Supplementary Table 1

	One-time Admissions (N=21568)	2–5 Admissions (N=13859)	5 or More Admissions (N=4488)	Over All (N=39915)
Gender				
Missing	1 (0.0%)	2 (0.0%)	1 (0.0%)	4 (0.0%)
Male	13247 (61.4%)	8618 (62.2%)	2521 (56.2%)	24387 (61.1%)
Female	8320 (38.6%)	5239 (37.8%)	1966 (43.8%)	15525 (38.9%)
Age				
Mean (SD)	61.8 (12.9)	61.8 (13.0)	58.9 (14.5)	61.5 (13.2)
Median [Min, Max]	63.0 [18.0, 99.0]	63.0 [18.0, 98.0]	59.0 [18.0, 98.0]	62.0 [18.0, 99.0]
Medical Reform Policy Phase I				
Mean (SD)	0.249 (0.433)	0.646 (0.970)	1.60 (4.74)	0.540 (1.77)
Median [Min, Max]	0 [0, 1.00]	0 [0, 5.00]	0 [0, 124]	0 [0, 124]
Medical Reform Policy Phase II				
Mean (SD)	0.504 (0.500)	1.04 (1.08)	4.64 (6.01)	1.16 (2.49)
Median [Min, Max]	1.00 [0, 1.00]	1.00 [0, 5.00]	3.00 [0, 86.0]	1.00 [0, 86.0]
Medical Reform Policy Phase III				
Mean (SD)	0.246 (0.431)	1.07 (1.22)	6.96 (9.48)	1.29 (3.86)
Median [Min, Max]	0 [0, 1.00]	1.00 [0, 5.00]	5.00 [0, 200]	0 [0, 200]

* Medical Reform Policy Phase I: 2010.3.12-2015.12.31;
Medical Reform Policy Phase II: 2015.12.31-2018.8.31;
Medical Reform Policy Phase III: 2018.8.31-2024.11.30

Supplementary Table 1.
Baseline characteristics of discovery cohort from Wuhan with three major time groups.

Supplementary Table 2

Shandong Cancer Hospital EHR Database

Model name	Precision	Recall	F1-score	Accuracy
Random Forest	0.81583976	0.855083867	0.83500096	0.732992356
XGBoost	0.790115433	1	0.88275361	0.790115433
LightGBM	0.845173601	0.689066915	0.75917845	0.654591181
Linear	0.817499754	0.594274168	0.68823907	0.57460757
Ridge	0.817401991	0.594685455	0.68848013	0.574791246
Hist Gradient Boosting	0.845234163	0.686456135	0.75761553	0.65295223
LSTM	0.79065815	0.998283323	0.88242221	0.789804598
GRU	0.790572581	0.998712492	0.88253652	0.789945886

MIMIC EHR Database

Model name	Precision	Recall	F1-score	Accuracy
Hist Gradient Boosting	0.58493796	0.59130435	0.557737	0.591304
Random Forest	0.58371384	0.59130435	0.564522	0.591304
Logistic Regression	0.59922034	0.60421607	0.580775	0.604216
Decision Tree	0.53178274	0.53175231	0.531767	0.531752
Gradient Boosting	0.58862665	0.58945982	0.536709	0.58946
SVM	0.59861659	0.60131752	0.567992	0.601318
RNN	0.56176128	0.56442688	0.562713	0.564427
LSTM	0.56699019	0.56943347	0.567879	0.569433
GRU	0.5720326	0.57944664	0.570761	0.579447
Transformer	0.57495556	0.56389987	0.565228	0.5639

