Supplementary Online Content

Khera R, Haimovich J, Hurley NC, et al. Use of machine learning models to predict death after acute myocardial infarction. *JAMA Cardiol.* Published online March 10, 2021. doi:10.1001/jamacardio.2021.0122

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods. Supplementary Methods

Mean squared prediction error

To calculate the prediction error, we created a Bernoulli distribution using the probability of the inverse logit of the prediction at an individual-level, so that a prediction for an individual is either 0 or 1, and we compared the actual value of 0 or 1. A root mean square of this prediction across all patients was reported as the mean squared prediction error.

Measures of calibration

We also calculate the F-score, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV). The F-score is the harmonic mean of the sensitivity and PPV at a certain risk threshold, which classifies an individual risk estimate as either a death (if above the threshold) or no death (if below the threshold). Once a risk threshold is set, the number of true positives, false positives, true negatives, and false negatives can be calculated and used to derive sensitivity, specificity, PPV, and NPV. Here, the risk threshold associated with the highest F-score was selected as the overall risk-threshold for the method, and used to determine the sensitivity, specificity, PPV, and NPV for the overall model. The risk threshold is therefore determined using a data-driven approach and optimized for each model. The F-score is the harmonic mean of the sensitivity and PPV at a certain risk threshold, which classifies an individual risk estimate as either a death (if above the threshold) or no death (if below the threshold).

eTable 1. Differences in Characteristics of Patients Excluded vs Included in the Analyses

	Excluded patients (N = 191,195)		Included patients (N = 755,402)	
	Mean (SD) or N (%)	**Missing**	**Mean (SD) or N (%)**	**Missing**
Demographics				
Age, years	64.4 (13.7)	0	64.6 (13.7)	0
Weight, kg	86.5 (22.3)	566	87.0 (22.1)	1421
Female	66,349 (34.7%)	0	260,200 (34.4%)	0
Race white	161,417 (84.4%)	0	640,995 (84.9%)	0
Race black	21,612 (11.3%)	0	87,089 (11.5%)	0
Medical History				
History of diabetes mellitus	64,869 (34.0%)	392	257,072 (34.0%)	144
History of hypertension	141,071 (73.9%)	257	562,423 (74.5%)	74
History of dyslipidemia	189 (78.1%)	190953	461,269 (61.1%)	127
Current/recent smoker	64,963 (34.0%)	246	253,829 (33.6%)	145
Current dialysis	29 (27.4%)	191089	68,086 (14.4%)	283305
History of MI	5,176 (2.7%)	603	19,055 (2.5%)	244
History of HF	60 (38.7%)	191040	188,297 (24.9%)	175
Prior PCI	57 (40.1%)	191053	94,897 (12.6%)	704
Prior CABG	-	191195	193,179 (25.6%)	0
History of atrial fibrillation	54 (36.2%)	191046	100,897 (13.4%)	393
Prior cerebrovascular disease	38 (24.5%)	191040	62,312 (8.3%)	519
Prior peripheral arterial disease	20,293 (10.6%)	372	91,723 (12.1%)	148
Presentation				
Presentation after cardiac arrest	8,499 (4.5%)	2047	29,458 (3.9%)	2581
In cardiogenic shock	8,180 (4.3%)	384	28,783 (3.8%)	584
In HF	22,569 (11.8%)	306	95,240 (12.6%)	529
Heart rate	84.1 (24.3)	1089	84.0 (23.9)	2216
SBP at presentation	145.7 (35.8)	1166	146.5 (35.2)	2678
Presentation ECG				

STEMI	85,634 (44.8%)	0	292,784 (38.8%)	0
New or presumed new ST-depressions	16,772 (8.8%)	0	83,555 (11.1%)	0
New or presumed new T-wave inversions	11,078 (5.8%)	0	56,791 (7.5%)	0
Transient ST-segment elevation lasting < 20 min	1,920 (1.0%)	0	8,279 (1.1%)	0
Initial laboratory values				
Troponin Ratio	7.8 (8.3)	4088	7.3 (8.1)	12071
Creatinine, mg/dl	1.3 (1.2)	1098	1.3 (1.2)	4404
Creatinine clearance, ml/min	85.4 (42.9)	1634	85.2 (42.5)	5756
Hemoglobin, g/dl	13.7 (2.2)	1141	13.8 (2.2)	4426
Outcome				
In-hospital Mortality	9432 (4.9%)	0	33,468 (4.4%)	0

eTable 2. List of Patient Variables Used in Modeling

	Model Variables
Demographics	Age*
	Weight, kg*
	BMI kg/m^2
	Sex*
	Race (White, Black, Asian, American Indian, Native Hawaiian)
	Hispanic origin
Medical History	History of diabetes mellitus*
	Diabetes control
	History of hypertension*
	History of dyslipidemia*
	Current/recent smoker*
	Current dialysis*
	Chronic lung disease*
	History of MI*
	History of heart failure*
	Prior PCI*
	Prior CABG*
	History of atrial fibrillation*
	Prior cerebrovascular disease*
	Prior peripheral artery disease*
	Prior stroke
	Prior transient ischemic attack
Presentation	After Cardiac Arrest*
	In Cardiogenic shock*
	In heart failure*
	Heart rate, bpm*
	SBP, mmHg*
Presentation ECG	ST-elevation myocardial infarction*
	New or presumed new ST-segment depression*
	New or presumed new T-wave inversion*
	Transient ST-segment elevation < 20 minutes*
	ST elevation
	Left bundle branch block
	Isolated posterior MI
Home Medications	Aspirin
	Clopidogrel
	ACE inhibitor
	Angiotensin receptor blocker
	Beta blocker
	Statin
	Non-statin lipid-lowering agent
	Prasugrel

	Warfarin
	Aldosterone blocking agent
Initial Laboratory Tests	Initial CKMB collected
	Initial Troponin collected
	Initial Creatinine collected
	Initial Hemoglobin collected
	Lipid panel collected
	Initial BNP collected
	Initial pro-BNP collected
	Troponin Ratio*
	Creatinine mg/dL*
	Creatinine Clearance*
	Hemoglobin, g/dL*

*denotes model variables used in McNamara et al. study.[9] Creatinine clearance calculated via Cockgroft-Gault equation.

eTable 3. Shift Table Representing Actual Observed Event Rates for Pairs of Models

3A: Logistic regression (LR) vs extreme gradient descent boosting (XGB), both trained with limited variables used by McNamara et al.

