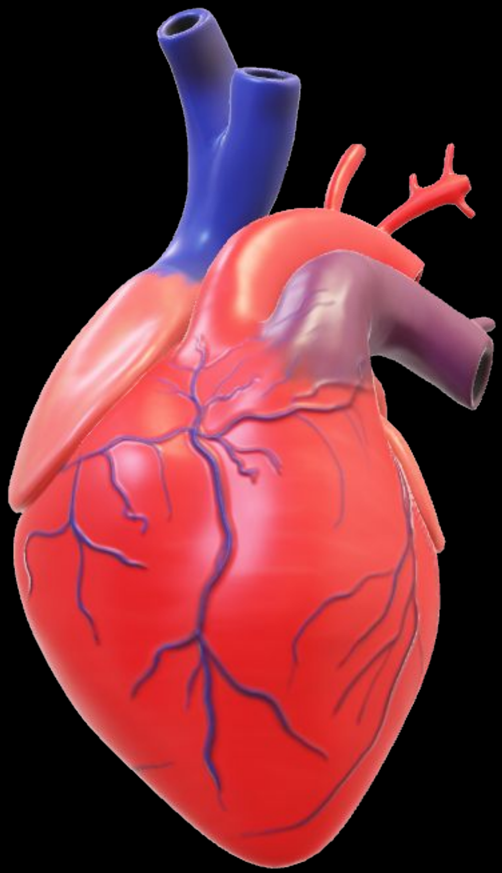
Predicting Atrial Fibrillation

Bella Falkenberg
Ava Klissouras
Manoj Kambara



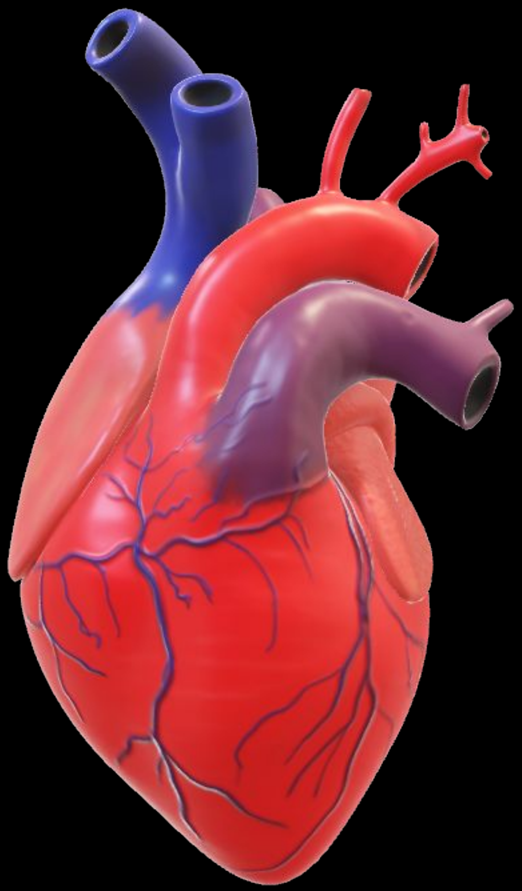
Background

Myocardial Infarction

Muscle of the heart fails to receive enough blood

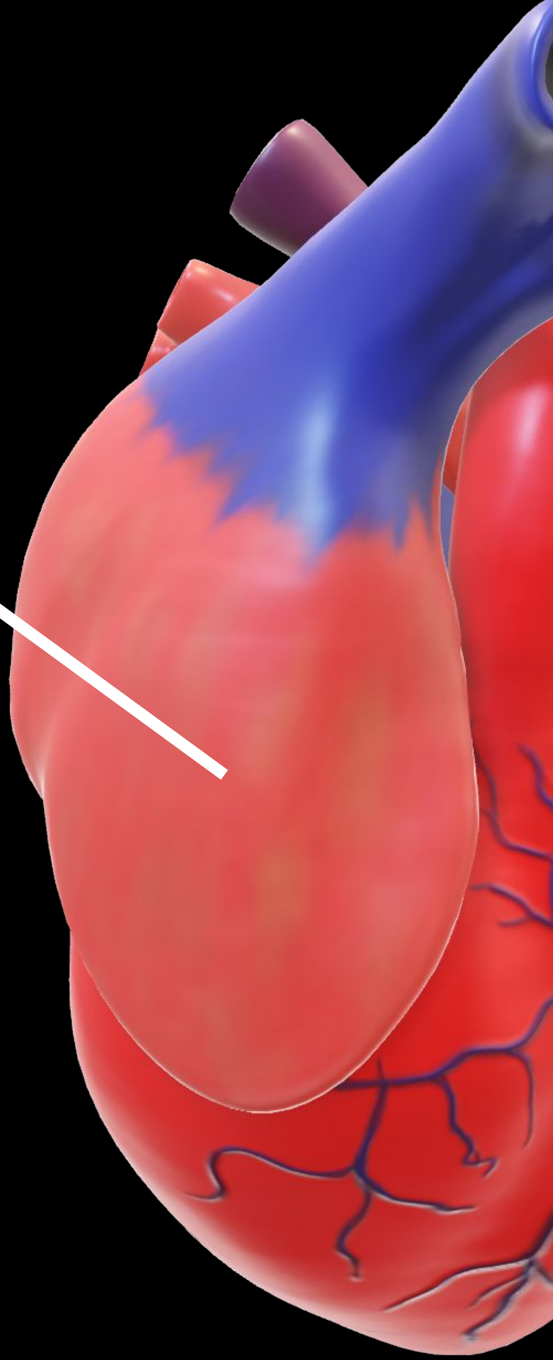
Major Complications of an MI:

- Atrial Fibrillation
- Pulmonary Edema
- Relapse of MI



Atrial Fibrillation

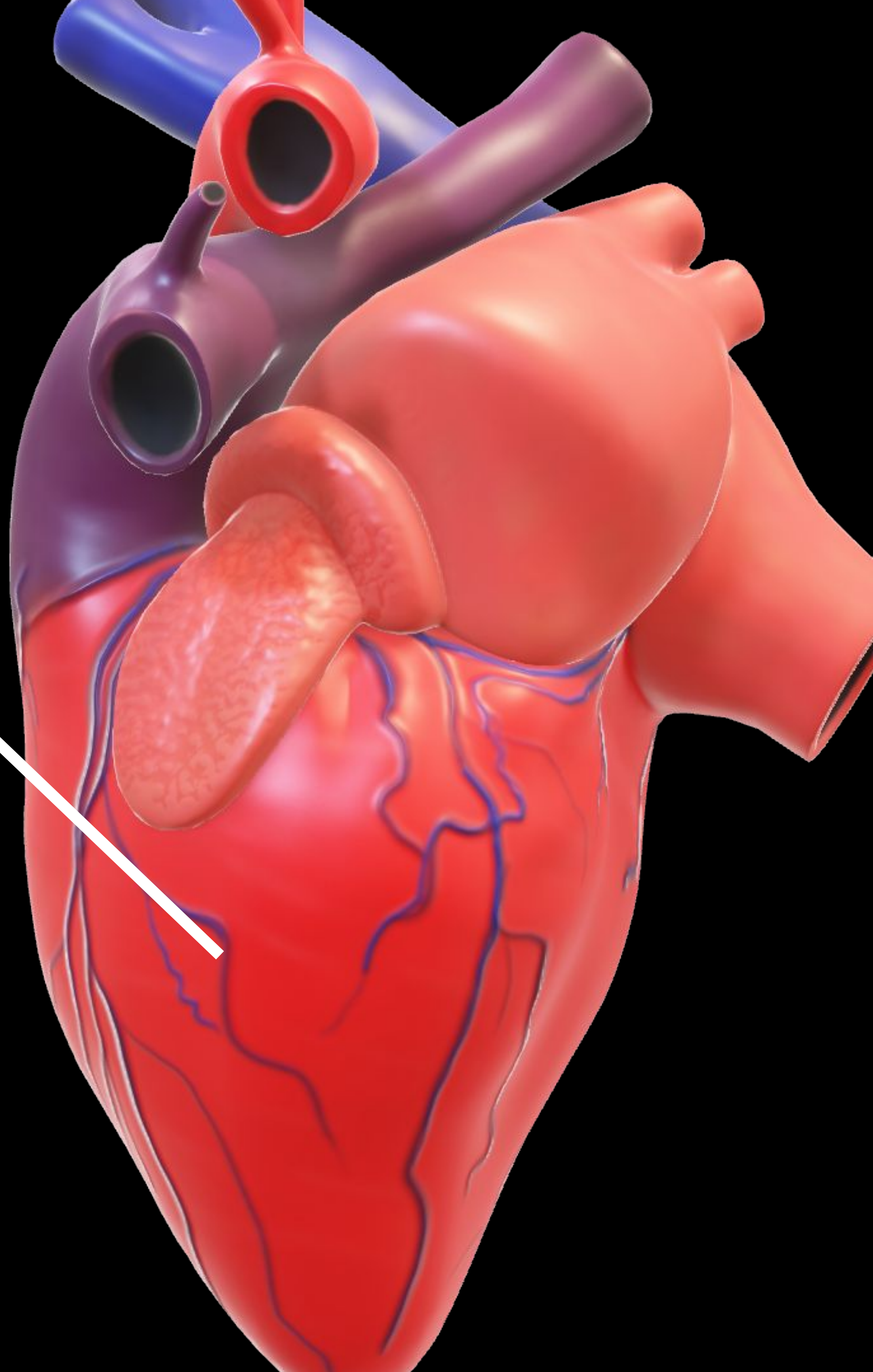
Signals originate from the SA node, located in the right atria of the heart