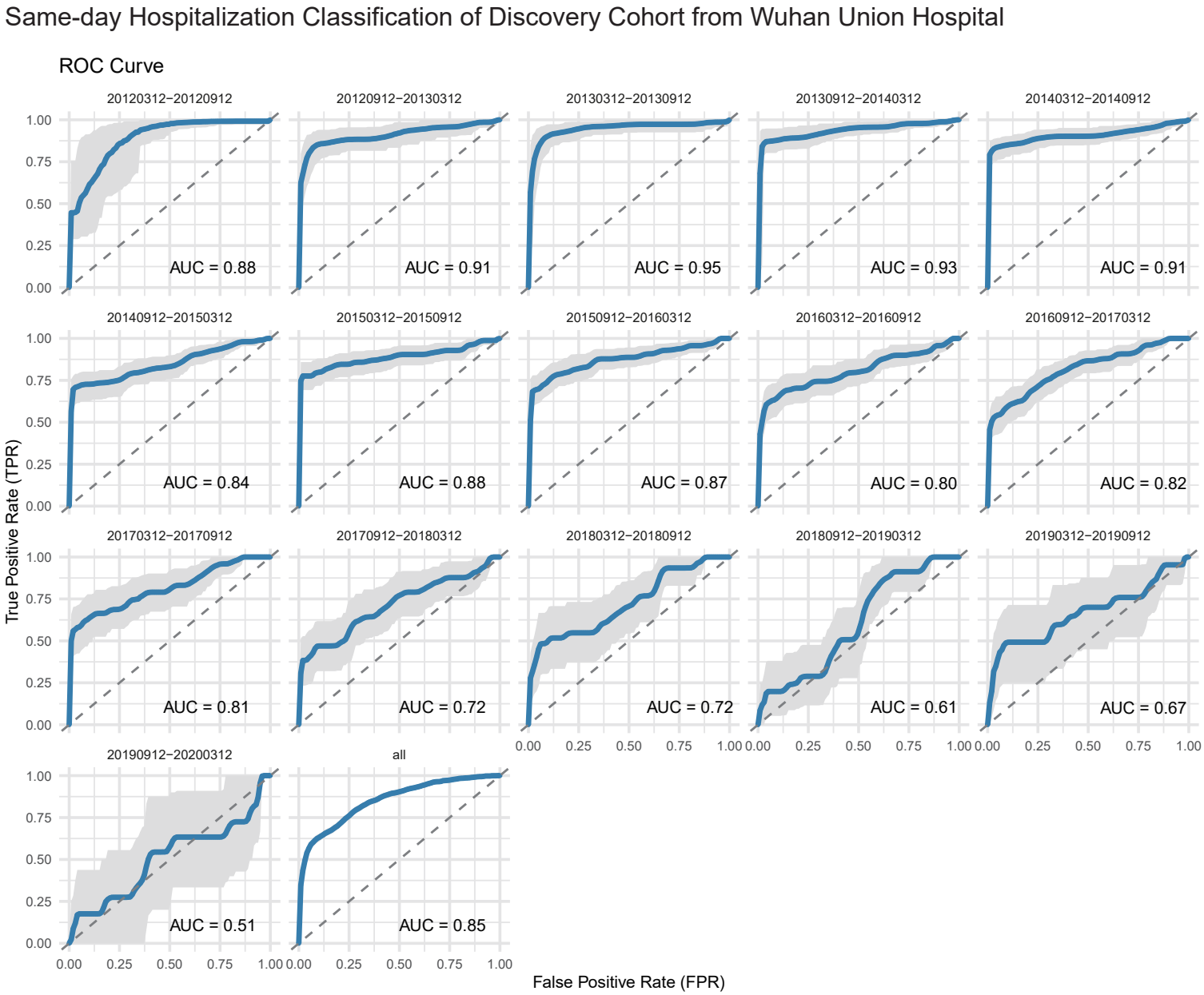
SupplementaryTable 2.
Performance of Predictive Models on EHR Data from Shandong Cancer Hospital and MIMIC-III.

Supplementary Table 3

	One-time Admission (N=14985)	2–5 Admissions (N=9516)	5 or More Admissions (N=6578)	Over All (N=31079)
Gender				
Male	7969 (53.2%)	5184 (54.5%)	3574 (54.3%)	16727 (53.8%)
Female	7016 (46.8%)	4331 (45.5%)	3004 (45.7%)	14351 (46.2%)
Missing	0 (0%)	1 (0.0%)	0 (0%)	1 (0.0%)
Age				
Mean (SD)	64.0 (10.4)	62.7 (9.58)	61.0 (9.54)	62.9 (10.1)
Median [Min, Max]	65.0 [5.00, 98.0]	63.0 [1.00, 99.0]	62.0 [1.00, 90.0]	63.0 [1.00, 99.0]
Missing	5 (0.0%)	1 (0.0%)	4 (0.1%)	10 (0.0%)

