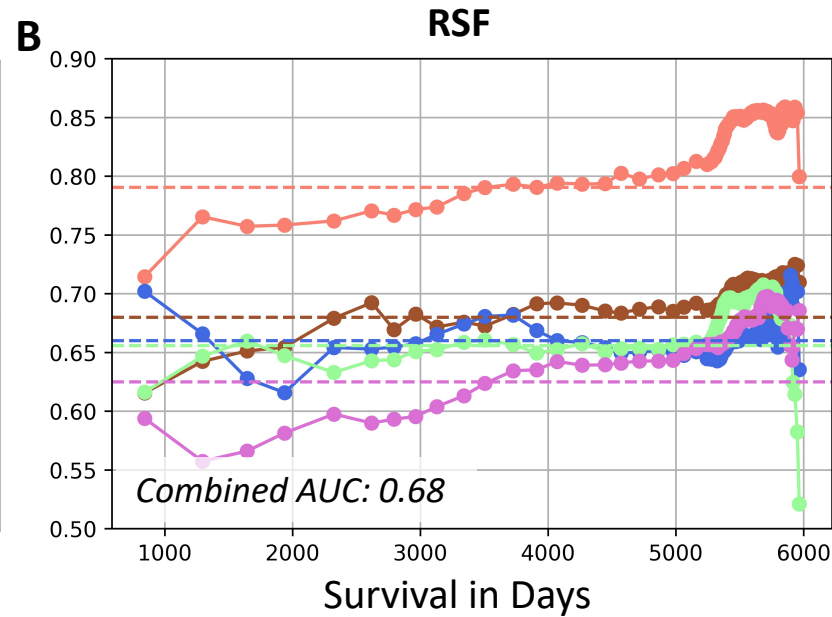
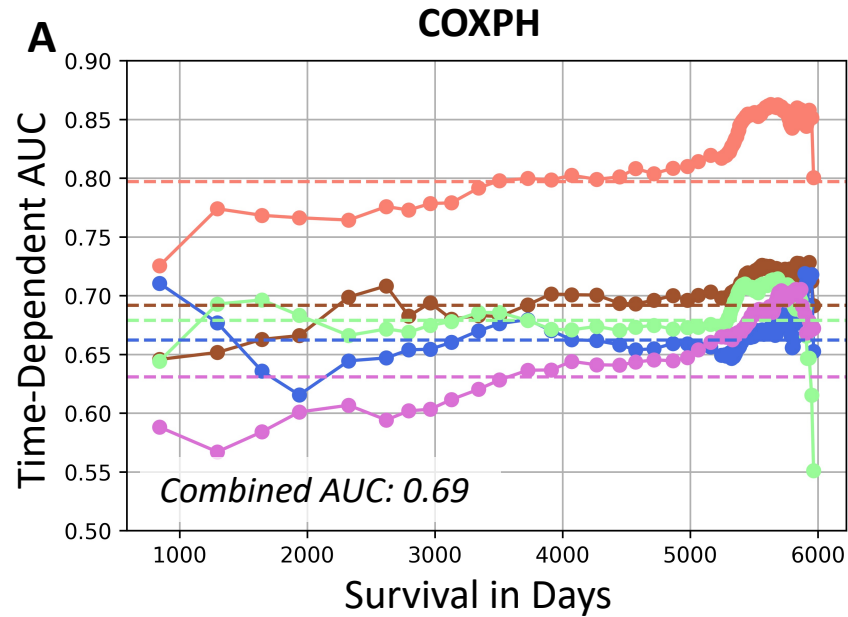
$$f(x) = \begin{cases} +1, & \text{if } x > 1 \\ -1, & \text{if } x < -1 \\ x, & \text{otherwise} \end{cases}$$