	McNamara LR			
	<1%	**1-5%**	**>5%**	**All**
McNamara XGB	**N Patients Observed Rate**	**N Patients Observed Rate**	**N Patients Observed Rate**	**N Patients Observed Rate**
<1%	68,818 0.33%	25,228 0.80%	179 0.00%	94,225 0.46%
1-5%	5,577 1.42%	51,584 2.47%	6,760 4.87%	63,921 2.63%
>5%	135 2.22%	4,809 6.09%	27,394 20.81%	32,338 18.54%
All	74,530 0.42%	81,621 2.16%	34,333 17.56%	190,484 4.26%

3B: Logistic regression (LR) vs Neural Network (NN), both trained with limited variables used by McNamara et al.

	McNamara LR			
	<1%	**1-5%**	**>5%**	**All**
McNamara NN	**N Patients Observed Rate**	**N Patients Observed Rate**	**N Patients Observed Rate**	**N Patients Observed Rate**
<1%	72,909 0.40%	30,143 1.13%	31 9.68%	103,083 0.62%
1-5%	1,596 1.13%	48,772 2.59%	7,009 6.38%	57,377 3.02%
>5%	25 0.00%	2,706 5.95%	27,293 20.44%	30,024 19.12%
All	74,530 0.42%	81,621 2.16%	34,333 17.56%	190,484 4.26%

3C: Logistic regression (LR) vs XGBoost (XGB), both trained with expanded variables from the Chest pain-MI registry.

	Expanded LR			
	<1%	**1-5%**	**>5%**	**All**
	N Patients	**N Patients**	**N Patients**	**N Patients**

Expanded XGB	Observed Rate	Observed Rate	Observed Rate	Observed Rate
<1%	65,193 0.27%	31,971 0.65%	422 1.18%	97,586 0.40%
1-5%	3,384 0.95%	44,486 2.21%	13,155 3.91%	61,025 2.51%
>5%	68 2.94%	2,899 6.21%	28,906 20.79%	31,873 19.42%
All	68,645 0.30%	79,356 1.73%	42,483 15.37%	190,484 4.26%

3D: Logistic regression (LR) vs Neural Network (NN), both trained with expanded variables from the Chest pain-MI registry.

	Expanded LR			
	<1%	1-5%	>5%	All
Expanded NN	N Patients Observed Rate	N Patients Observed Rate	N Patients Observed Rate	N Patients Observed Rate
<1%	63,271 0.27%	18,909 0.67%	91 4.40%	82,271 0.37%
1-5%	5,349 0.67%	53,097 1.79%	7,307 4.83%	65,753 2.04%
>5%	25 0.00%	7,350 4.00%	35,085 17.59%	42,460 15.23%
All	68,645 0.30%	79,356 1.73%	42,483 15.37%	190,484 4.26%

3E: Logistic regression (LR) trained with limited variables vs LR trained with expanded variables from the Chest pain-MI registry.

	McNamara LR			
	<1%	**1-5%**	**>5%**	**All**
Expanded LR	**N Patients Observed Rate**	**N Patients Observed Rate**	**N Patients Observed Rate**	**N Patients Observed Rate**
<1%	62,341 0.28%	6,303 0.48%	1 0.00%	68,645 0.30%
1-5%	12,029 0.96%	63,489 1.69%	3,838 4.77%	79,356 1.73%
>5%	160 11.88%	11,829 5.60%	30,494 19.17%	42,483 15.37%
All	74,530 0.42%	81,621 2.16%	34,333 17.56%	190,484 4.26%

3F: Logistic regression (LR) trained with limited variables vs XgBoost (XgB) trained with expanded variables from the Chest pain-MI registry.

	McNamara LR			
	<1%	**1-5%**	**>5%**	**All**
Expanded XGB	**N Patients Observed Rate**	**N Patients Observed Rate**	**N Patients Observed Rate**	**N Patients Observed Rate**
<1%	67,202 0.29%	30,104 0.62%	280 2.14%	97,586 0.40%
1-5%	7,050 1.38%	45,152 2.36%	8,823 4.17%	61,025 2.51%
>5%	278 7.19%	6,365 8.09%	25,230 22.42%	31,873 19.42%
All	74,530 0.42%	81,621 2.16%	34,333 17.56%	190,484 4.26%

3G: Logistic regression (LR) trained with limited variables vs Neural Network (NN) trained with expanded variables from the Chest pain-MI registry.

	McNamara LR			
	<1%	**1-5%**	**>5%**	**All**
Expanded NN	**N Patients Observed Rate**	**N Patients Observed Rate**	**N Patients Observed Rate**	**N Patients Observed Rate**
<1%	66,541 0.31%	15,688 0.60%	42 0.00%	82,271 0.37%
1-5%	7,892 1.25%	54,195 1.94%	3,666 5.21%	65,753 2.04%
>5%	97 4.12%	11,738 5.31%	30,625 19.07%	42,460 15.23%
All	74,530 0.42%	81,621 2.16%	34,333 17.56%	190,484 4.26%

3H: Logistic regression (LR) trained with limited variables vs Meta classifier model (Meta) trained with expanded variables from the Chest pain-MI registry.