Irregular contractions of the atria caused by imbalanced signals

Atrial Fibrillation

This causes mistimed contractions of the left ventricle, which pumps blood into the aorta

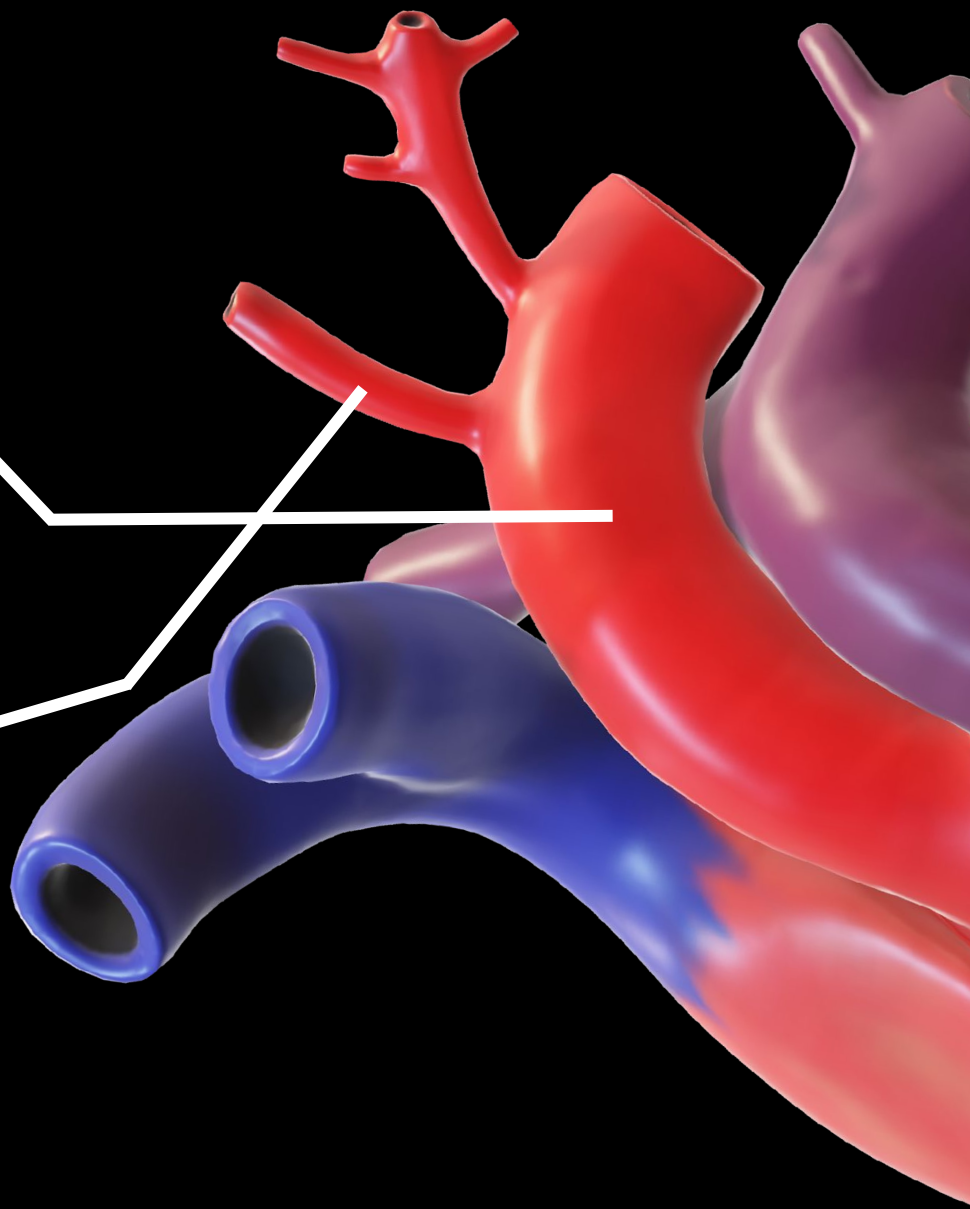


Atrial Fibrillation

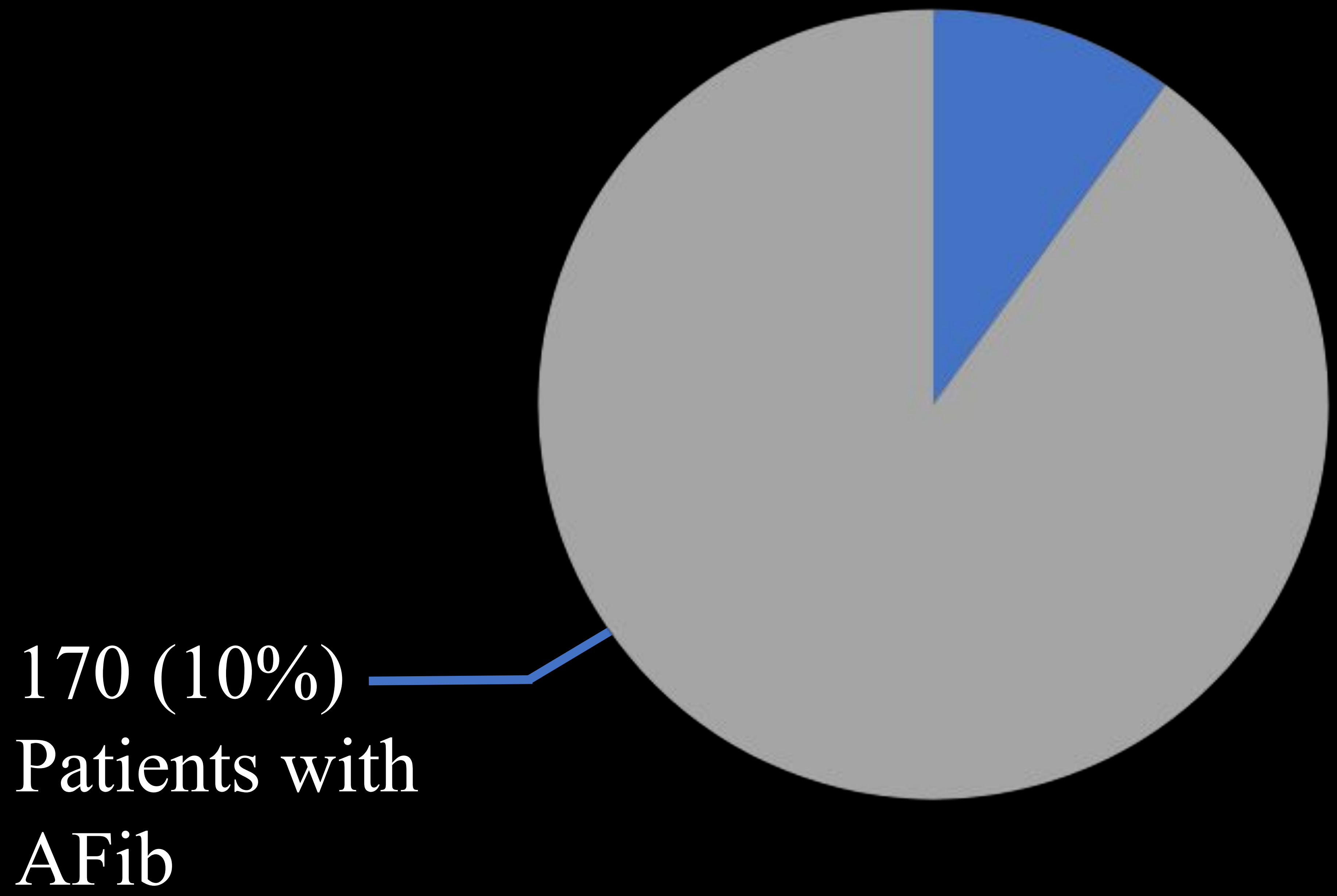
AFib can lead to blocks in the aorta

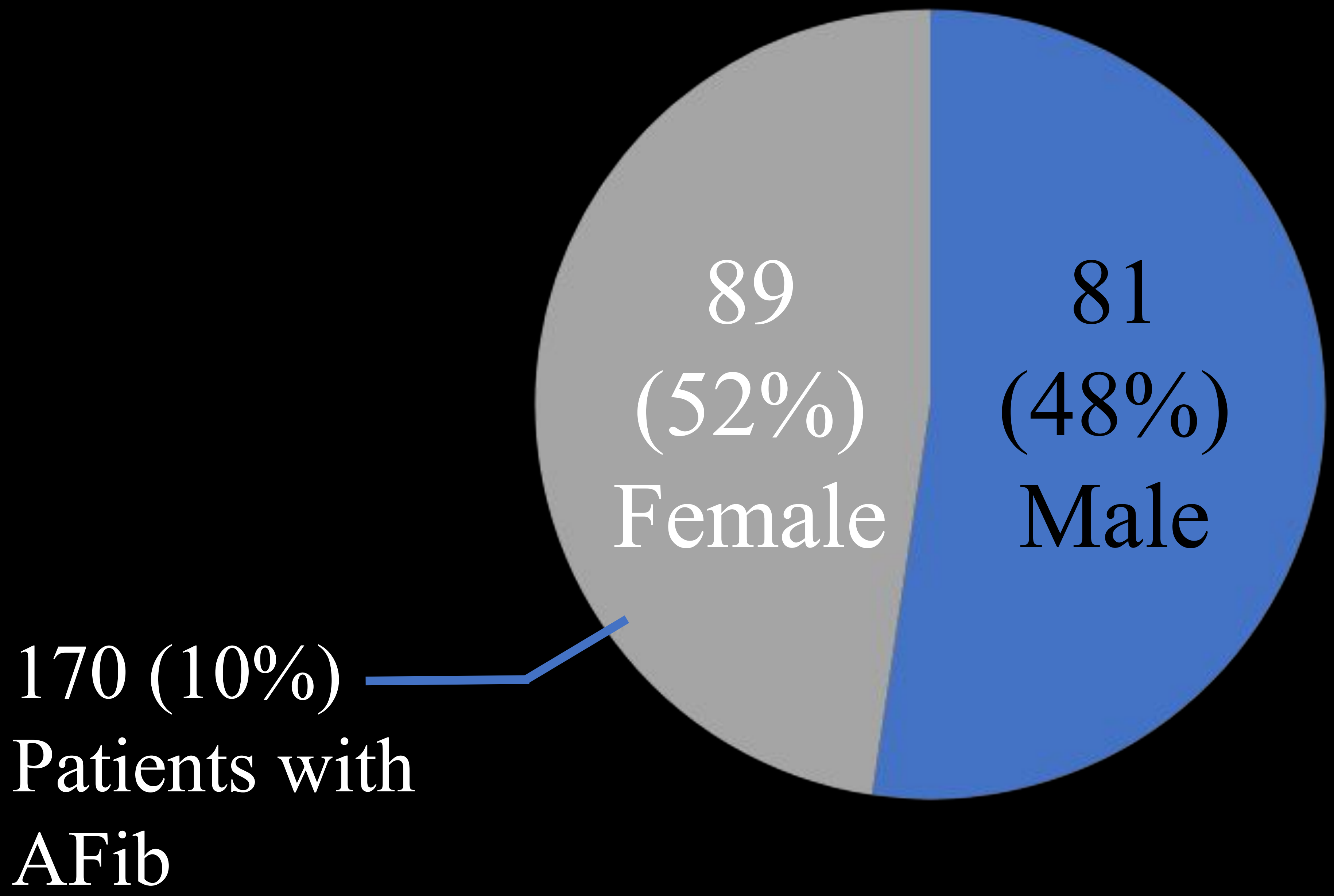