# Supplemental Figure 4

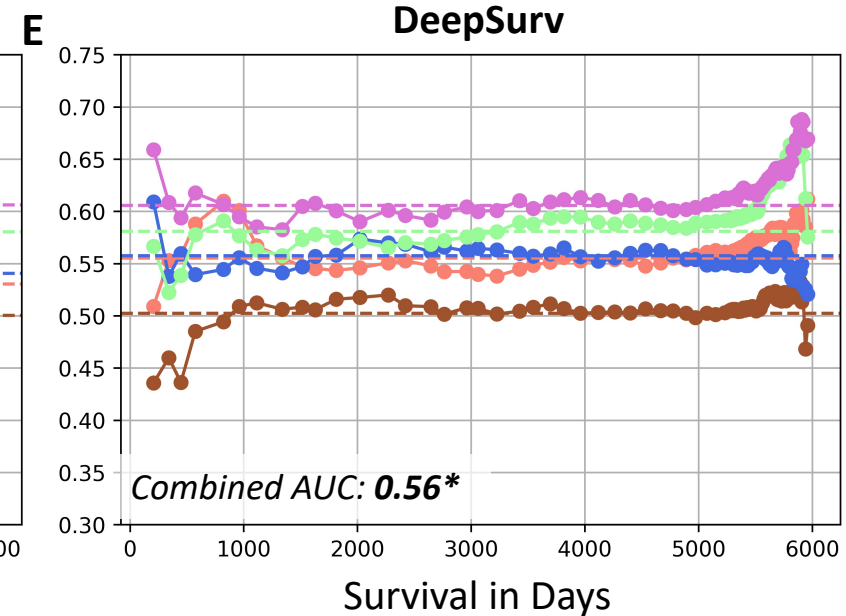
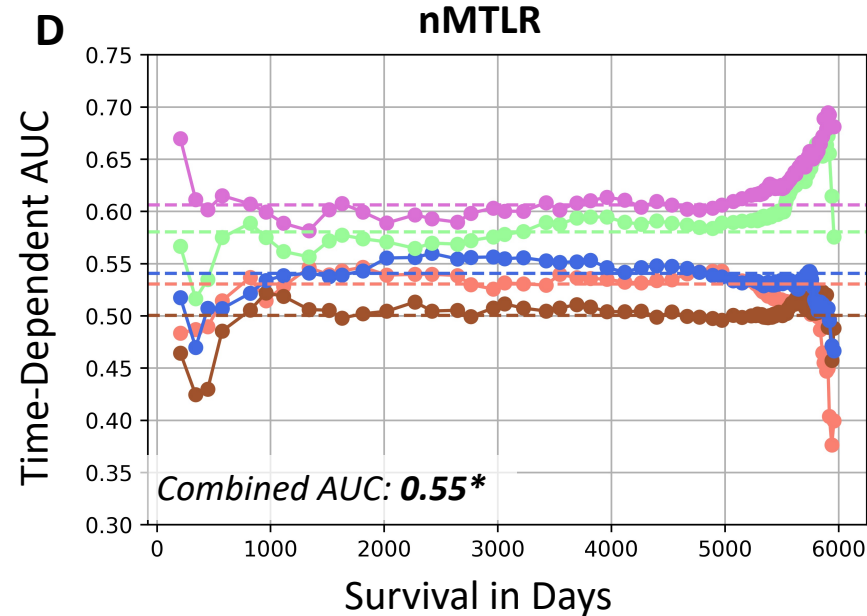
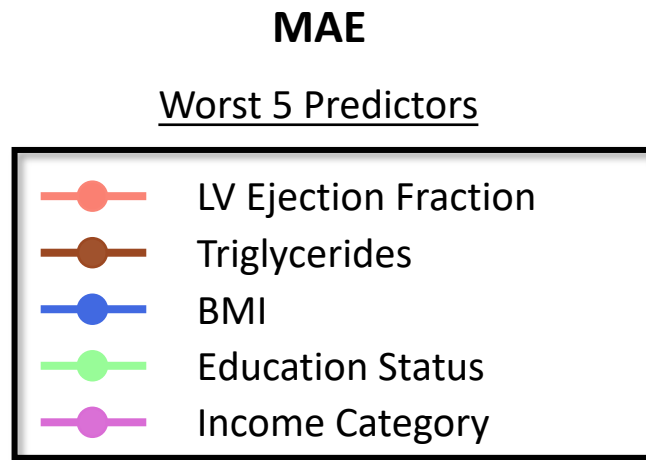
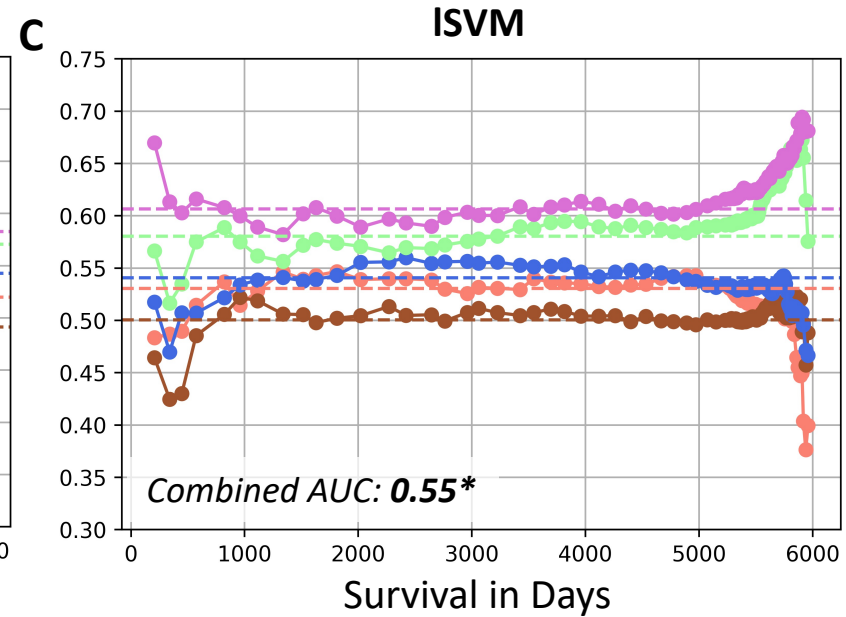