	McNamara LR			
	<1%	**1-5%**	**>5%**	**All**
Expanded Meta	**N Patients Observed Rate**	**N Patients Observed Rate**	**N Patients Observed Rate**	**N Patients Observed Rate**
<1%	67,974 0.29%	28,428 0.62%	128 0.00%	96,530 0.39%
1-5%	6,387 1.49%	45,122 2.24%	6,180 3.87%	57,689 2.33%
>5%	169 9.47%	8,071 7.19%	28,025 20.66%	36,265 17.61%
All	74,530 0.42%	81,621 2.16%	34,333 17.56%	190,484 4.26%

Three categories of predicted risk based on the logistic regression are compared against the predicted risk for the same patients using XGBoost model, neural network, and the meta-classifier (bottom third), further stratified based on the variables used to train the model (variables in the model by McNamara et al vs the expanded variable set). Event rate is reported as a percentage for each cohort, and the cohort size is shown in parentheses.

eTable 4. Shift Tables for Sensitivity Analysis With Risk Thresholds for Shift Tables Set at Risk <1.5%, 1.5-3%, >3%, Each Model Trained With Expanded Variables From the Chest Pain-MI Registry

XGBoost vs Logistic Regression				
	Expanded LR			
	<1.5%	1.5-3%	>3%	All
Expanded XGBoost	Patients, N Observed rate	Patients, N Observed rate	Patients, N Observed rate	Patients, N Observed rate
<1.5%	87,507 0.36%	22,235 0.94%	5,779 1.32%	115,521 0.52%
1.5-3%	3,085 1.78%	10,828 2.19%	13,417 2.83%	27,330 2.46%
>3%	577 2.95%	2,806 3.56%	44,250 15.18%	47,633 14.35%
All	91,169 0.42%	35,869 1.52%	63,446 11.31%	190,484 4.26%

Meta-classifier vs Logistic Regression				
	Expanded LR			
	<1.5%	1.5-3%	>3%	All
Expanded Meta	Patients, N Observed rate	Patients, N Observed rate	Patients, N Observed rate	Patients, N Observed rate
<1.5%	88,360 0.37%	22,698 0.99%	4,680 1.15%	115,738 0.52%
1.5-3%	2,378 2.14%	9,836 2.09%	9,636 2.30%	21,850 2.19%
>3%	431 3.02%	3,335 3.51%	49,130 14.04%	52,896 13.28%
All	91,169 0.42%	35,869 1.52%	63,446 11.31%	190,484 4.26%

Neural Network vs Logistic Regression				
	Expanded LR			
	<1.5%	1.5-3%	>3%	All

Expanded Neural Network	Patients, N Observed rate	Patients, N Observed rate	Patients, N Observed rate	Patients, N Observed rate
<1.5%	84,335 0.37%	14,126 1.01%	2,432 2.38%	100,893 0.51%
1.5-3%	6,047 1.12%	14,359 1.73%	8,386 2.59%	28,792 1.85%
>3%	787 1.02%	7,384 2.11%	52,628 13.11%	60,799 11.62%
All	91,169 0.42%	35,869 1.52%	63,446 11.31%	190,484 4.26%

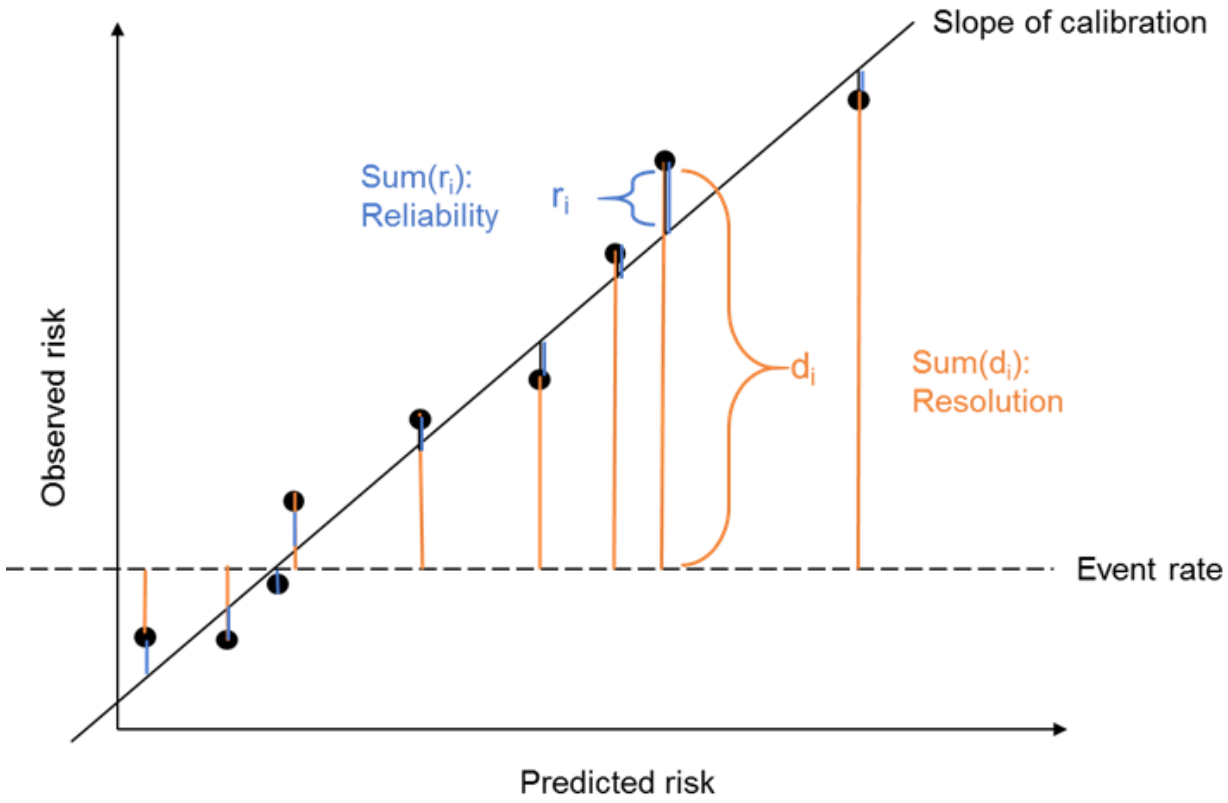
eTable 5. Area Under the Receiver Operator Characteristic Curve for the 5-Fold Multiple Imputation. Values in square brackets represents 95% confidence intervals.