Blocks in the aorta can keep blood from the brain, causing a stroke[1]

AFib can lead to heart failure[1]

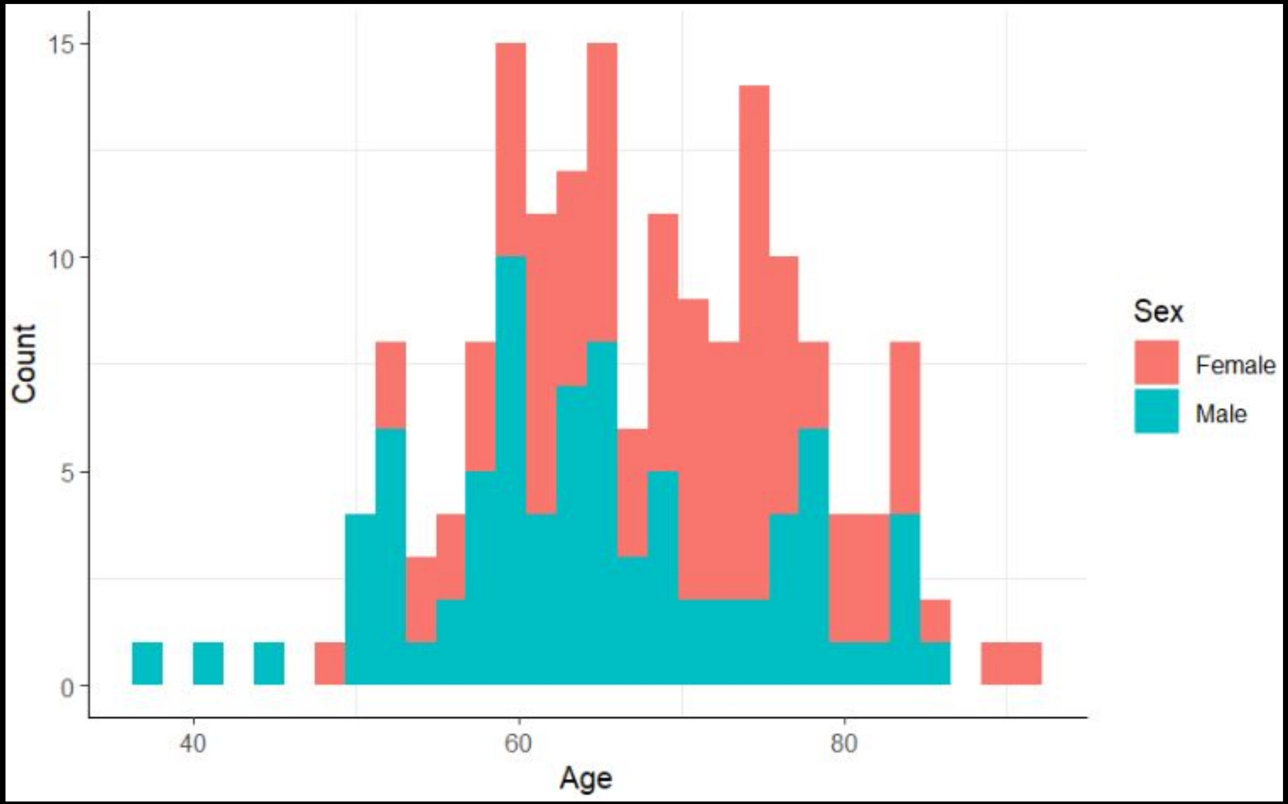


1700
Patients


170 (10%)
Patients with
AFib


89
(52%)
Female
81
(48%)
Male
170 (10%)
Patients with
AFib

Mean: 67.124

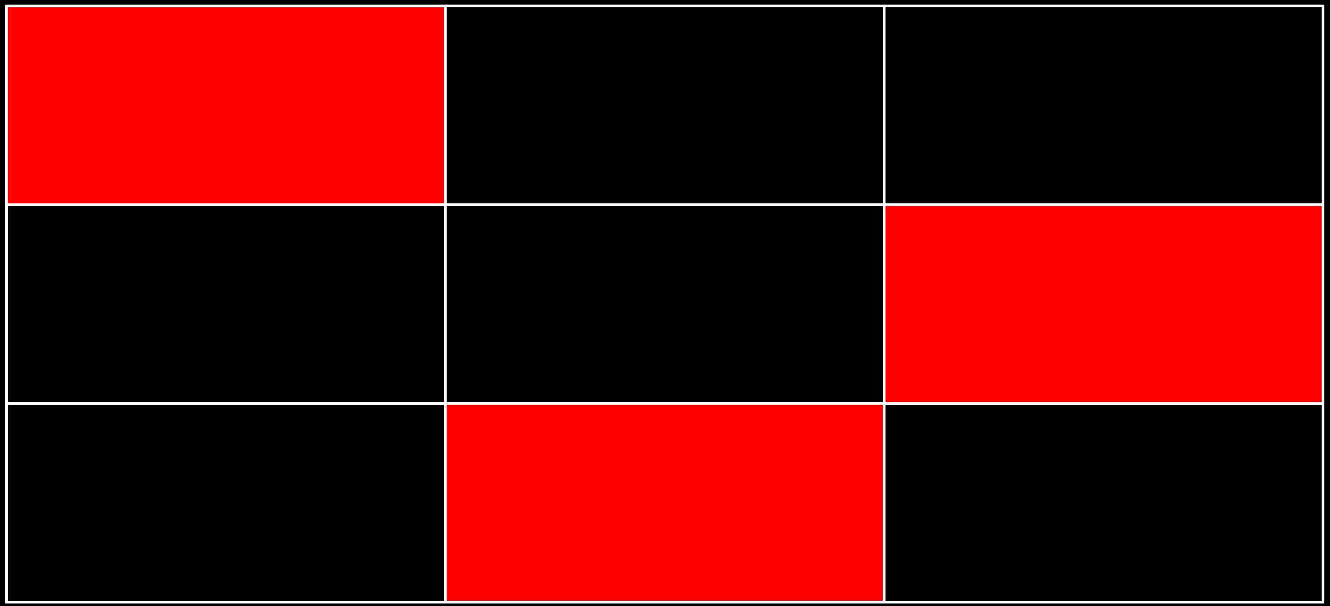


Distribution of ages of patients that developed AFib as a complication of their MI.

