Mahidol University
Wisdom of the Land

Use of Machine Learning Models to Predict Death After Acute Myocardial Infarction

Jitsama Tanlamai

Journal Club 2021 June 18

JAMA Cardiol. doi:10.1001/jamacardio.2021.0122
Published online March 10, 2021.

Background and Objective

Existing risk prediction models developed in the prediction of AMI outcomes have been limited

- Lack of inclusion of nonlinear effects and complex interactions among variables in national samples
- Only evaluated the effects in small patient groups

"To evaluate whether
contemporary machine learning methods
can improve **prediction of in-hospital death**
after hospitalization for acute myocardial infarction (AMI)
by including a larger number of variables
and **identifying complex relationships**
between predictors and outcomes"

Mahidol University
Wisdom of the Land

Data Collection

- Cohort study
- Participants
  - Inclusion
    - Admit 1,128 participating hospitals for AMI
      @ 1st Jan 2021 - 31st Dec 2016 (6 years)
    - ST-elevation myocardial infarction (STEMI) or not
  - Population = 993,905

Mahidol University
Wisdom of the Land

Data Collection

- Data auditing - NCDR data quality program 2012
  - **Completeness**

    Proportion of missing data within fields
  - **Consistency**

    Logically related fields contain values consistent with other fields
  - **Accuracy**

    Agreement between registry data and the contents of original charts from the hospitals submitting data

Mahidol University
Wisdom of the Land

Data

- Output

  Categorical variable - death from **any cause** during hospitalization

- Variables
  - Categorical and Continuous variables
  - 29 variables from NCDR standard
    : **Current standard model uses 9 variables** from 29 candidate variables (use LR for the selection)
  - Additional variables
    Available variables to a practitioner at the time of hospital presentation for AMI with < 1% missing variable rate

Mahidol University
Wisdom of the Land

Data

- Variables (Cont.)

	NCDR standard	Additional	NCDR standard + Addition
Demographic	3	3	6
Medical history	13	3	16
Presentation (e.g., after cardiac arrest)	5	-	5
Presentation ECG	4	3	7
Home medications	-	10	10
Initial laboratory tests	4	7	11
Total	**29**	26	**55**

Data

- **Set #1** – Model development
  - Exclusion
    - Patients transferred to another facility for management
    - Missing a key risk factor included in the current standard for predicting mortality outcomes - history of percutaneous coronary intervention
  - Remained = 755,402
  - Imputation: Median and mode
- **Set #2** – Sensitivity test
  - Exclusion
    - Patients transferred to another facility for management
    - ~~Missing a key risk factor included in the current standard for predicting mortality outcomes - history of percutaneous coronary intervention~~
    - **(Drop variable)** Covariates with missingness > 5%
  - Remained = 946,597
  - Imputation: 5-fold multiple imputation – chained equations method (regression-based approach)

Mahidol University
Wisdom of the Land

Model Development

Data set #1 (Total n = 755,402)

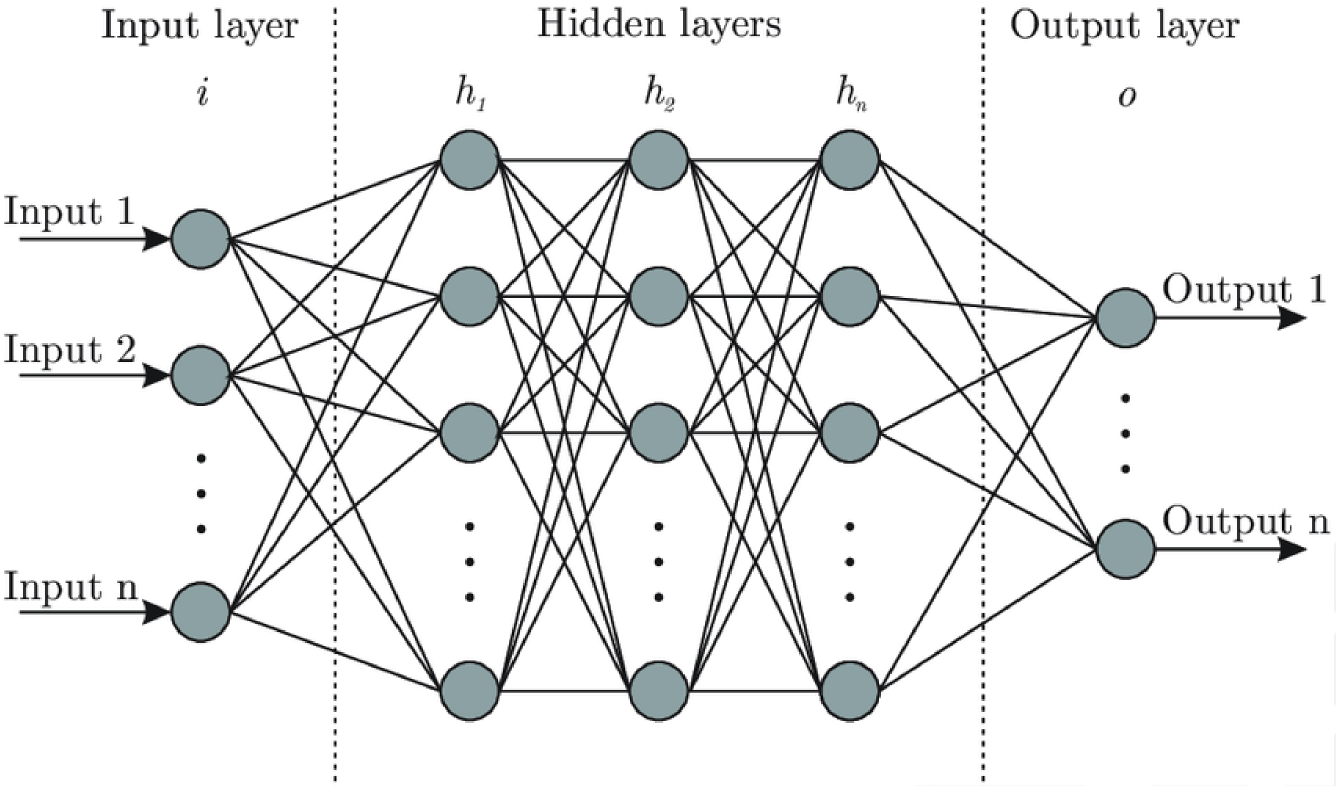
Derivation cohort (training, #1) and validation cohort (test, #2)