Model	Models Constructed using Limited variables	Models Constructed using Expanded variables
Logistic Regression	0.877 [0.877-0.877]	0.888 [0.888-0.888]
Neural Network	0.874 [0.873-0.875]	0.886 [0.884-0.888]
XGBoost	0.885 [0.884-0.885]	0.897 [0.897-0.898]
Meta-classifier	0.885 [0.885-0.886]	0.898 [0.897-0.898]

eTable 6. Model Calibration Slopes in Patient Subgroups

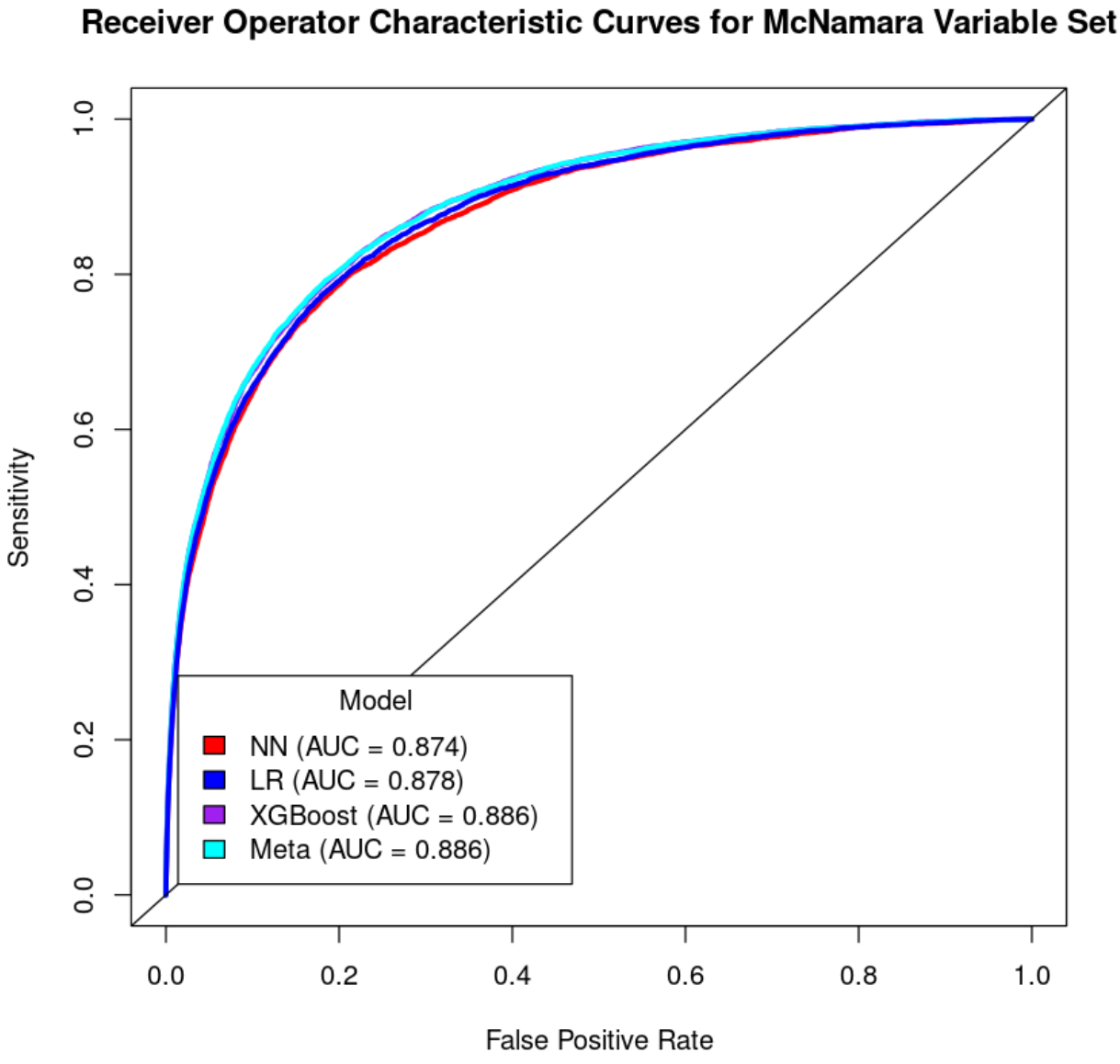
Group	Logistic regression	Neural network	XGBoost	Metaclassifier
Overall	0.93 [0.91, 0.95]	0.83 [0.82, 0.85]	0.98 [0.96, 1.00]	0.99 [0.98, 1.00]
Age in years				
18-44	0.90 [0.87, 0.93]	0.81 [0.77, 0.84]	0.98 [0.95, 1.00]	0.97 [0.94, 1.00]
45-64	0.93 [0.92, 0.94]	0.83 [0.82, 0.85]	0.97 [0.96, 0.98]	0.98 [0.96, 1.00]
≥65	0.94 [0.91, 0.97]	0.83 [0.81, 0.86]	0.99 [0.96, 1.03]	1.00 [0.99, 1.01]
Sex				
Male	0.94 [0.92, 0.95]	0.84 [0.82, 0.85]	0.98 [0.97, 1.00]	0.99 [0.98, 1.01]
Female	0.92 [0.89, 0.95]	0.82 [0.80, 0.85]	0.97 [0.94, 1.00]	0.97 [0.96, 0.99]
Race/ethnicity				
White	0.93 [0.92, 0.95]	0.83 [0.82, 0.84]	0.98 [0.96, 1.00]	0.99 [0.97, 1.00]
Black	0.95 [0.89, 1.00]	0.86 [0.83, 0.90]	1.00 [0.94, 1.06]	1.01 [0.97, 1.04]