Aim

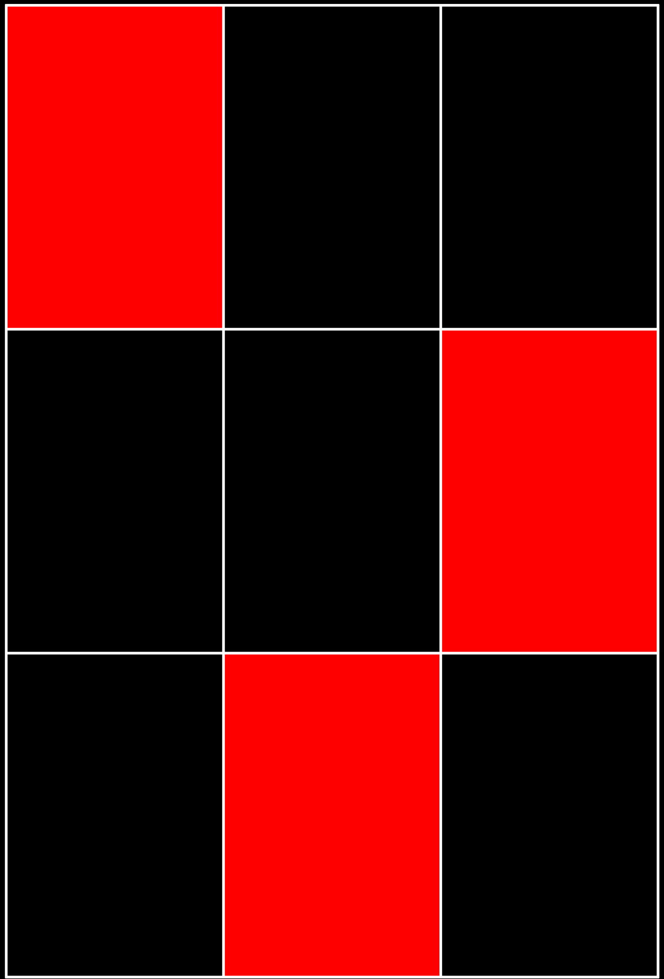
Create models to predict probability of AFib as a complication following treatment for an MI.

1. Find and apply a method to clean the data
2. Use R methods to make a model
3. Evaluate the models for calibration and discrimination



Missing Data Imputation
Multiple Imputation by Chained Equations
(MICE)

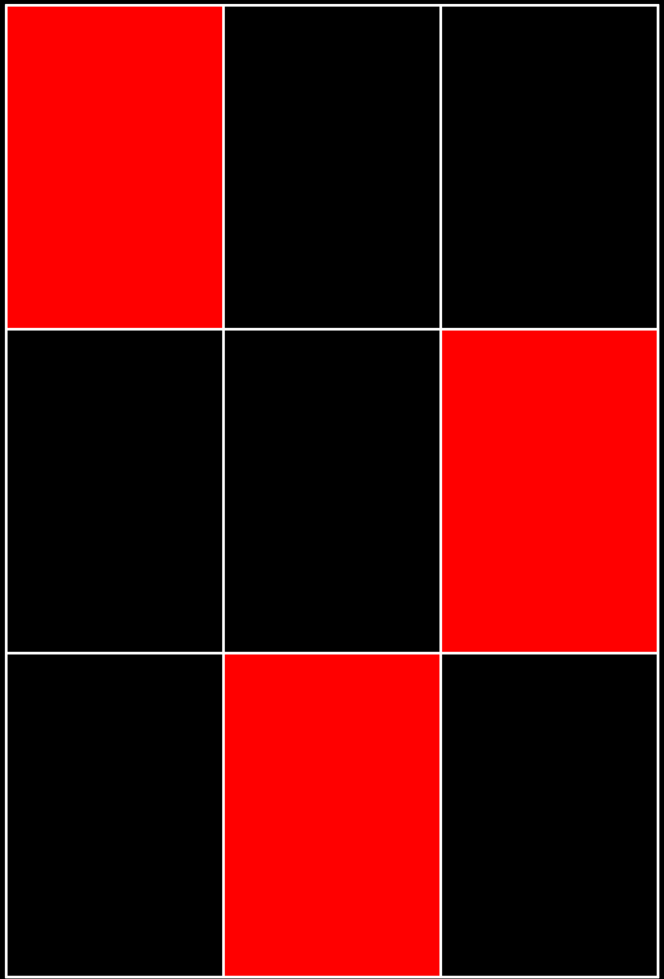
MICE Imputation[2]: Data Preparation



Removed columns in which greater than 50% of values were missing

- KFK_Blood
  - Serum PCK content
- IBS_NASL
  - Heredity on CHD
- S_AD_KBRIG/D_AD_KBRIG
  - Systolic/Diastolic by emergency cardiology team

MICE Imputation[2]: Data Preparation



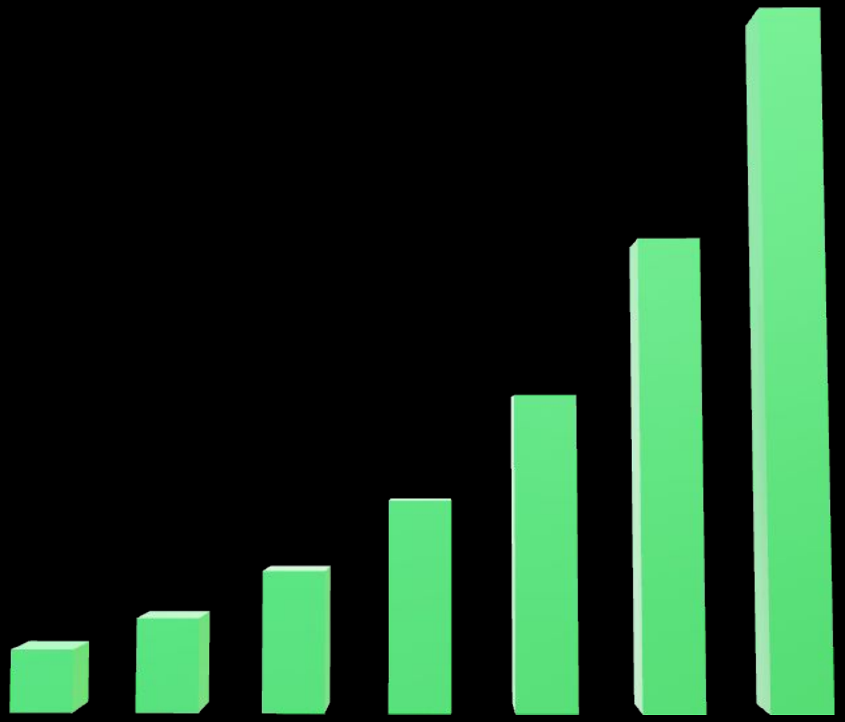
Removed a highly correlated column

- S_AD_ORIT
  - Systolic BP by ICU
- Used `findCorrelation()`
  - Used a cutoff of 0.8
- Assumed that data is Missing at Random (MAR)
- Applied the function `mice()` to the data set

MICE Imputation[2]: Data Preparation

Removed a highly correlated column

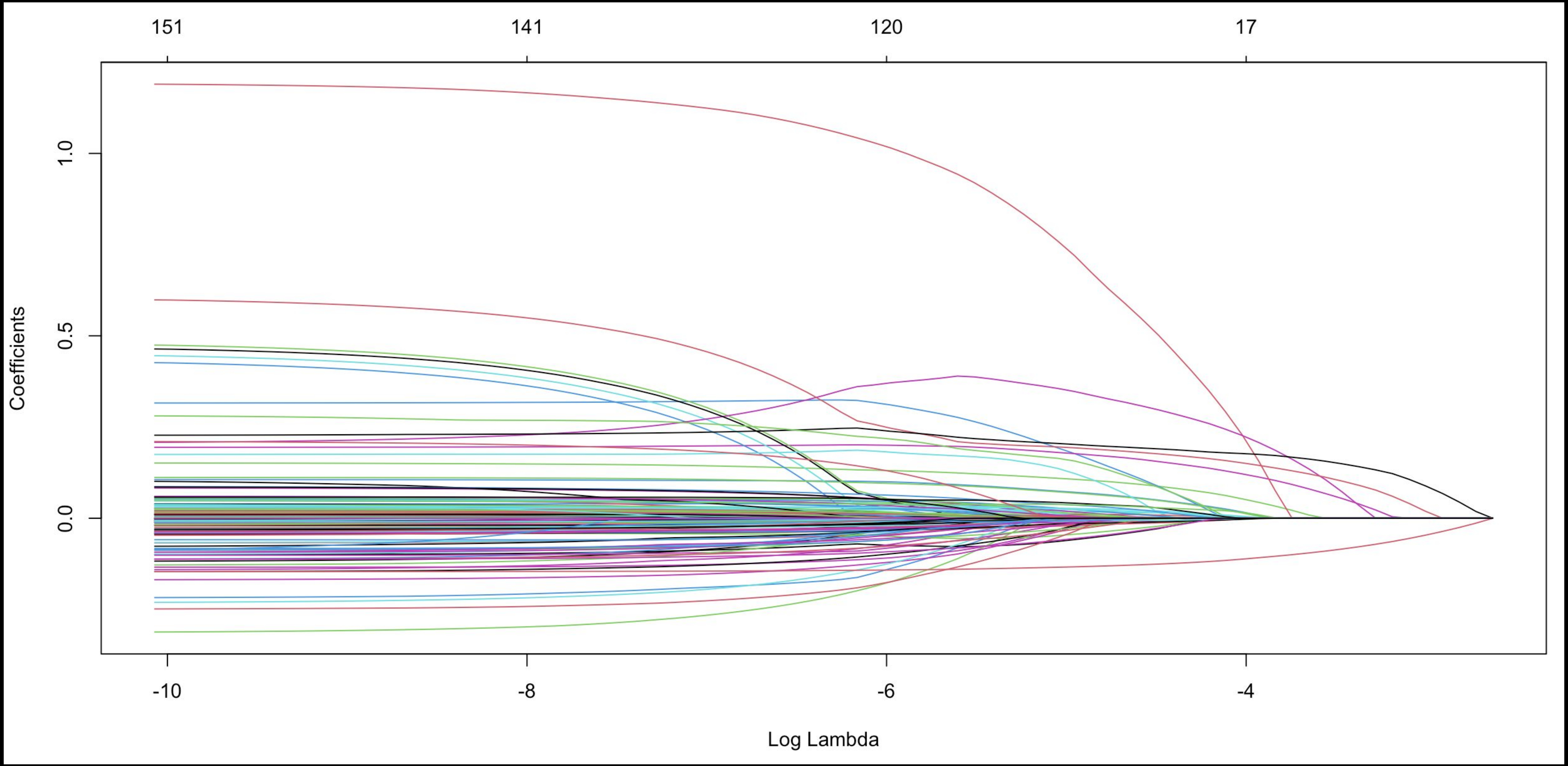
- S_AD_ORIT
  - Systolic BP by ICU
- Used `findCorrelation()`
  - Used a cutoff of 0.8
- Assumed that data is Missing at Random (MAR)
- Applied the function `mice()` to the data set