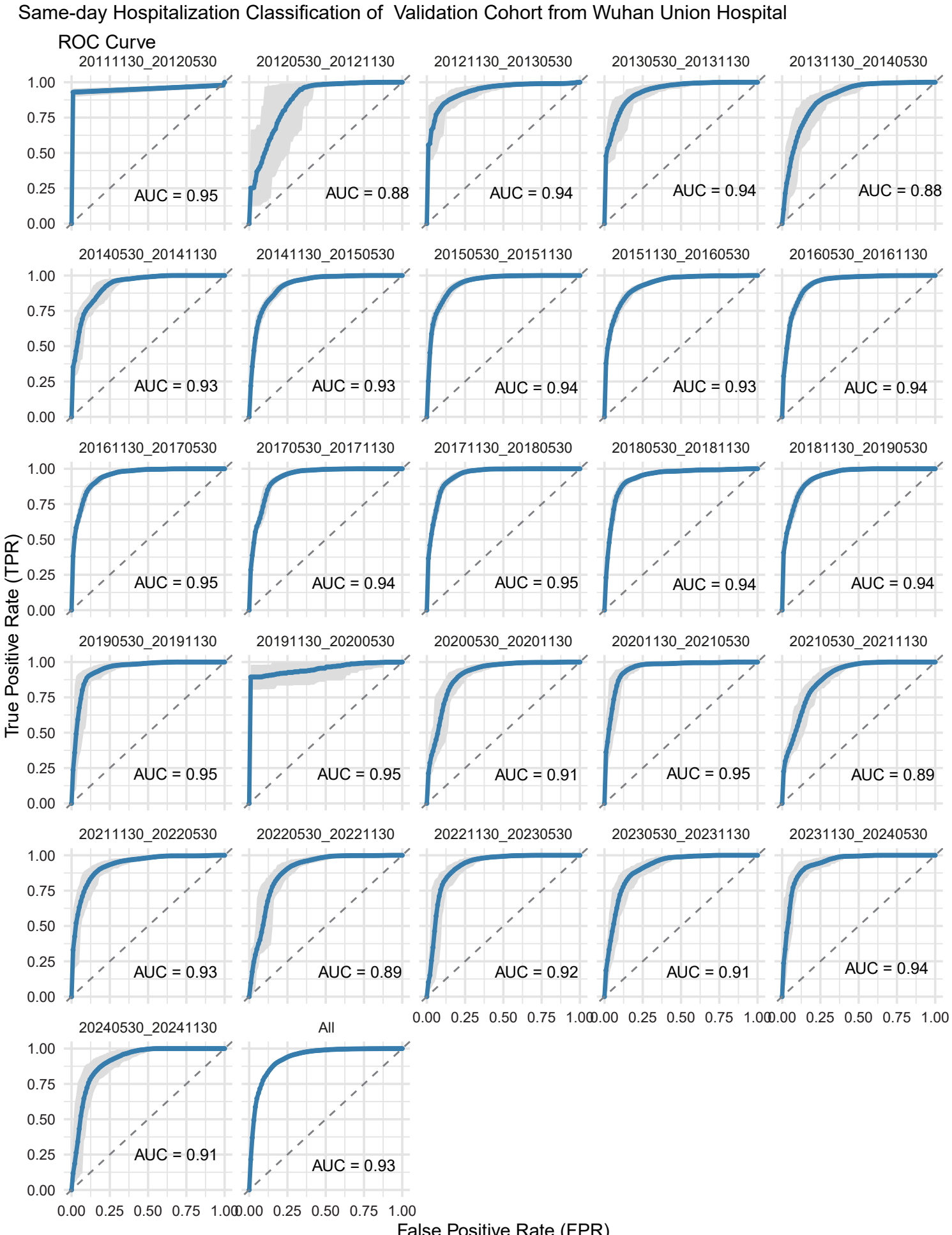
Supplementary Table 3.
Basic Description of the EHR Data from Shandong Cancer Hospital.