eFigure 1. Derivation of Brier Score Components Based on Calibration Curve

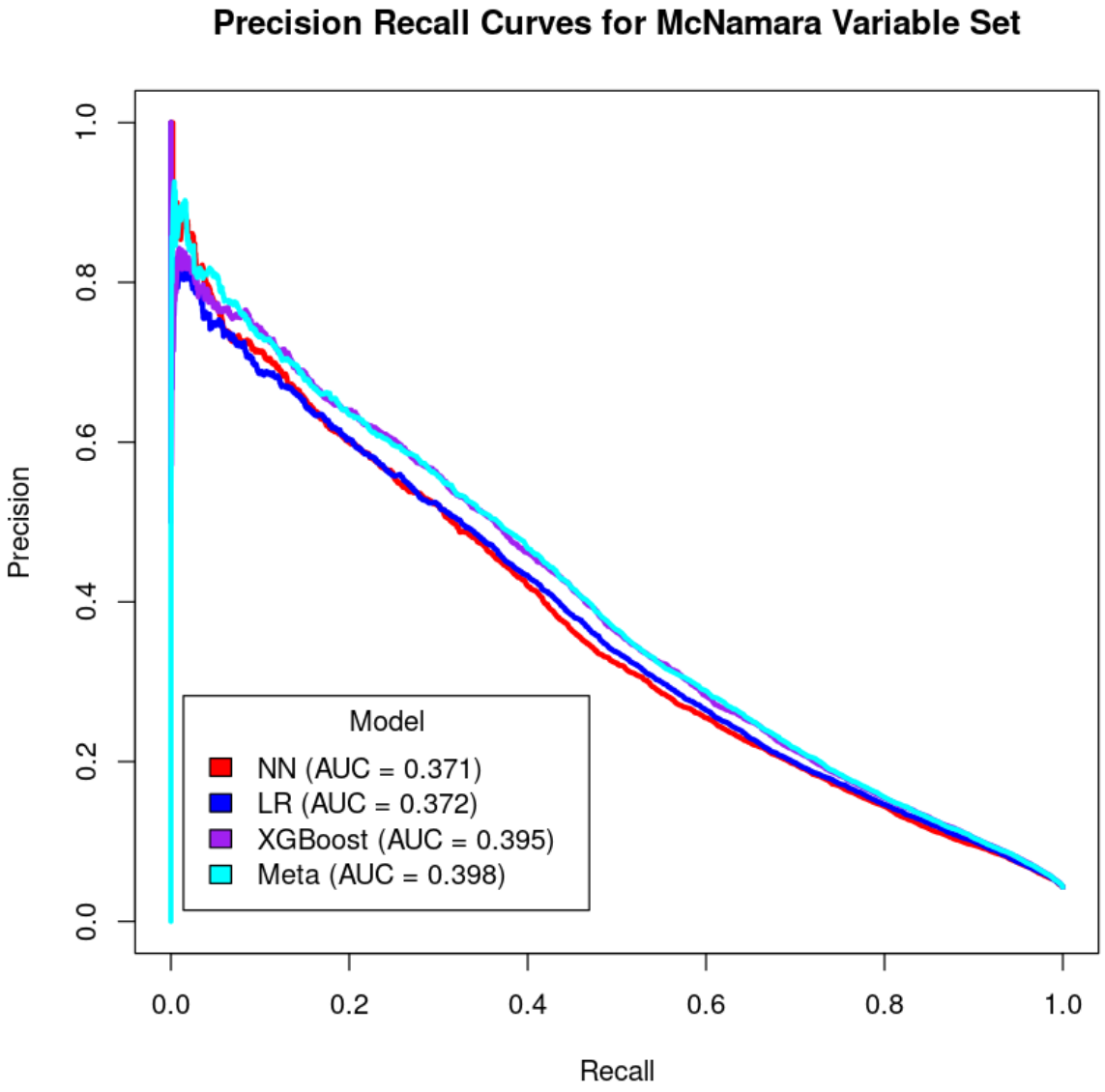


In the figure, each point represents the predicted versus Observed risk at a given decile of risk. Reliability is the sum of the mean-squared error between the deciles of predicted risk and Observed risk, and Resolution is the mean-squared error between deciles of predicted risk and the event rate of the entire cohort