Modeling

LASSO Regression

Introduction to LASSO

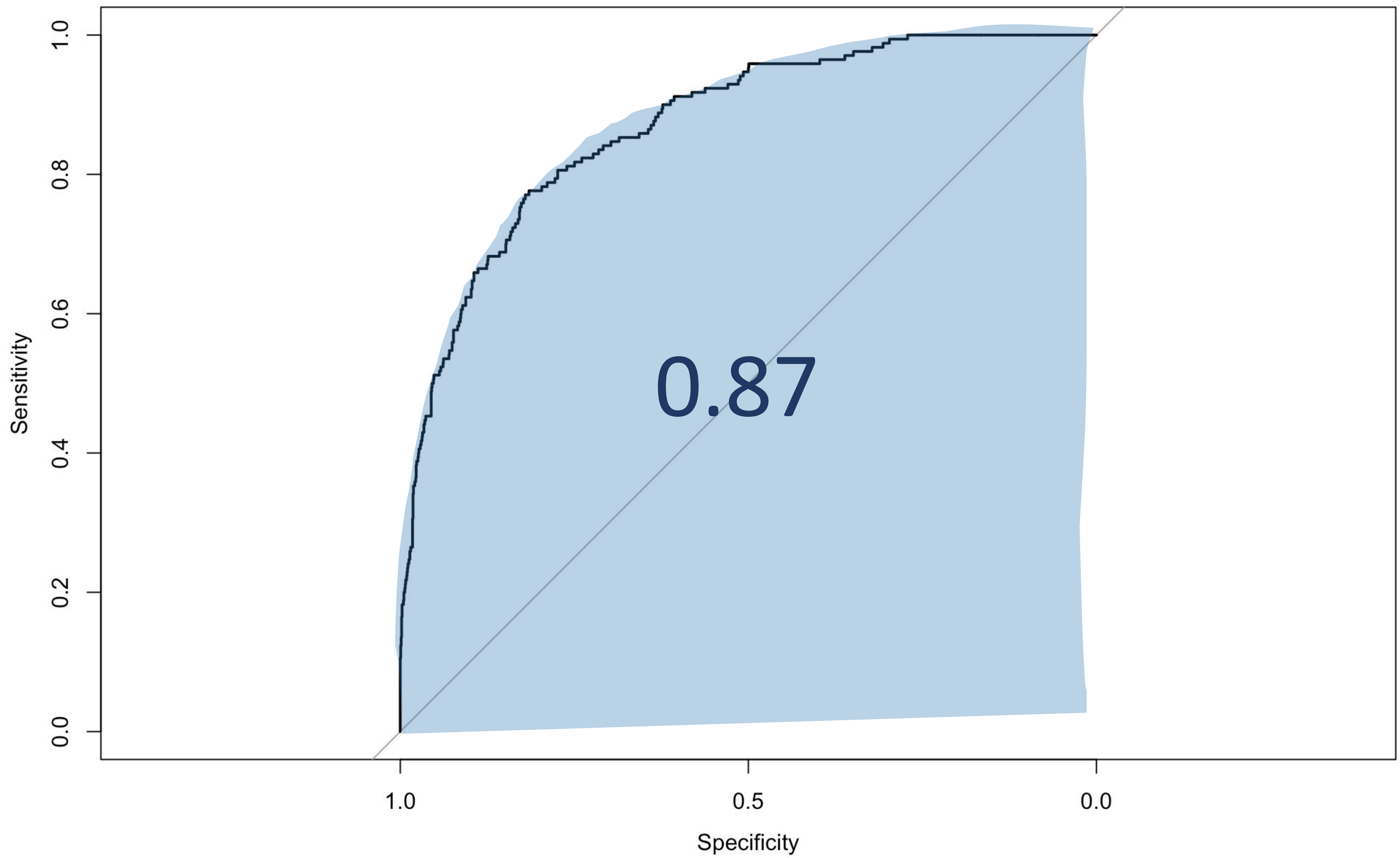


Our LASSO Regression

- Categorical variables --> dummy variables
- p-value = 6.59E-9
- 50 significant covariates

Table 1. Significant covariates for LASSO regression

Covariate	Log Odds	Probability	p-value
Age (AGE)	0.037	0.51	0.003
Afib in anamnesis (nr03)	1.73	0.85	<0.001
ECG rhythm @ hospital (bradycardia) (ritm_ecg_p_08)	2.33	0.91	<0.001
ECG rhythm @ hospital (Afib) (n_r_ecg_p_05)	1.82	0.86	<0.001
NSAID use by ECT (NOT_NA_KB)	-0.41	0.12	<0.001



0.87
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
Specificity
1.0
0.5
0.0

Evaluating LASSO

- Calibration plot
- Discrimination plot
- AUC of ROC plot
- Confusion Matrix
  - Accuracy = 0.8112
  - Sensitivity = 0.9704
  - Specificity = 0.3181

		Actual	
		AFib Complication	No AFib Complication
Predicted	AFib Complication	1247	283
	No AFib Complication	38	132

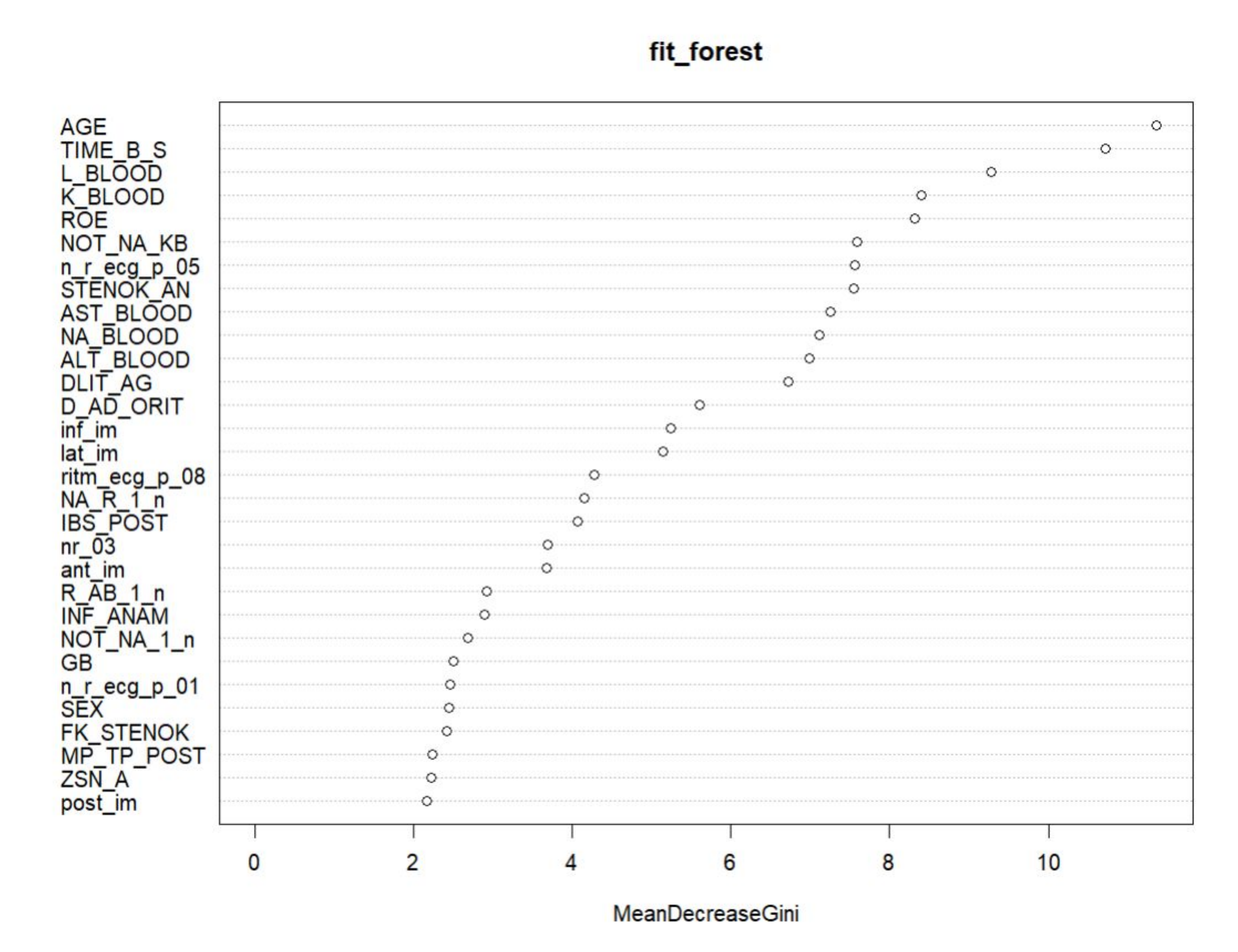
Random Forest

Random Forest

- Supervised machine learning model
- Combines the output of multiple decision trees in order to make a single result
- Can perform both regression and classification
- Random forests can be greedy