Set #	Period	%	n	Objective
1.1a	1st Apr 2011 – 30th Sep 2013	75	564,918	Model development Lv. 1
1.1b	1st Oct 2013 – 30th Sep 2015			Model development Lv. 2
1.2	1st Oct 2015 – 31st Dec 2016	25	190,484	Model testing

- Model comparison – 3 models with baseline
  - Baseline model: LR + LASSO
  - Tested models
    1) NN : 1.1a+b, 1.2
    2) XGBoost : 1.1a+b, 1.2
    3) Meta classifier : 1.1a, 1.1b, 1.2

Model Development

- Tested model
  1) NN

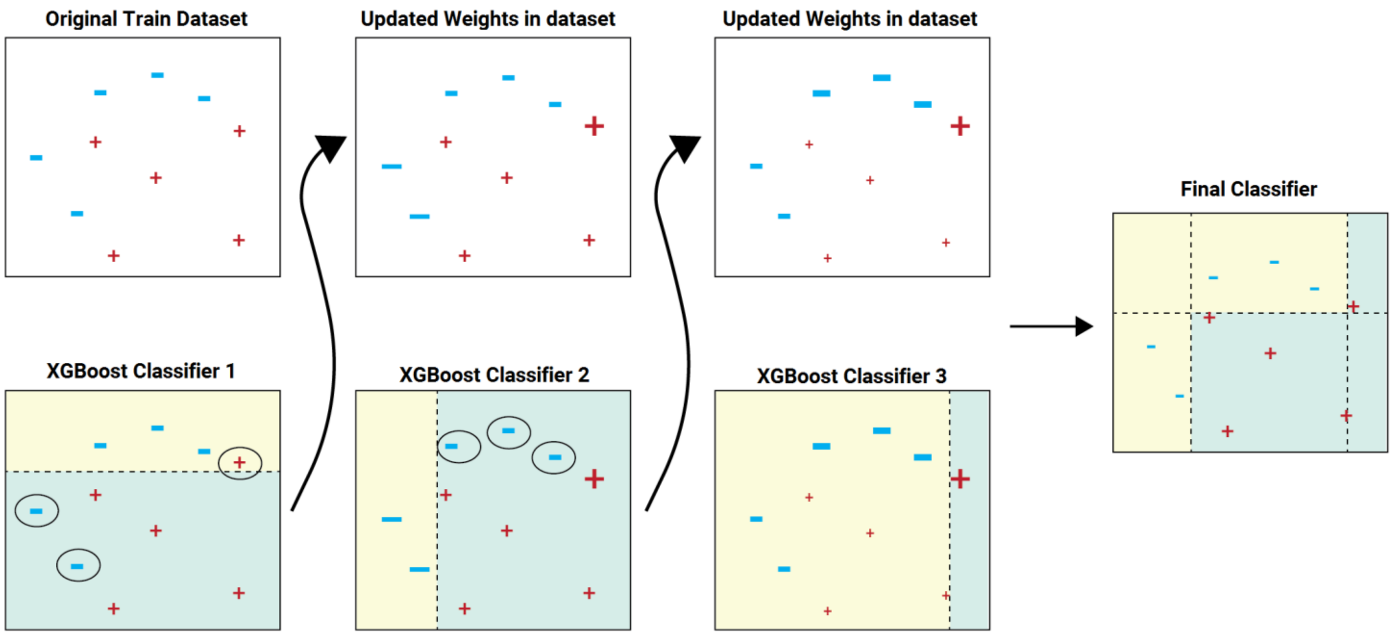


Model Development

- Tested model (cont.)

  2) XGBoost

  - Series of decision trees
    - Interpretability
    - Can capture higher-order interactions and account for complex **nonlinear** relationships between model variables and outcomes
  - Loss function and regularization - noise robustness and less overfitting