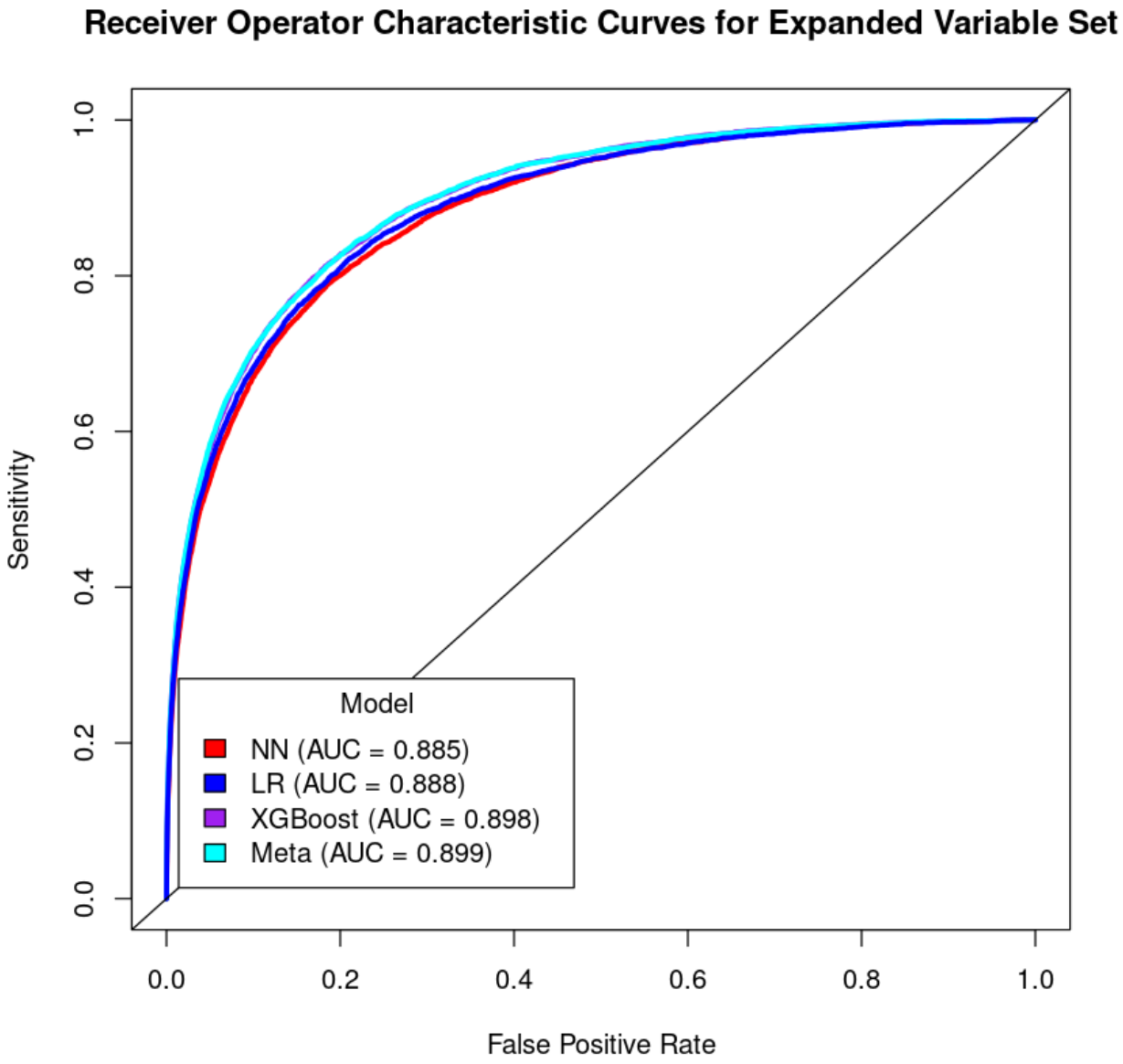
eFigure 2. Receiver Operator Characteristic for Logistic Regression, Neural Network (NN), XGBoost, and Meta-Classifier Models, Developed Using Variables Included in the Current Model for In-hospital Mortality by McNamara et al
They plot the model sensitivity against the false positive rate across a range of all possible risk thresholds for deciding the binary mortality outcome. The black line shows the performance of an imperfect (random) classifier. Area under the curve (or c-statistic) for each model is shown in the legend.