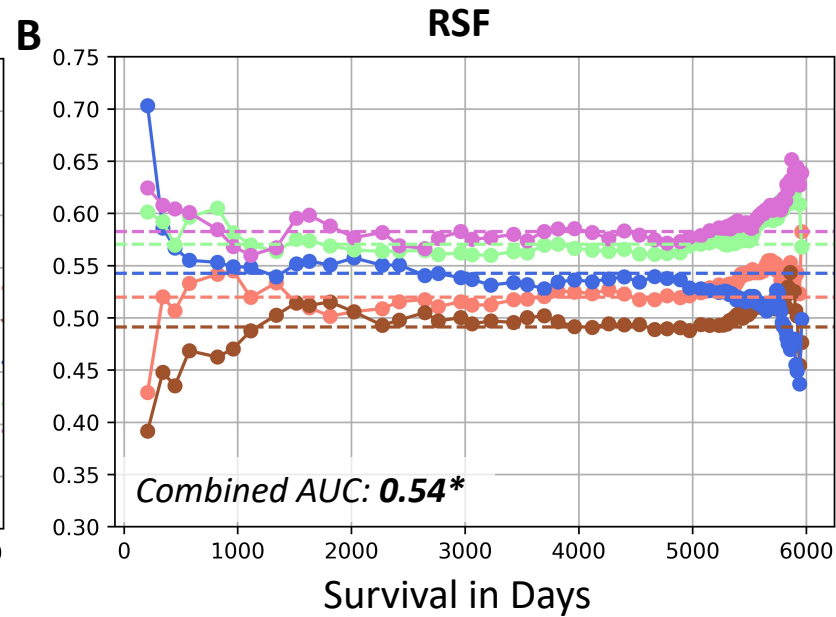
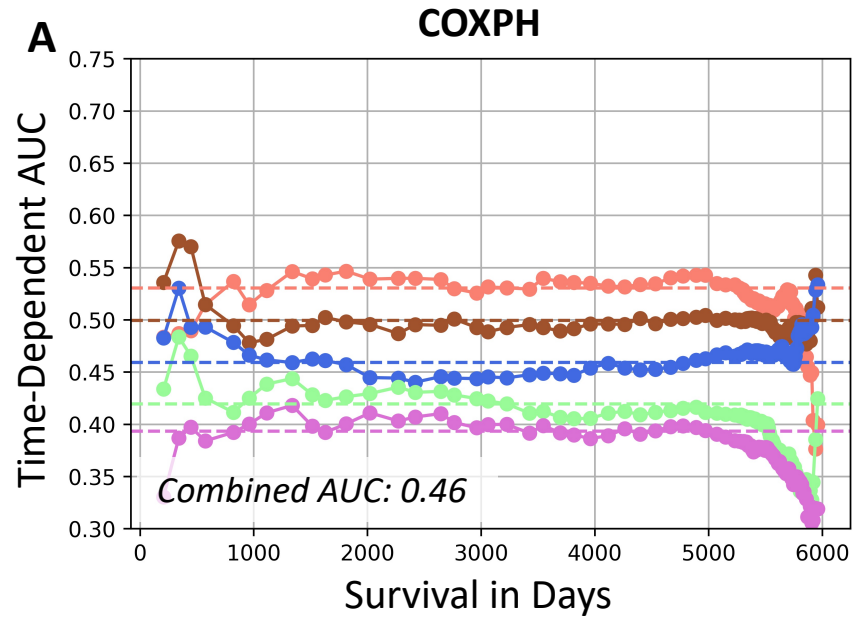




# Supplemental Figure 5



# Supplemental Figure 6





# Supplemental Figure 7