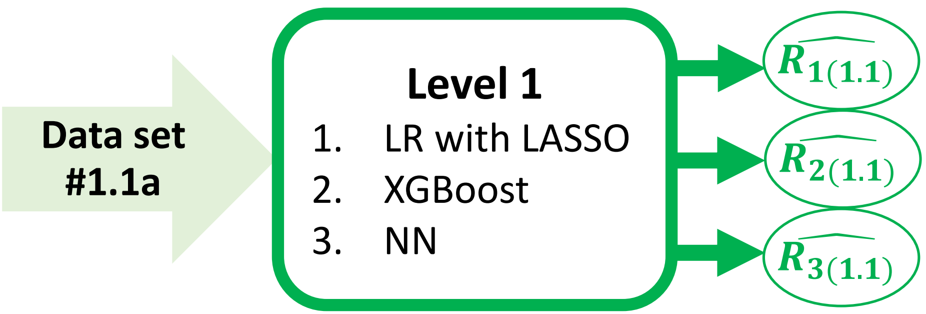
Mahidol University
Wisdom of the Land

Model Development

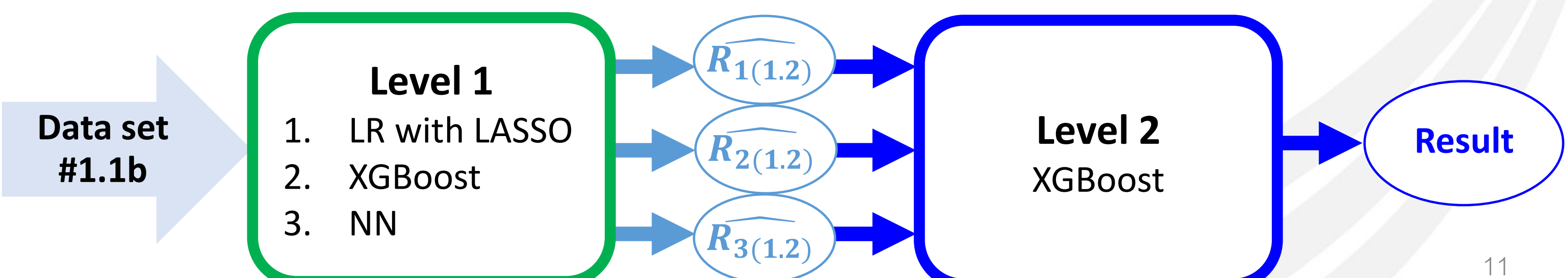
- Tested model (cont.)

3) Meta classifier

Step1 :



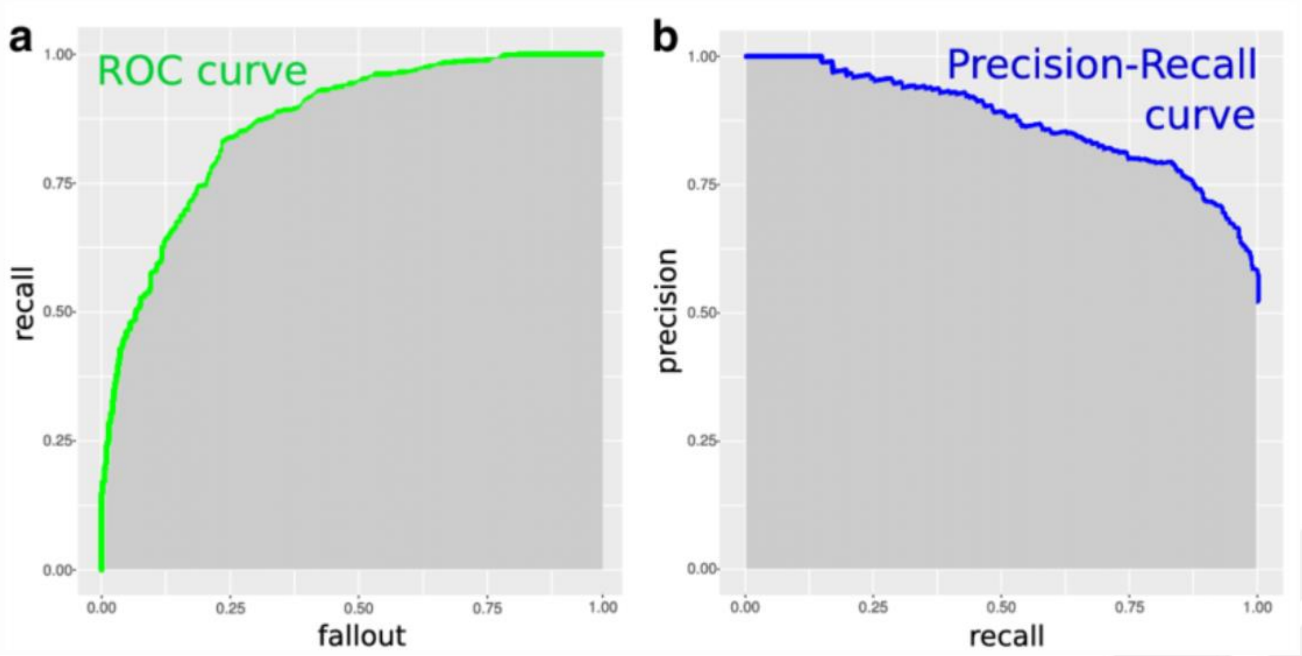
Step2 :



Mahidol University
Wisdom of the Land

Model Evaluation

- F1 score, precision, recall, PPR, NPR
- AU ROC/ C statistics and its 95% CI
- PR curve



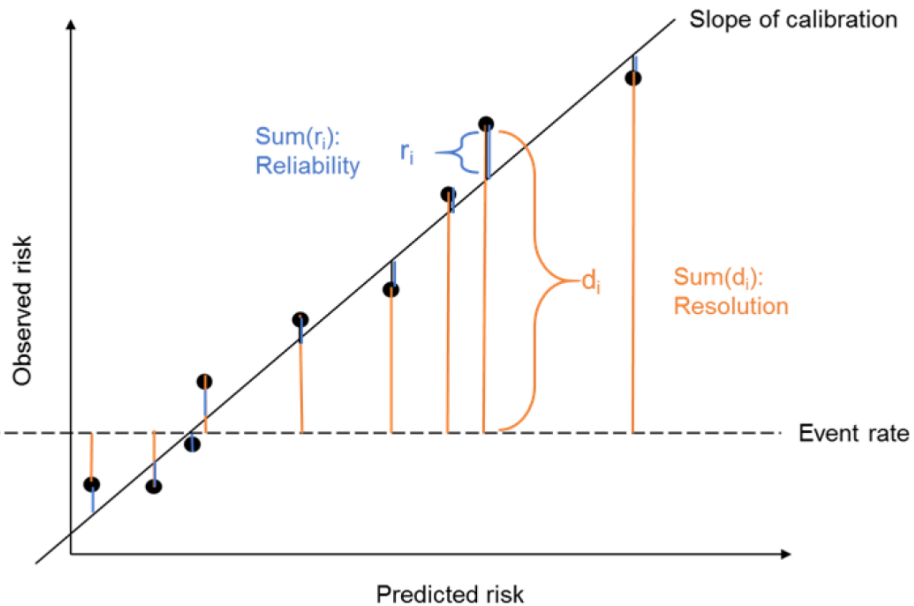
- MSE

Model Calibration

- Calibration slope
- Brier score (3-component decomposition)