Supplementary Figure 1



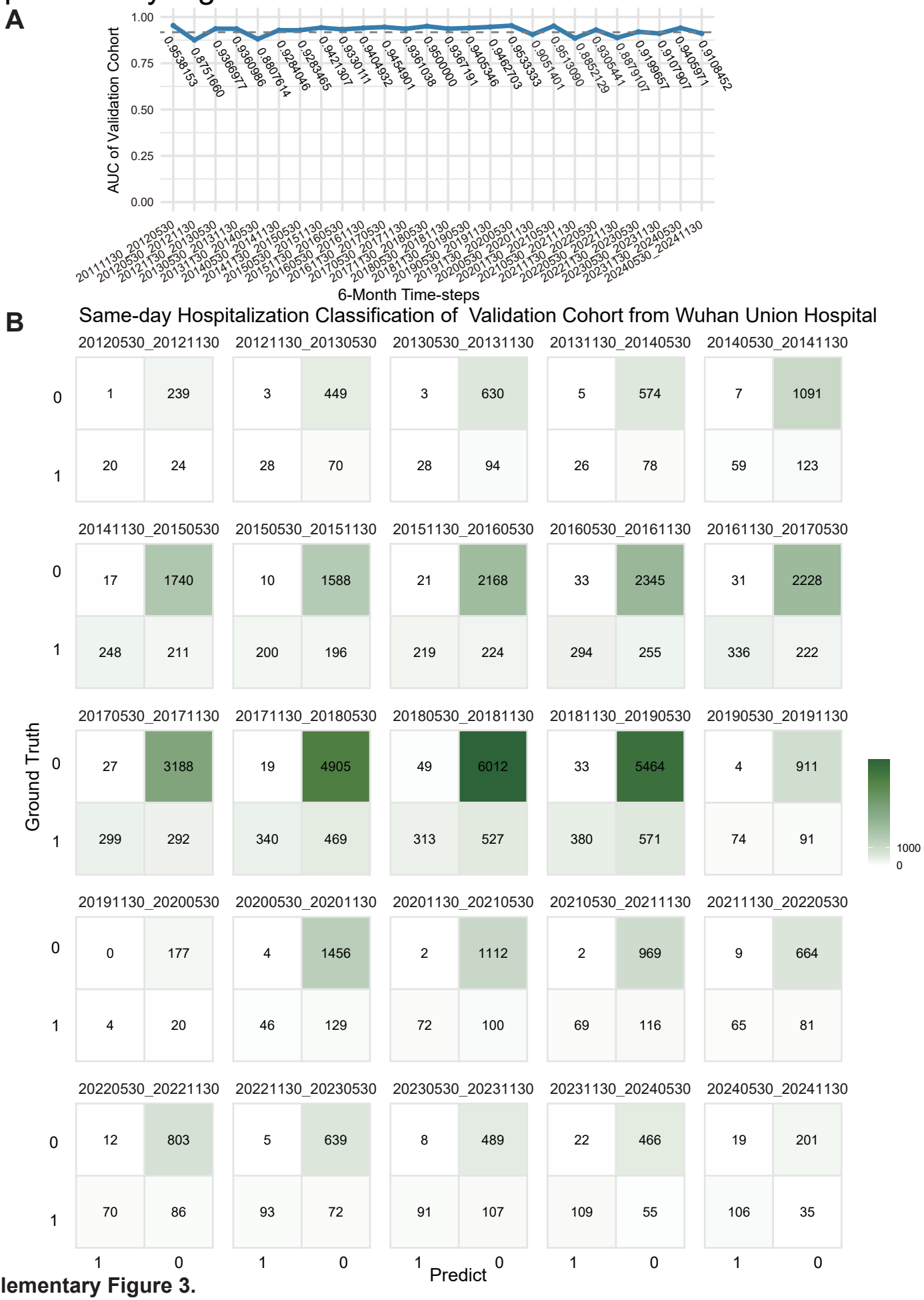
SupplementaryFigure 1.
Evaluation of Predictive Performance on Wuhan Union Hospital EHR Discovery Cohort Data Using the Same-day Admission Classification Task. Receiver Operating Characteristic (ROC) curves were used to assess prediction performance at 6-month timesteps. The blue line represents the ROC curve, and the shaded grey area indicates the 95% confidence interval.

Supplementary Figure 2



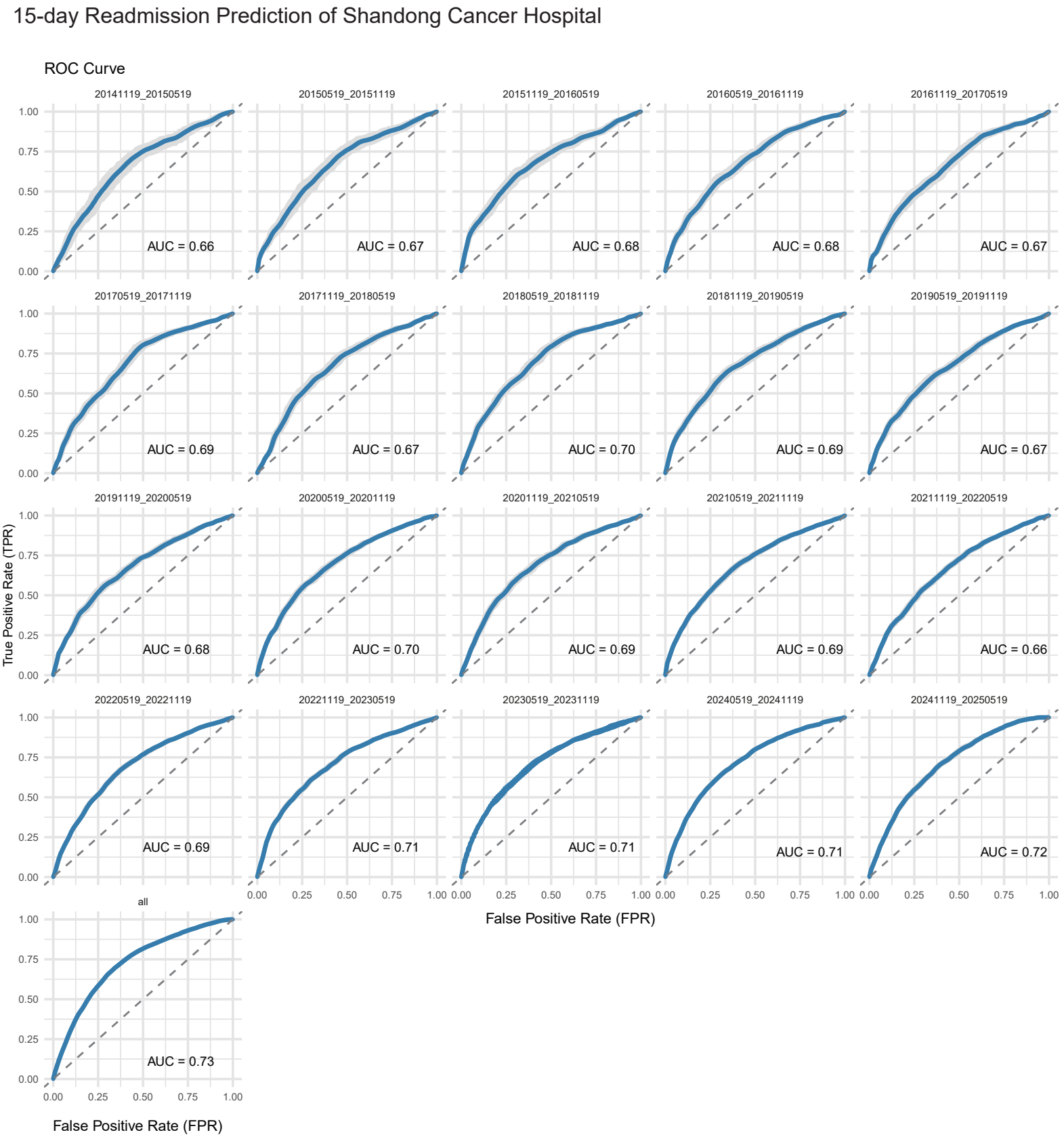
Supplementary Figure 2.
Evaluation of Predictive Performance on Wuhan Union Hospital EHR Validation Cohort Using the Same-day Hospitalization Classification Task. ROC curves were used to assess prediction performance at 6-month timesteps.The blue line represents the ROC curve, and the shaded grey area indicates the 95% confidence interval.