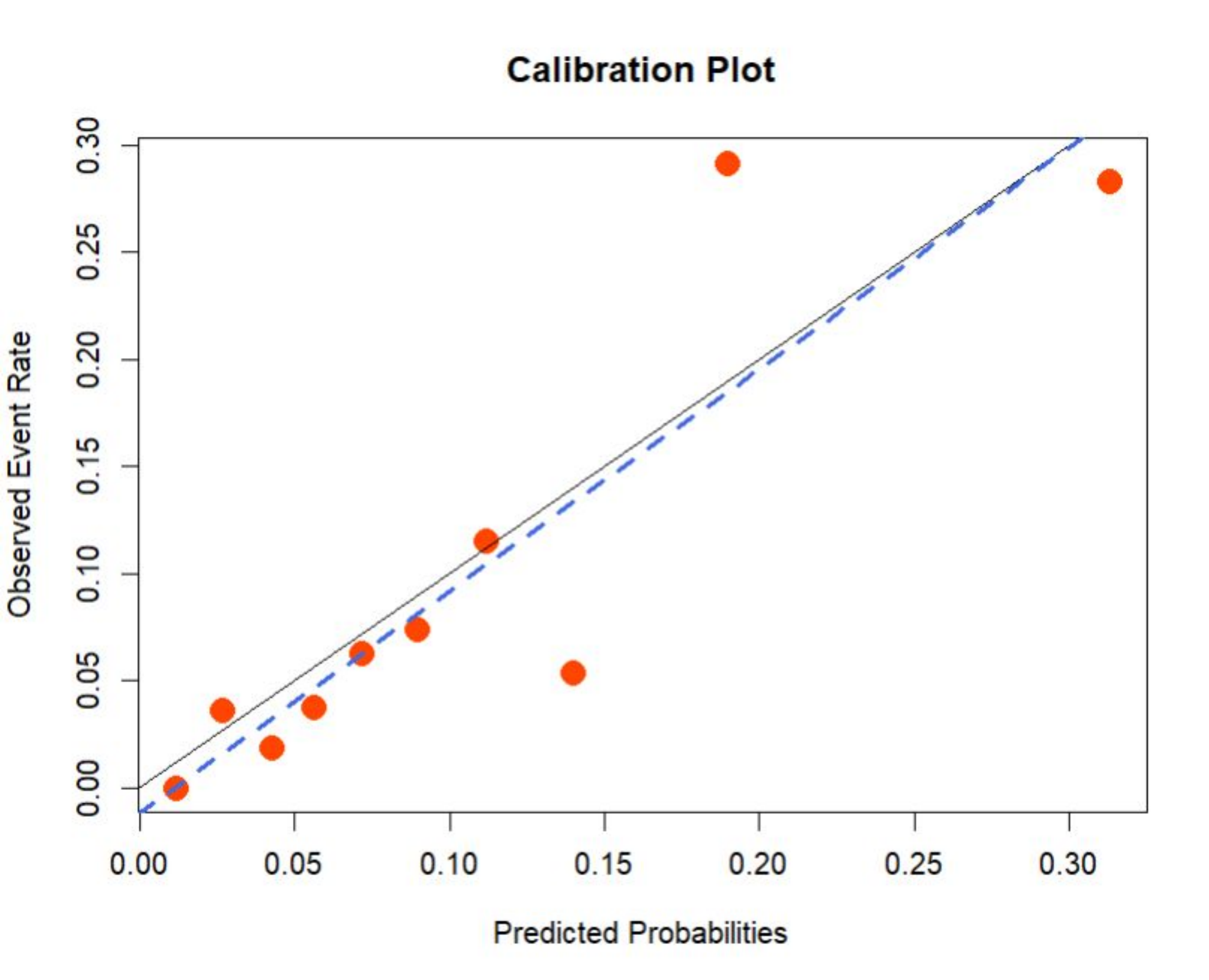
Our Random Forest Model

- We split the data into a training set (70%) and a testing set (30%)
  - Most significant:
    - Age
    - Time to hospital
    - Blood content



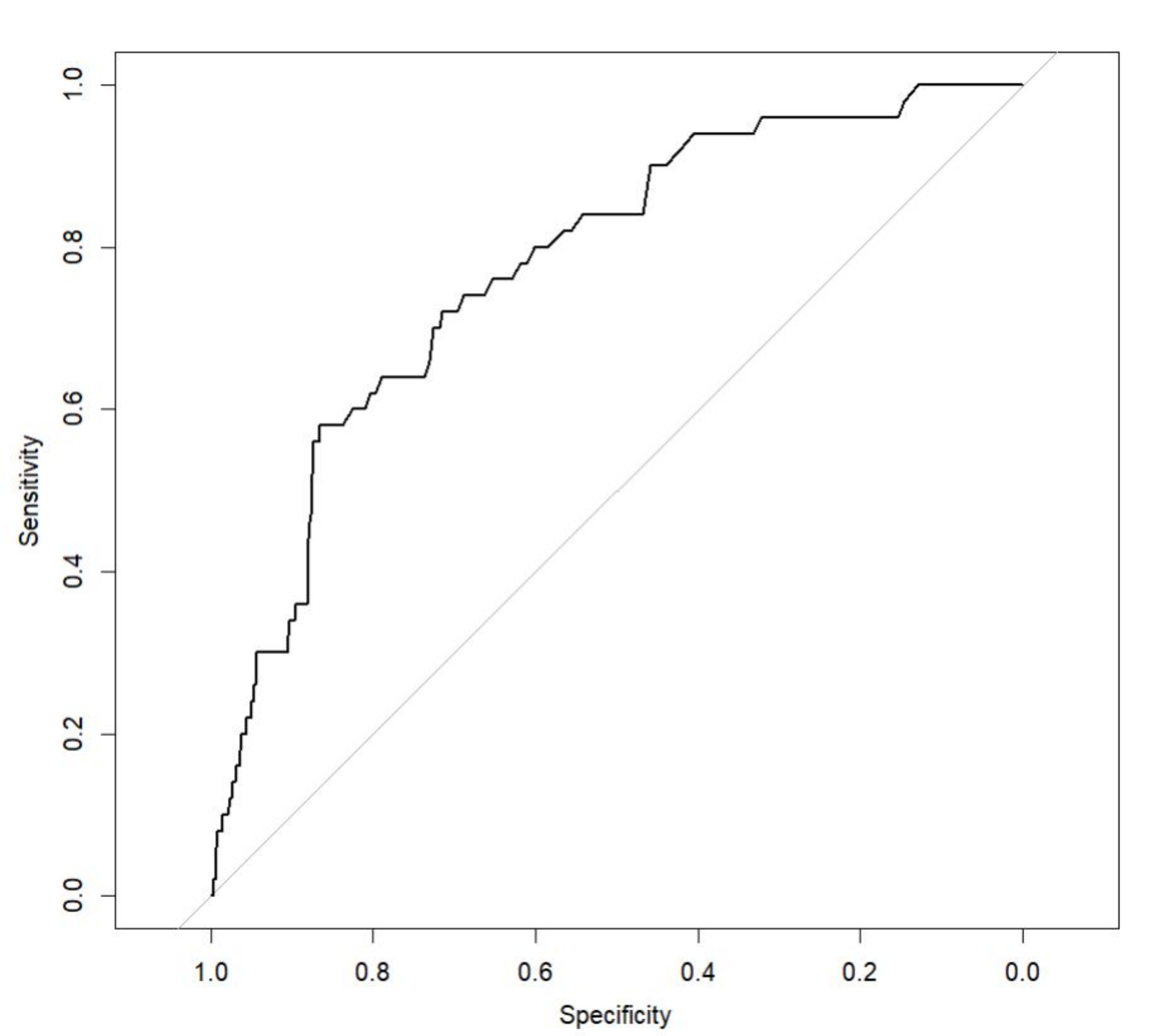
Our Random Forest Model

- We split the data into a training set (70%) and a testing set (30%)
- Calibration Plot
  - Slope
    - 1.04 [0.62,1.45]
  - Y-intercept
    - -0.01 [-0.07,0.05]



Our Random Forest Model

- We split the data into a training set (70%) and a testing set (30%)
- Calibration Plot
- ROC Curve
  - AUC = 0.78



Our Random Forest Model

- We split the data into a training set (70%) and a testing set (30%)
- Calibration Plot
- ROC Curve
  - AUC = 0.78
- Confusion Matrix
  - Accuracy = 0.8381
  - Sensitivity = 0.9514
  - Specificity = 0.3118