$$BS = REL - RES + UNC$$

$$BS = \frac{1}{N}\sum_{k=1}^{K} n_k(\mathbf{f_k} - \bar{\mathbf{o}}_\mathbf{k})^2 - \frac{1}{N}\sum_{k=1}^{K} n_k(\bar{\mathbf{o}}_\mathbf{k} - \bar{\mathbf{o}})^2 + \bar{\mathbf{o}}(1 - \bar{\mathbf{o}})$$



Mahidol University
Wisdom of the Land

Model Calibration

- Shift table
  - Validation cohort (set B: 190,484)
  - Categorize target
    - Low risk: < 1%
    - Moderate risk: 1% - 5%
    - High risk: > 5%
- Sensitivity analysis with 3 thresholds
  - < 1.5%
  - 1.5% - 3%
  - > 3%
- Subgroup analysis

Result – Model Discrimination

- n = 755,402 (derivation + validation cohort)
- Overall mortality rate = 4.4 %

NCDR standard model

- 9 variables
- LR
- AUC ROC = 0.867

Mahidol University
Wisdom of the Land

Result - Model Discrimination

- **F1 score, precision, recall, PPR, NPR** – Limited variable

Baseline

Characteristic	Logistic regression	LASSO	Neural network	XGBoost	Meta-classifier
Variables included in the model of McNamara et al[21]					
Model performance metrics					
AUROC (95% CI)	0.878 (0.875-0.881)	0.874 (0.870-0.879)	0.874 (0.870-0.878)	0.886 (0.882-0.890)	0.886 (0.882-0.890)
Precision-recall AUC	0.372	0.367	0.371	0.395	0.398
F score	0.415	0.408	0.411	0.432	0.432
Sensitivity	0.42 (0.41-0.43)	0.43 (0.42-0.45)	0.41 (0.40-0.42)	0.44 (0.43-0.45)	0.43 (0.42-0.44)
Specificity	0.97 (0.97-0.97)	0.97 (0.97-0.97)	0.97 (0.97-0.97)	0.97 (0.97-0.97)	0.98 (0.97-0.98)
PPV	0.41 (0.40-0.42)	0.38 (0.37-0.39)	0.41 (0.40-0.42)	0.42 (0.41-0.43)	0.44 (0.43-0.45)
NPV	0.97 (0.97-0.97)	0.97 (0.97-0.98)	0.97 (0.97-0.97)	0.98 (0.97-0.98)	0.97 (0.97-0.98)