Supplementary Figure 3



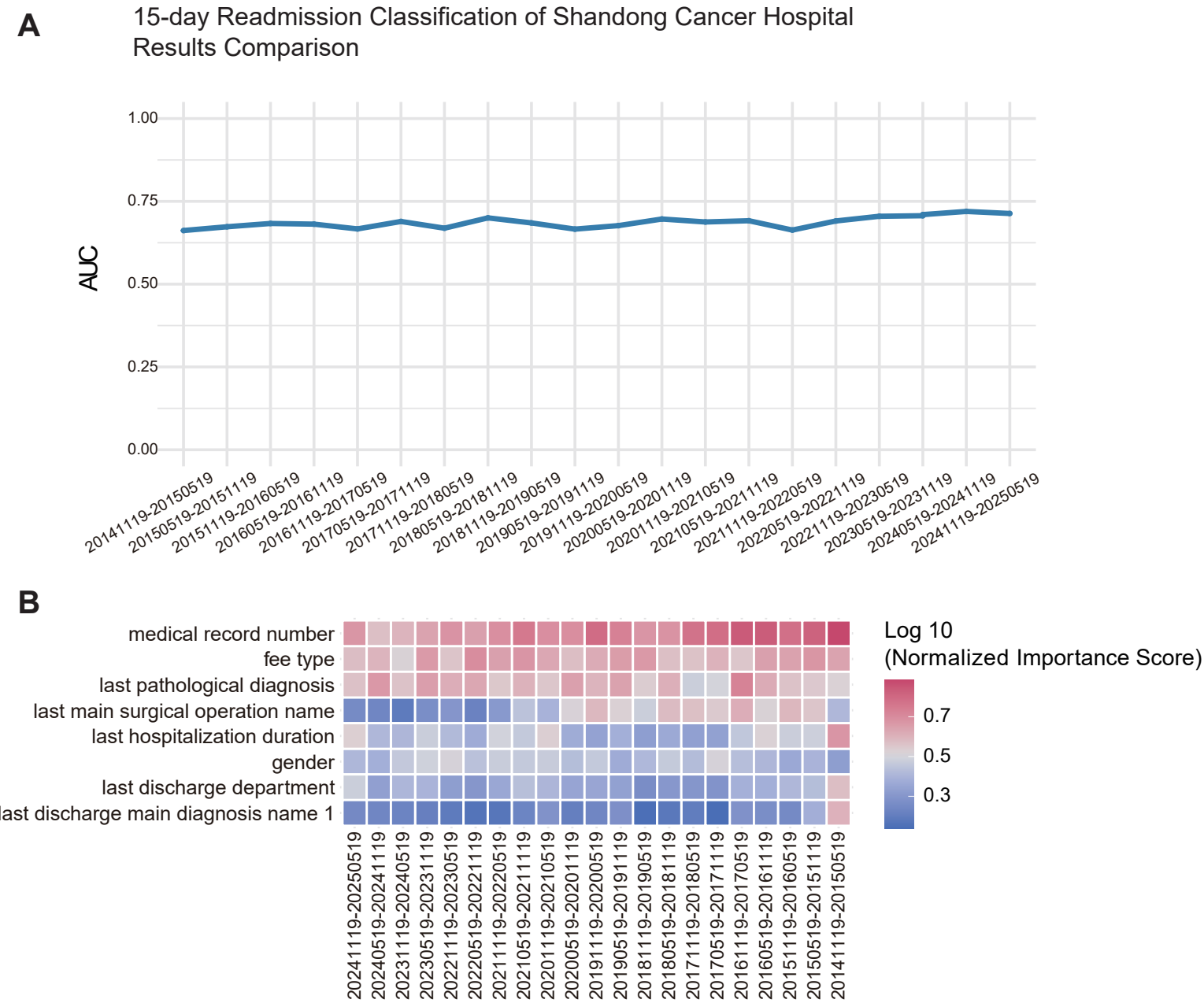
Supplementary Figure 3.
Performance of same-day Hospitalization classification across 6-month time intervals in the validation cohort.
(A) AUC for each 6-month timestep in the validation cohort, spanning from 2012 to 2024.
AUC values remained consistently above 0.80 in earlier phases and declined slightly in later stages.
(B) Confusion matrices for each 6-month timestep. Each matrix represents the model's performance in classifying same-day admissions, with the intensity of green indicating the number of true positive and true negative predictions.