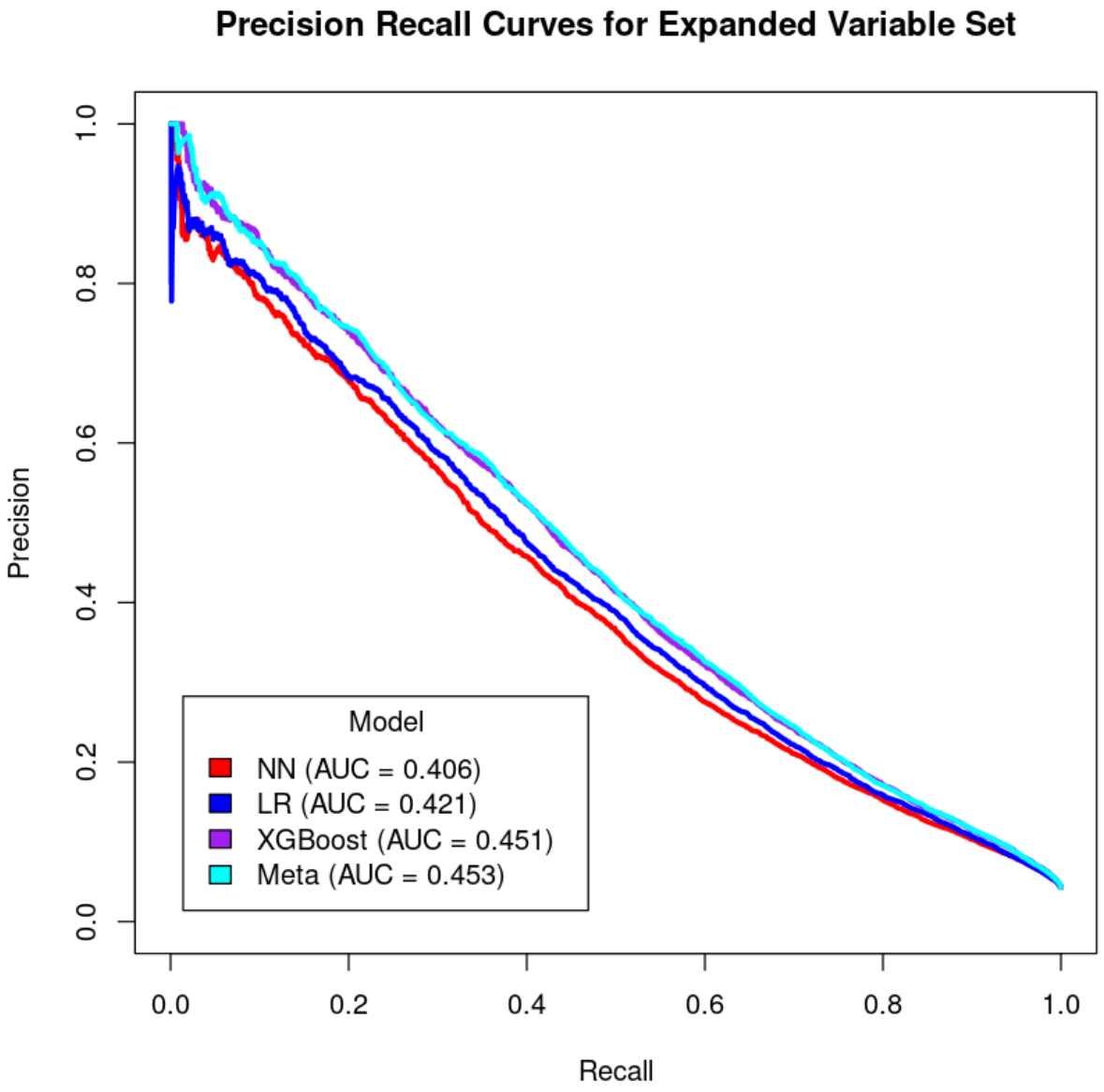
eFigure 3. Precision-Recall Curves for Logistic Regression, Neural Network (NN), XGBoost, and Meta-Classifier Models, Developed Using Variables Included in the Current Model for In-hospital Mortality by McNamara et al
Models with precision-recall curves nearest to the top right-hand corner of the graph have the best performance. Area under the curve for each model is shown in the legend.



eFigure 4. Receiver Operator Characteristic for Logistic Regression, Neural Network (NN), XGBoost, and Meta-Classifier Models, Developed Using the Expanded Set of Variables in the Chest Pain-MI Registry
They plot the model sensitivity against the false positive rate across a range of all possible risk thresholds for deciding the binary mortality outcome. The black line shows the performance of an imperfect (random) classifier. Area under the curve (or c-statistic) for each model is shown in the legend.

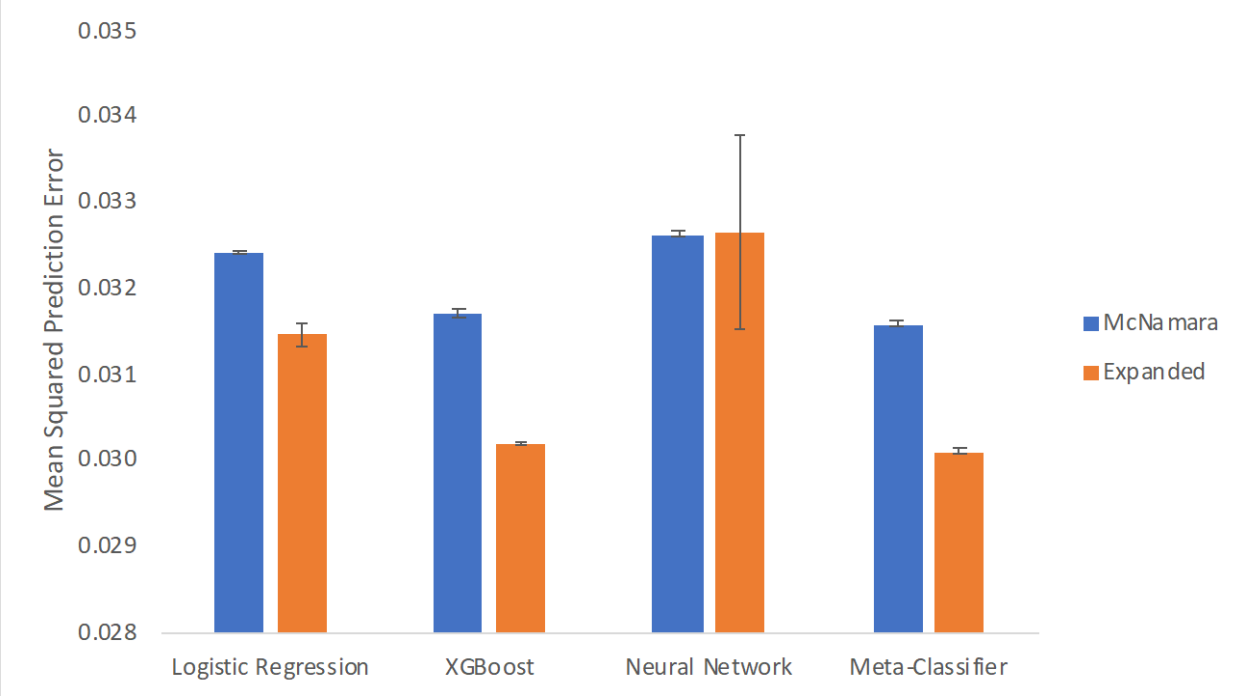


eFigure 5. Precision-Recall Curves for Logistic Regression, Neural Network (NN), XGBoost, and Meta-Classifier Models, Developed Using the Expanded Set of Variables in the Chest Pain-MI Registry
Models with precision-recall curves nearest to the top right-hand corner of the graph have the best performance. Area under the curve for each model is shown in the legend.