Data set #1.2 = 190,484

Result – Model Discrimination

- **F1 score, precision, recall, PPR, NPR** – Expanded variable set

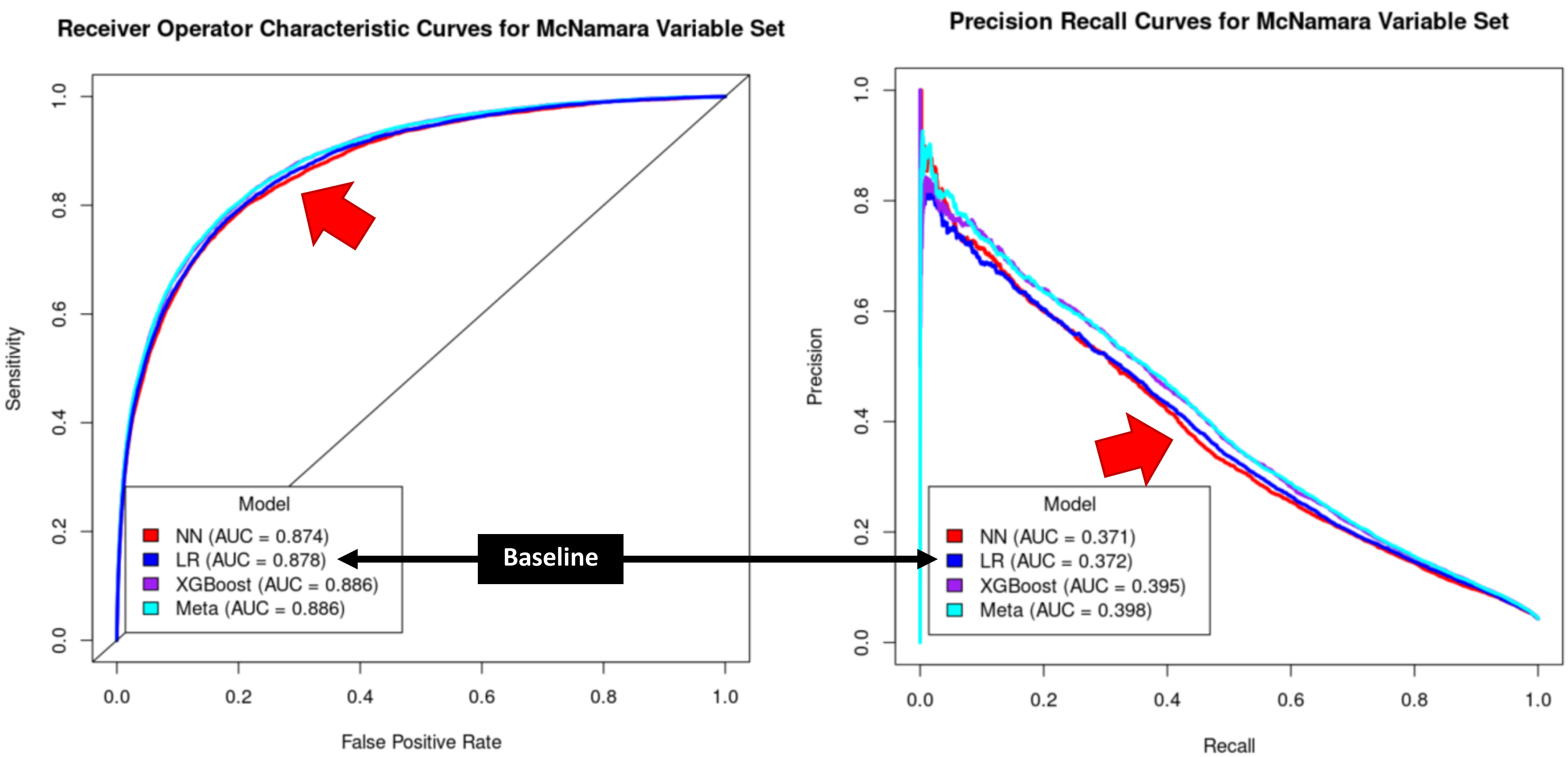
Baseline

Characteristic	Logistic regression	LASSO	Neural network	XGBoost	Meta-classifier
Expanded variables included from the CP-MI Registry					
Model performance metrics					
AUROC (95% CI)	0.888 (0.884-0.892)	0.886 (0.882-0.890)	0.885 (0.881-0.889)	0.898 (0.894-0.902)	0.899 (0.895-0.903)
Precision-recall AUC	0.421	0.415	0.406	0.451	0.453
F score	0.436	0.436	0.428	0.458	0.459
Sensitivity	0.47 (0.45-0.48)	0.42 (0.41-0.43)	0.43 (0.42-0.44)	0.45 (0.44-0.47)	0.43 (0.42-0.44)
Specificity	0.97 (0.97-0.97)	0.98 (0.98-0.98)	0.97 (0.97-0.98)	0.98 (0.98-0.98)	0.98 (0.98-0.98)
PPV	0.41 (0.40-0.42)	0.45 (0.44-0.46)	0.43 (0.42-0.44)	0.46 (0.45-0.47)	0.49 (0.48-0.50)
NPV	0.98 (0.98-0.98)	0.97 (0.97-0.98)	0.97 (0.97-0.98)	0.98 (0.98-0.98)	0.97 (0.97-0.98)