Supplementary Figure 4



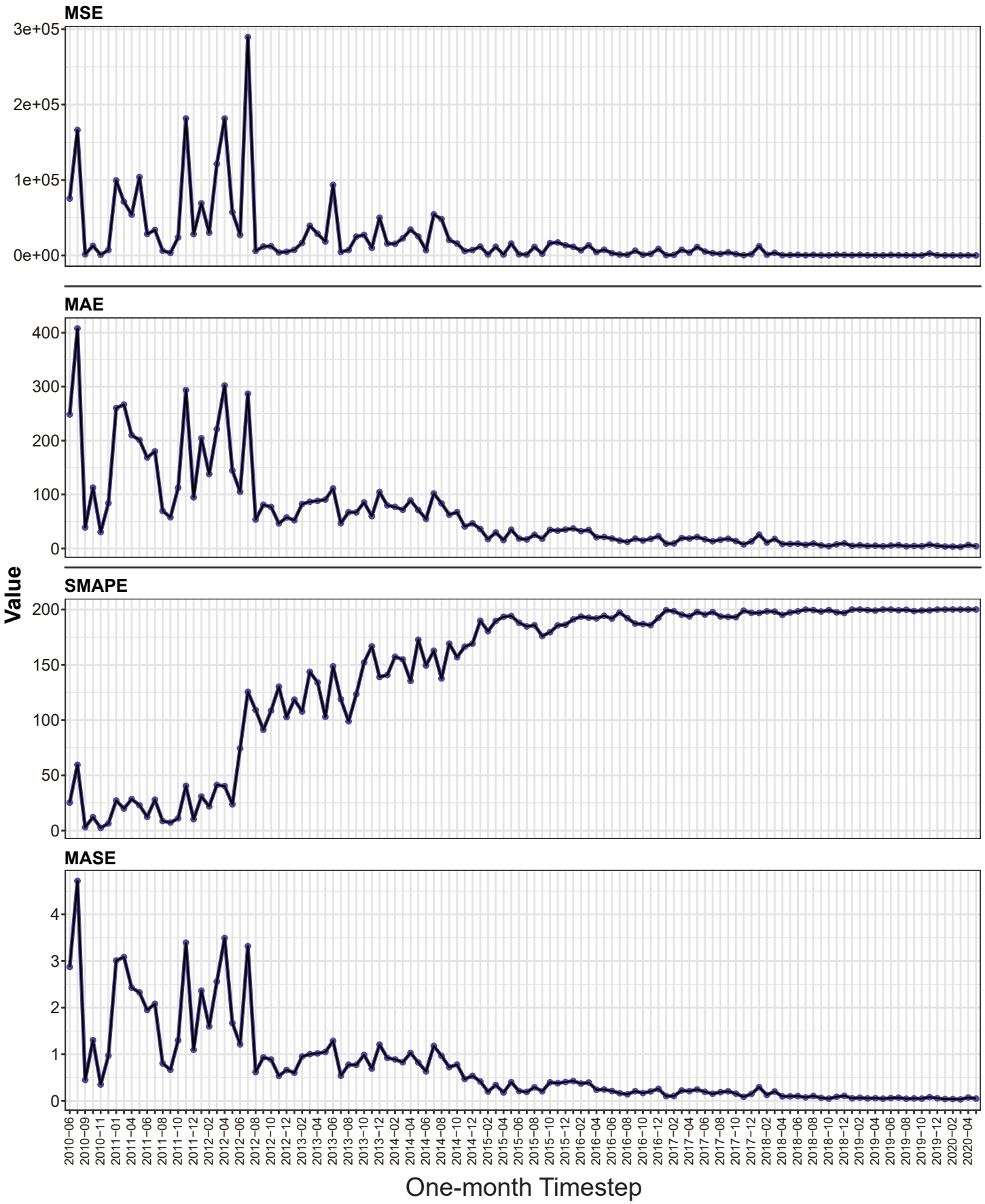
Supplementary Figure 4.
Evaluation of 15-day readmission prediction task Performance on Shandong Cancer Hospital EHR Data Using the Same-day Hospitalization Classification Task. Receiver Operating Characteristic (ROC) curves were used to assess prediction performance at 6-month timesteps. The blue line represents the ROC curve, and the shaded grey area indicates the 95% confidence interval.