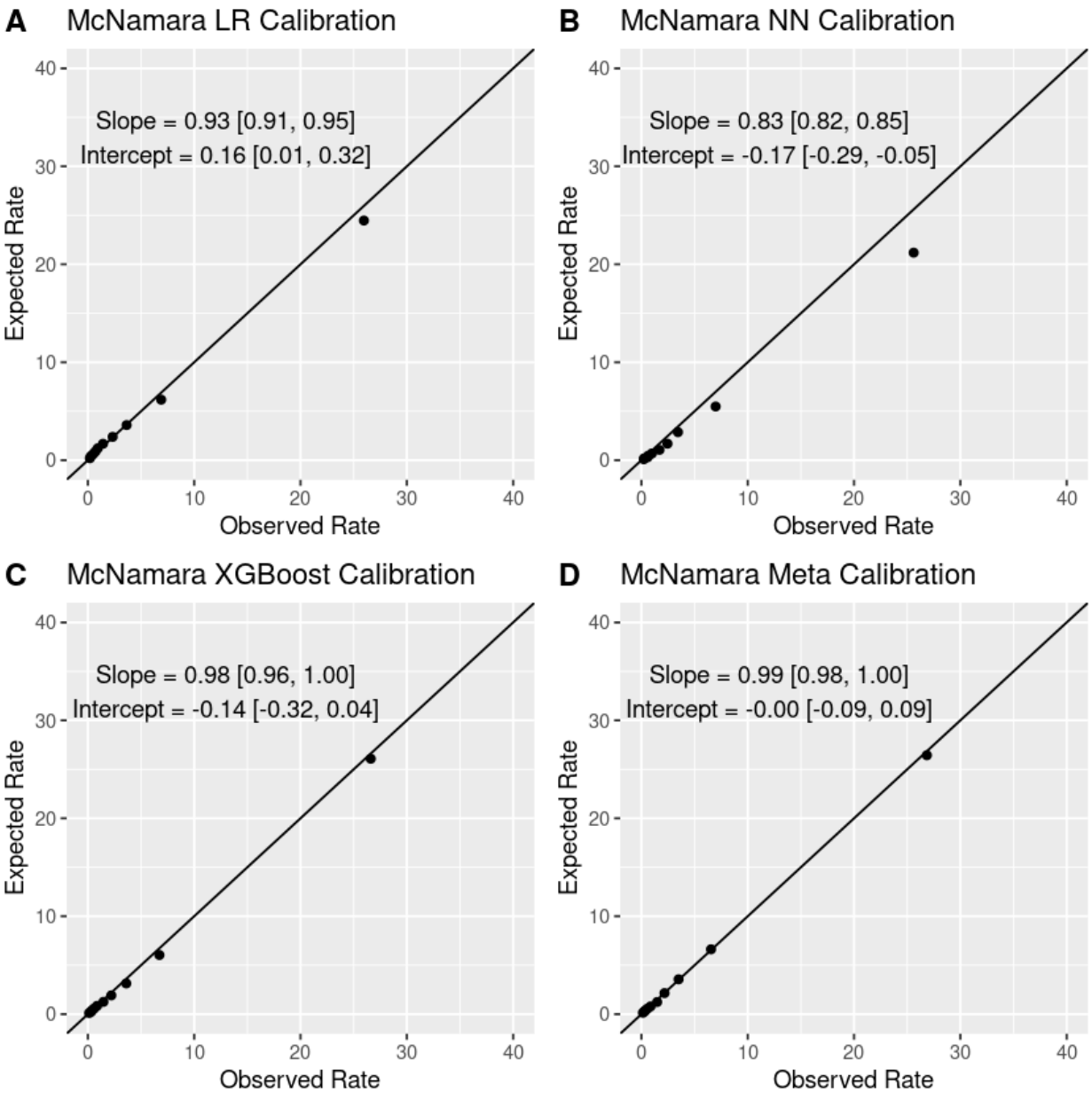
eFigure 6. Mean Squared Prediction Error of Machine Learning Models Compared With Logistic Regression

The mean squared prediction error for all machine learning models was lower than logistic regression applied to the same set of variables, including the variables used by the current standard (McNamara et al) and all variable available in the Chest pain-MI registry.



eFigure 7. Calibration of Models Developed Using Limited Number of Variables Included in the Current Standard (McNamara et al)

Calibration curves for logistic regression (LR, A), Neural Network (B), XGBoost (C) and Meta-Classifier (D) models for validation cohort predictions. Slope of 1 represents perfect model calibration with values greater than 1 suggesting overestimation of risk and less than 1 suggesting underestimation of risk.



eFigure 8. Calibration of Models Developed Using Expanded Number of Variables Included in the Chest Pain-MI Registry

Calibration curves for logistic regression (LR, A), Neural Network (B), XGBoost (C) and Meta-Classifier (D) models for validation cohort predictions. Slope of 1 represents perfect model calibration with values greater than 1 suggesting overestimation of risk and less than 1 suggesting underestimation of risk.