Data set #1.2 = 190,484

Result - Model Discrimination

- **ROC** – Limited variable

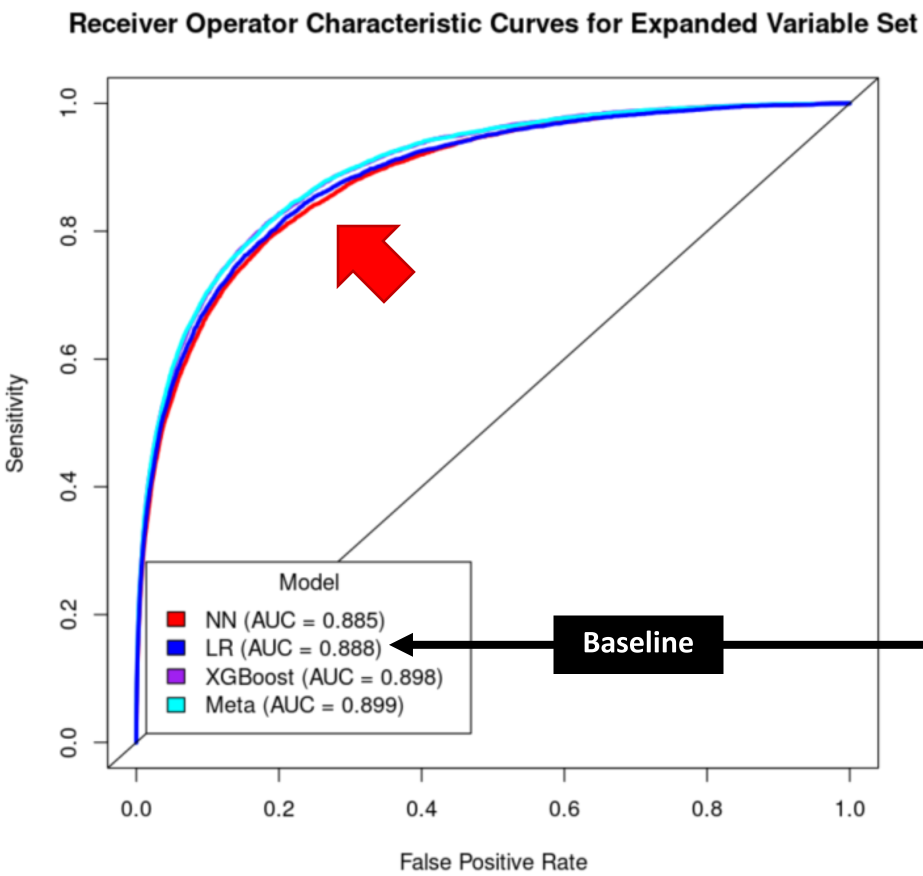


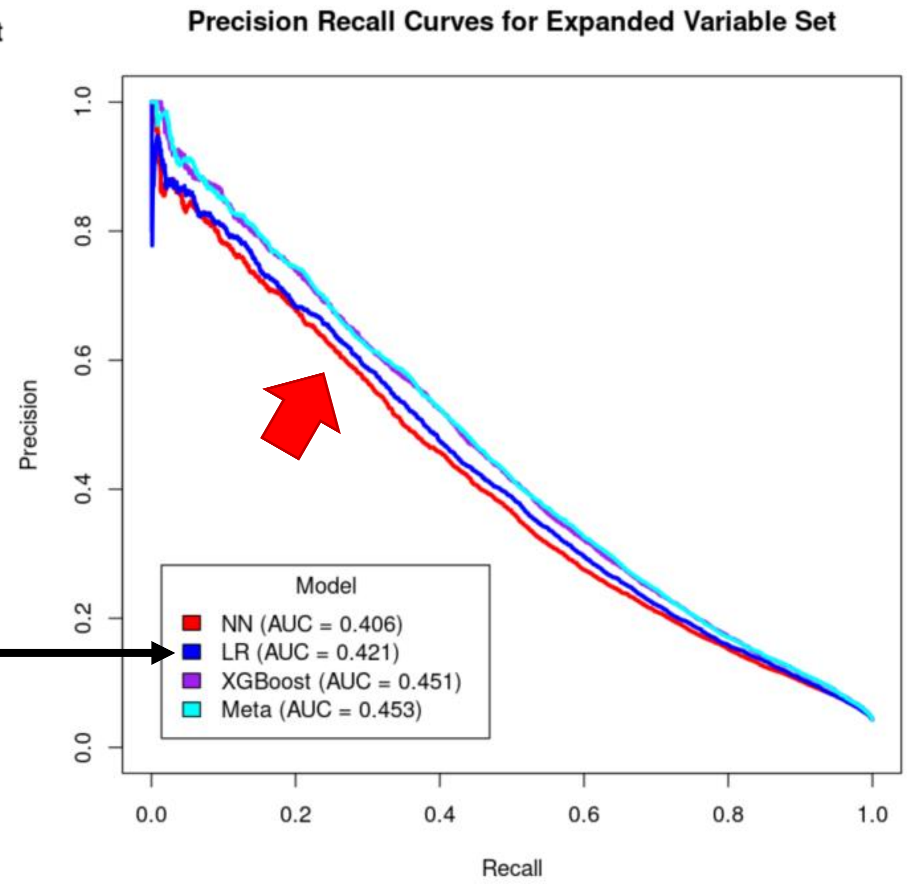
Data set #1.2 = 190,484

Mahidol University
Wisdom of the Land

Result - Model Discrimination

- **ROC** – Expanded variable set

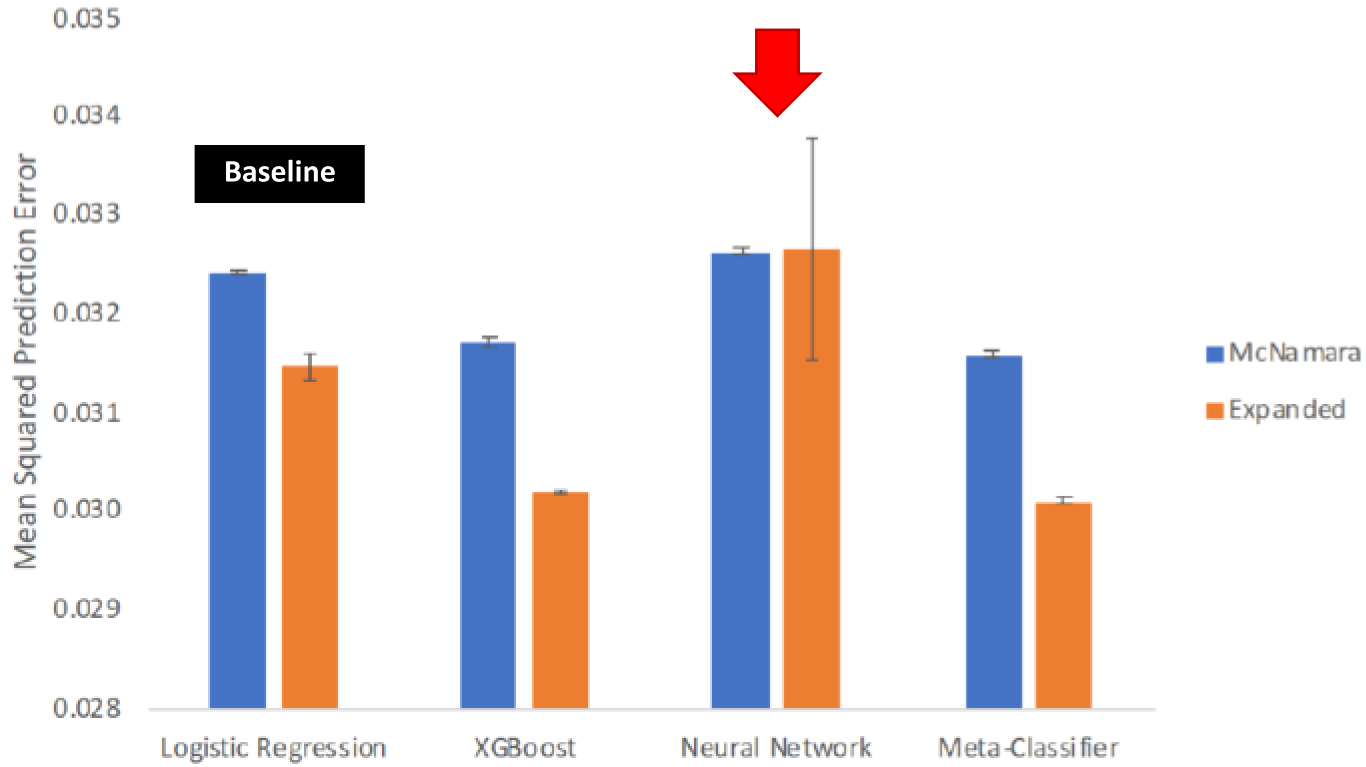




Data set #1.2 = 190,484

Result – Model Discrimination

- **MSE**

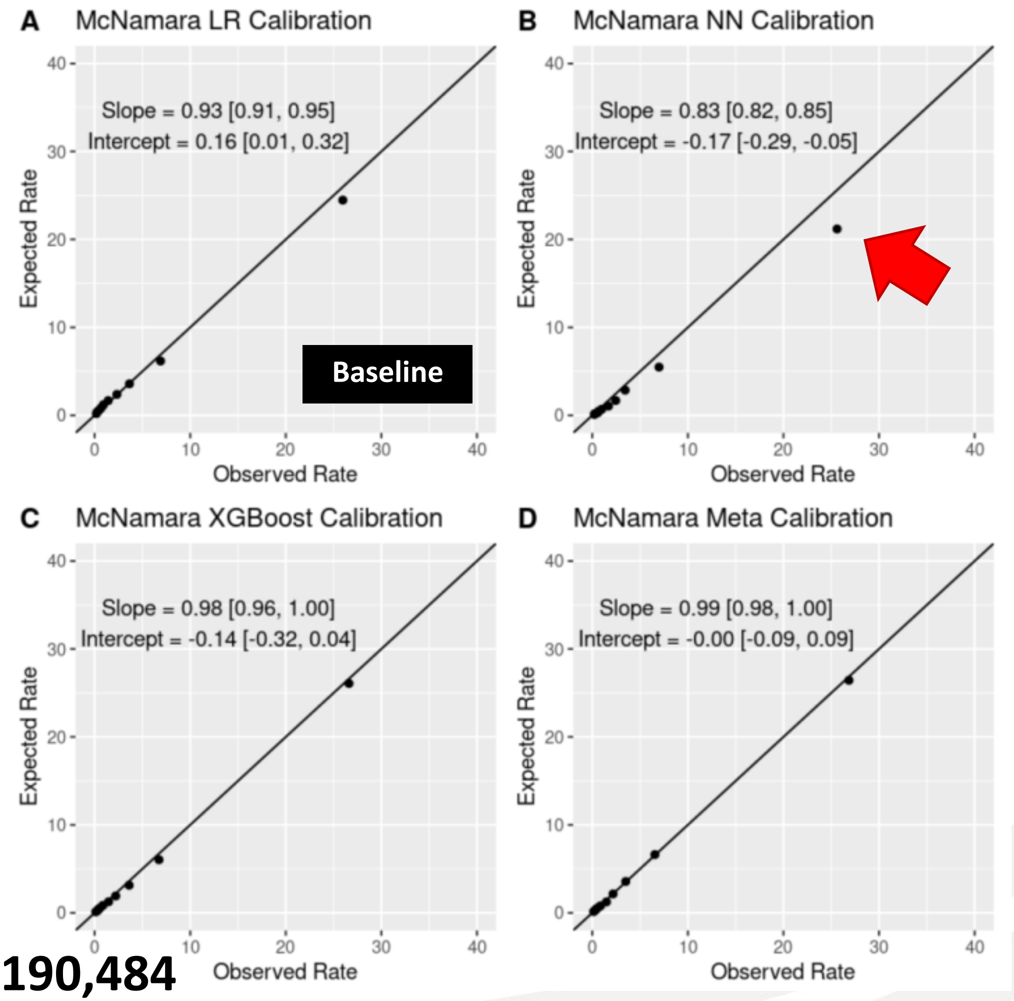


Data set #1.2 = 190,484

Mahidol University
Wisdom of the Land

Result - Model Calibration

- **Calibration slope** – Limited variable

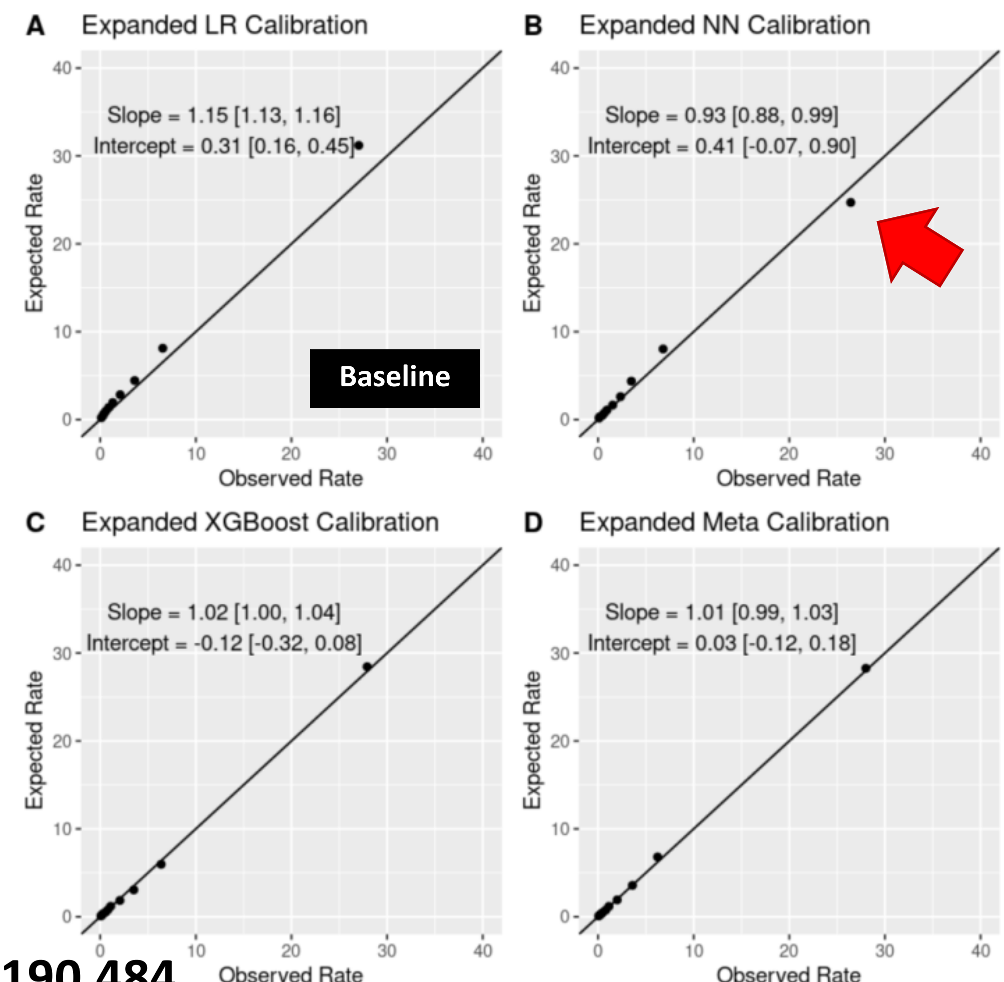


Data set #1.2 = 190,484

Mahidol University
Wisdom of the Land

Result - Model Calibration

- **Calibration slope** – Expanded variable set



Data set #1.2 = 190,484

Mahidol University
Wisdom of the Land

Result - Model Calibration

- **Brier score** – Limited variable

Baseline

Characteristic	Logistic regression	LASSO	Neural network	XGBoost	Meta-classifier
Variables included in the model of McNamara et al[21]					
Model performance metrics					
Brier score					
Reliability, mean (SD), $\times 10^{-6}$	28.4 (9.2)	96.3 (16.5)	224.0 (26.1)	9.5 (3.8)	2.3 (2.1)