		Actual	
		AFib Complication	No AFib Complication
Predicted	AFib Complication	411	64
	No AFib Complication	21	29

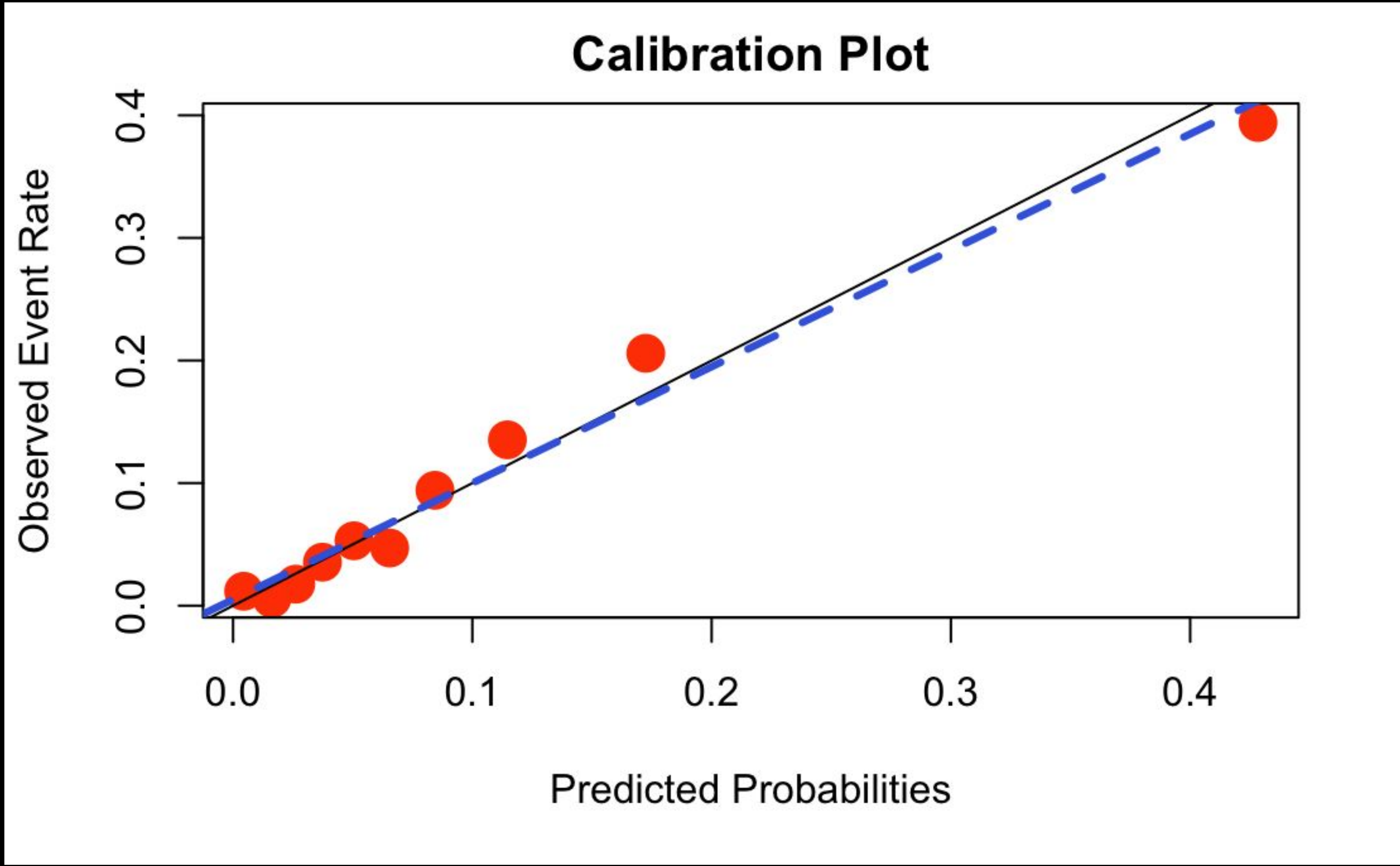
Forward Selection

Our Forward Selection

- We chose forward selection because:
  - Step and backwards selection were computationally intensive
- Used $\alpha = 0.05$
- Final model had 25 predictors

Evaluating Forward Selection

- Calibration Plot

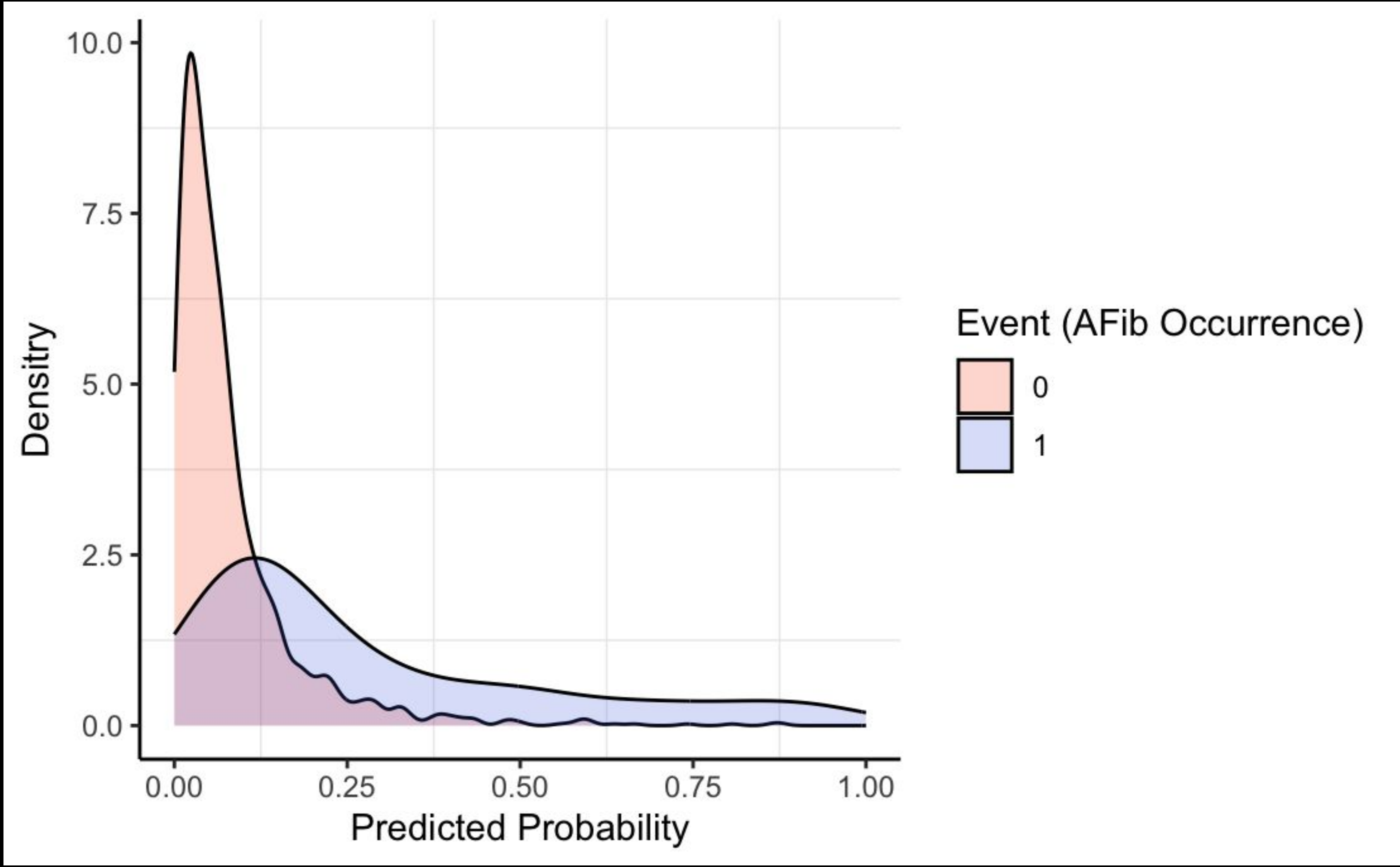


Slope: point estimate = 0.949
confidence interval [0.831, 1.068]

Intercept: point estimate = 0.005
confidence interval [-0.013, 0.024]