Supplementary Figure 5



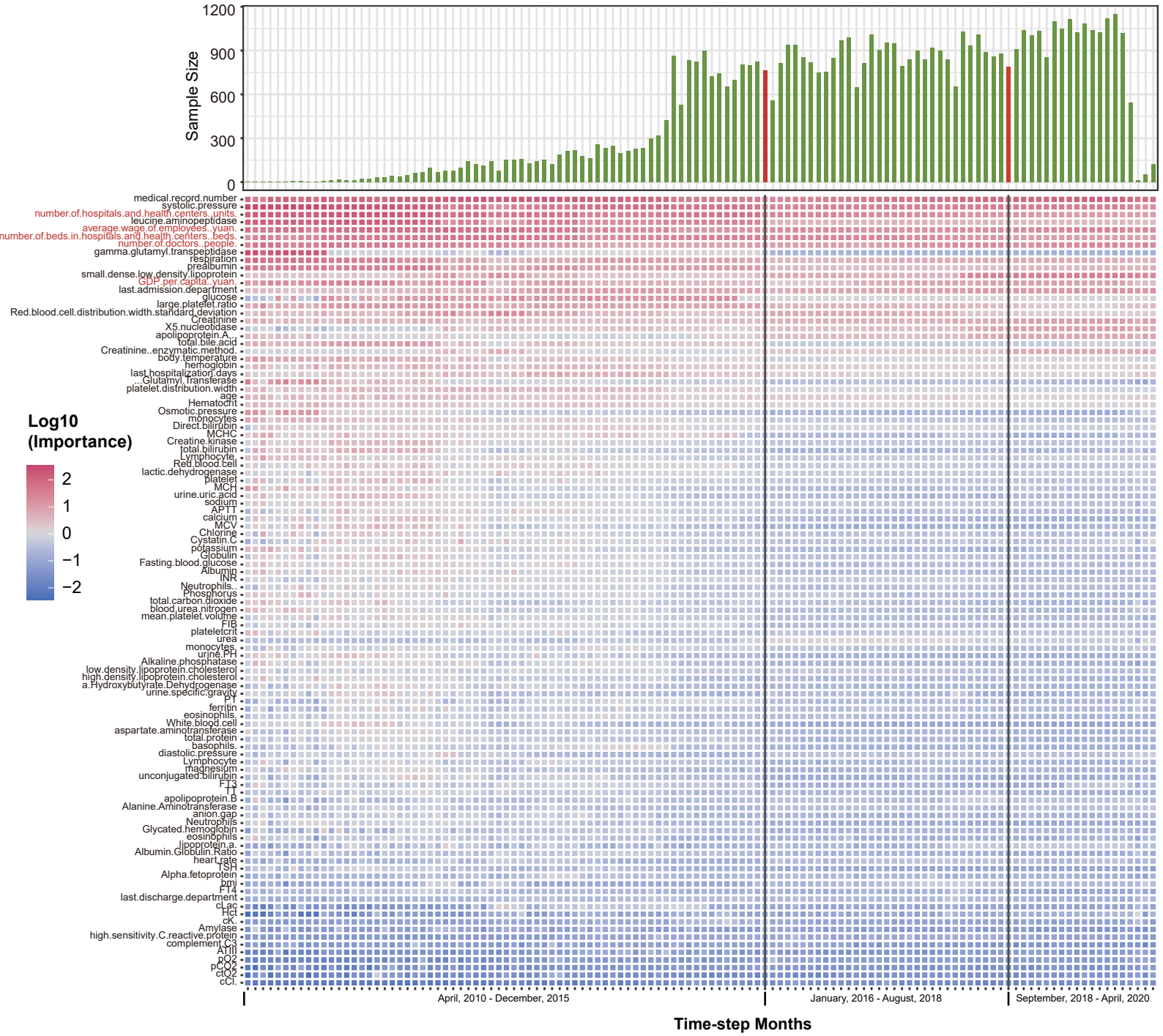
Supplementary Figure 5.
(A)Comparison of Predictive Performance Across Different Timesteps. Each point represents the Area Under the Curve (AUC) corresponding to the prediction performance at a specific timestep.
(B) Heatmap of Feature Contribution Scores Used in the Same-day Hospitalization Classification task. This heatmap illustrates the contribution of each feature to the model's output, evaluated by importance scores standardized using the STD score.