Resolution, mean (SD), $\times 10^{-3}$	5.6 (0.1)	5.5 (0.1)	5.4 (0.1)	5.8 (0.1)	5.9 (0.1)
Uncertainty	0.04	0.04	0.04	0.04	0.04
Overall, $\times 10^{-2}$	3.52	3.54	3.56	3.49	3.48

Data set #1.2 = 190,484

Mahidol University
Wisdom of the Land

Result - Model Calibration

- **Brier score** – Expanded variable set

Baseline

Characteristic	Logistic regression	LASSO	Neural network	XGBoost	Meta-classifier
Variables included in the model of McNamara et al[21]					
Model performance metrics					
Brier score					
Reliability, mean (SD), $\times 10^{-6}$	229.4 (25.6)	40.6 (10.3)	55.7 (11.2)	6.5 (3.5)	4.3 (2.6)
Resolution, mean (SD), $\times 10^{-3}$	6.0 (0.1)	5.9 (0.1)	5.8 (0.1)	6.4 (0.2)	6.5 (0.2)
Uncertainty	0.04	0.04	0.04	0.04	0.04
Overall, $\times 10^{-2}$	3.50	3.49	3.50	3.43	3.42

Data set #1.2 = 190,484

Result – Model Calibration

- **Shift table [expanded, better than limited]**

Table 3. Performance of the XGBoost and Meta-Classifier Models Compared With Logistic Regression[a]

Model	Expanded LR, No. of patients (% observed mortality)			
	<1%	1%-5%	>5%	All
XGBoost vs LR				
Expanded XGBoost				
<1%	65 193 (0.27)	31 971 (0.65)	422 (1.18)	97 586 (0.40)
1%-5%	3384 (0.95)	44 486 (2.21)	13 155 (3.91)	61 025 (2.51)
>5%	68 (2.94)	2899 (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)