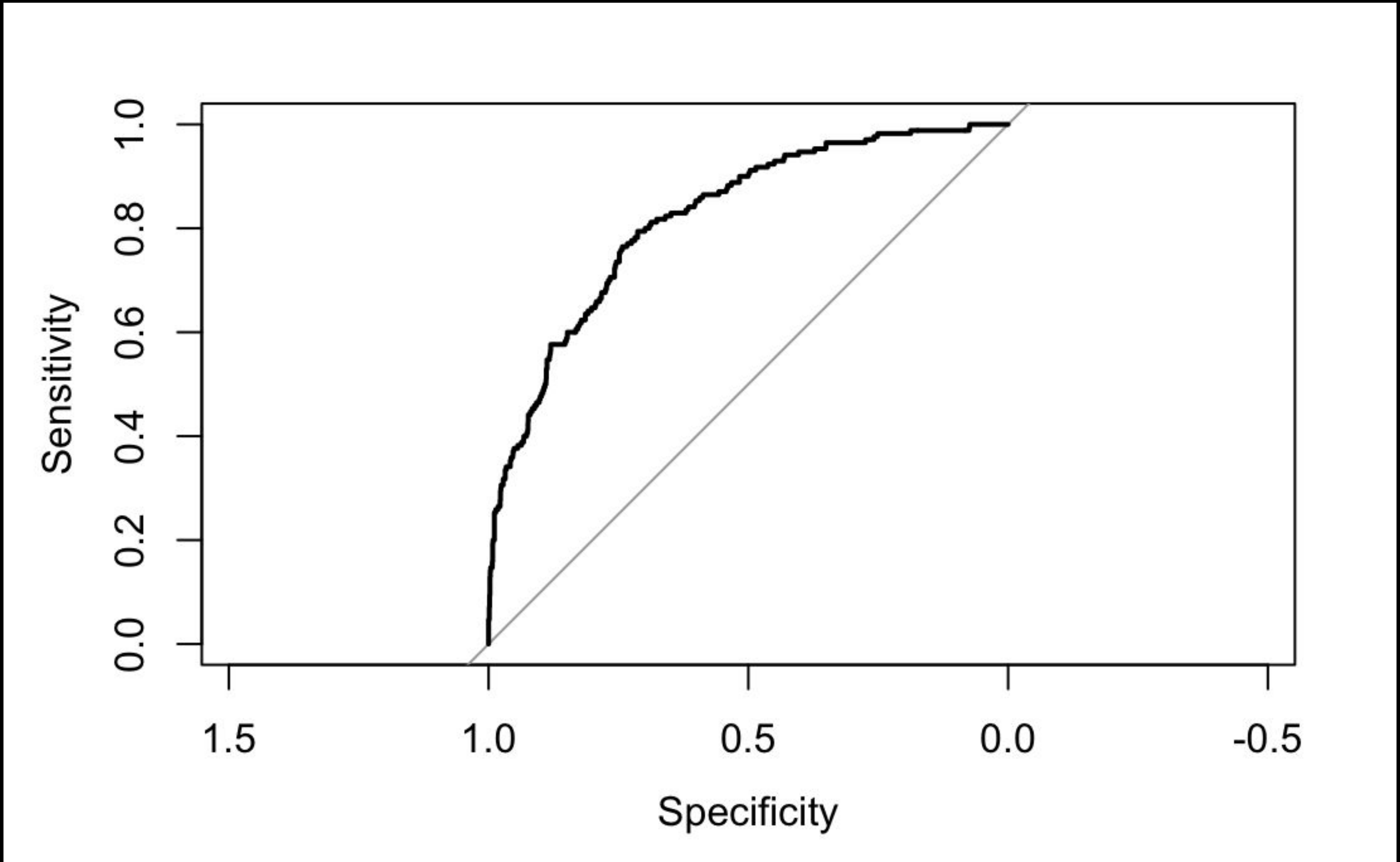
Evaluating Forward Selection

- Calibration Plot
- Discrimination Plot


Density
10.0
7.5
5.0
2.5
0.0
0.00
0.25
0.50
0.75
1.00
Predicted Probability
Event (AFib Occurrence)
0
1

Evaluating Forward Selection

- Calibration Plot
- Discrimination Plot
- AUC of ROC Plot



AUC: point estimate = 0.8207

Confidence Interval [0.7880, 0.8534]

Evaluating Forward Selection

- Calibration Plot
- Discrimination Plot
- AUC of ROC Plot
- Confusion Matrix
  - Accuracy = 0.8229
  - Sensitivity = 0.9443
  - Specificity = 0.2934

		Actual	
		AFib Complication	No AFib Complication
Predicted	AFib Complication	1306	224
	No AFib Complication	77	93

Best Model

Forward Selection

Why Forward Selection?

- Well-calibrated
- AUC under ROC = 0.82
- High accuracy (0.823)

Covariates: Anamnesis

Table 2. Significant covariates of anamnesis from the forward selection model			
Covariate	Log Odds	Probability	p-value
Atrial Fibrillation (nr03)	1.75	0.85	<0.001
Persistent Atrial Fibrillation (nr04)	-2.26	0.09	0.048
Ventricular Fibrillation (nr07)	17.14	0.99	0.974
Arrhythmia (nr11)	1.12	0.75	0.011
Chronic Bronchitis (zab_leg_01)	-0.88	0.29	0.055
Bronchial Asthma (zab_leg_03)	1.21	0.77	0.001
Exertional Angina Pectoris (IBS_POST1)	0.32	0.58	0.193
Unstable Angina Pectoris (IBS_POST2)	-0.45	0.39	0.082

Covariates: Demographic & Testing

Table 3. Significant covariates of demographics and testing from the forward selection model

Covariate	Log Odds	Probability	p-value
Sex (SEX)	-0.45	0.39	<0.001
Age (AGE)	0.04	0.51	0.03
Hypokalemia (GIPO_K)	-0.79	0.31	0.006
Blood Potassium (K_BLOOD)	-0.78	0.31	<0.001

Covariates: Treatments

Table 4. Significant covariates of treatments from the forward selection model

Covariate	Log Odds	Probability	p-value
Opioids Used on Day 3 Once (NA_R_3_n1)	1.13	0.76	0.001
Opioids Used on Day 3 Twice (NA_R_3_n2)	-0.16	0.46	0.880
NSAID Use by ECT (NOT_NA_KB)	-1.89	0.13	<0.001
Lidocaine Use in ICU (LID_S_n)	0.70	0.67	<0.001
Heparin use in ICU (GEPAR_S_n)	0.45	0.61	0.041

Covariates: ECG Measurements

Table 5. Significant covariates of ECG measurements from the forward selection model

Covariate	Log Odds	Probability	p-value
Paroxysms of Atrial Fibrillation on ECG (n_r_ecg_p_05)	1.29	0.78	<0.001
Premature Atrial Contractions on ECG (n_r_ecg_p_01)	1.12	0.75	0.002
Idioventricular on ECG (ritm_ecg_p_06)	2.99	0.95	0.047
Bradycardia on ECG (ritm_ecg_p_08)	2.05	0.89	<0.001

Conclusions

Conclusions

- The model developed from forward selection was the best model
- This model included many different factors, from patient history, patient biometrics, treatments, and demographics
- Is well calibrated, discriminates well, and had a high accuracy
- Limitations
  - This model is done solely on a Russian cohort which is not fully representative of a wide range of patient populations
  - The data set was missing 7.6% of values. Using imputation adds more variance to the model

Acknowledgements

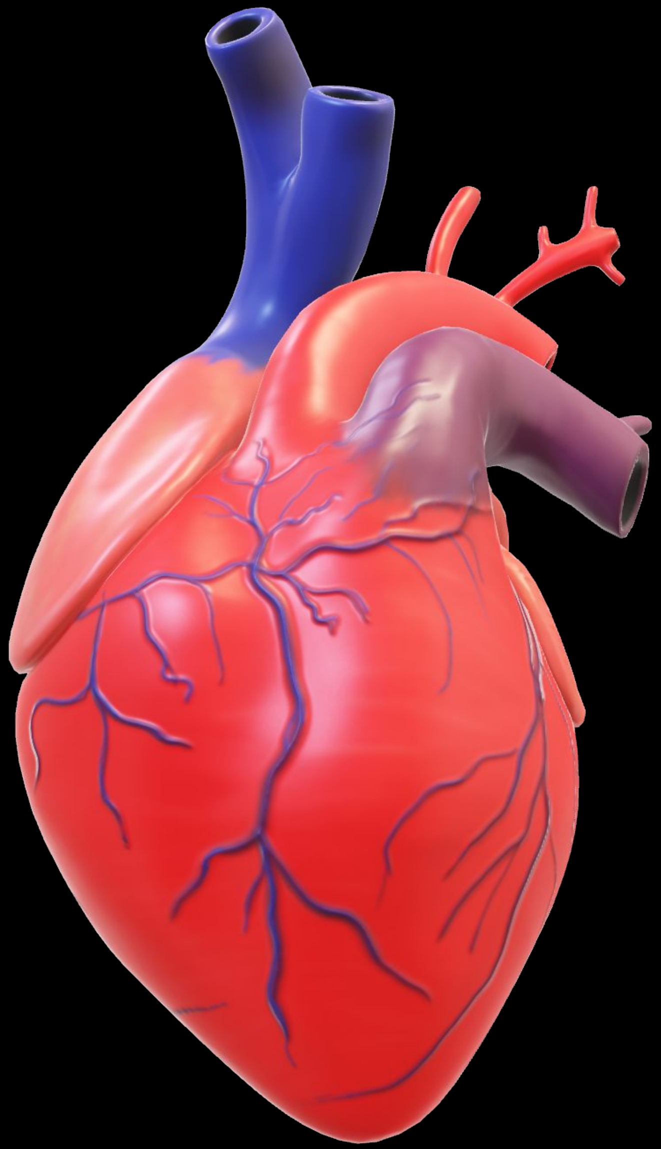
SIBS Faculty

SIBS Mentors

Dr. Sujit Ghosh

Dr. Herle McGowan

Dr. Megan Neely



References

1. Atrial fibrillation. Heart and Stroke Foundation of Canada. Accessed July 14, 2023. https://www.heartandstroke.ca/heart-disease/conditions/atrial-fibrillation#:~:text=People%20with%20atrial%20fibrillation%20have,get%20pumped%20to%20the%20brain.
2. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work?. Int J Methods Psychiatr Res. 2011;20(1):40-49. doi:10.1002/mpr.329