Meta-classifier vs LR				
Expanded meta-classifier				
<1%	65 694 (0.27)	30 661 (0.65)	175 (0.00)	96 530 (0.39)
1%-5%	2930 (1.06)	45 726 (2.17)	9033 (3.55)	57 689 (2.33)
>5%	21 (0.00)	2969 (6.03)	33 275 (18.66)	36 265 (17.61)
All	68 645 (0.30)	79 356 (1.73)	42 483 (15.37)	190 484 (4.26)

XGB

Baseline

Meta-classifier

Baseline

Data set #1.2 = 190,484

Mahidol University
Wisdom of the Land

Result - Model Calibration

- **Sensitivity analysis**

Data set #2 = 946,597 (multiple imputation)

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]

Baseline

Mahidol University
Wisdom of the Land

Result – Model Calibration

- Subgroup analysis

Baseline

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]

Mahidol University
Wisdom of the Land

Conclusion and limitation

- None of the tested ML models were **substantive** improvement in the discrimination of in-hospital mortality after AMI
- **XGB** and **meta-classifier models** improved accuracy of risk for high-risk patients (compared with LR)
- Better clarify the individual risk for adverse outcomes

- Relevant information, such as duration of comorbidities was not captured
- Certain prognostic characteristics of the patients' general health are not included