Supplementary Figure 6



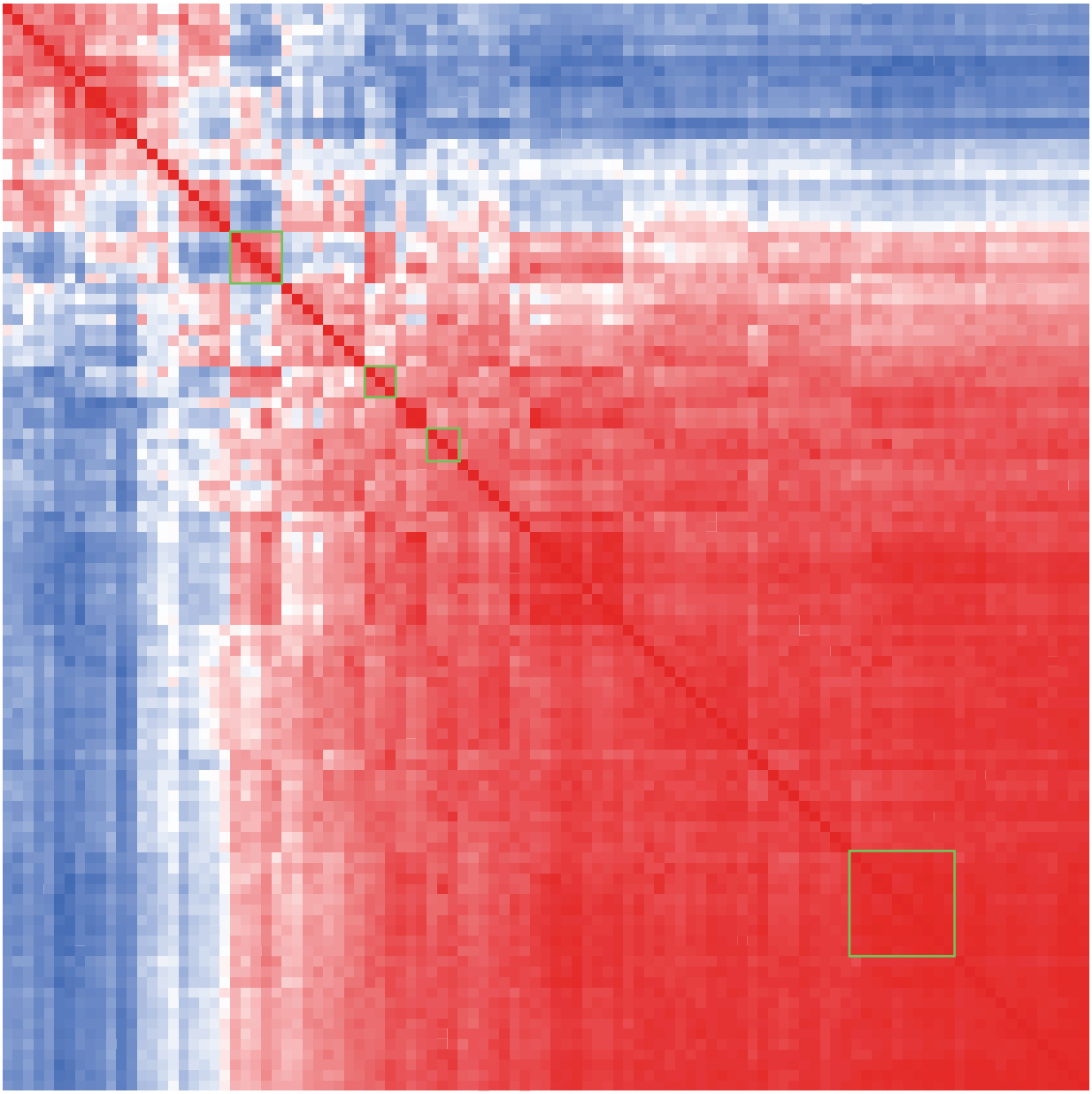
Supplementary Figure 6.
Regression of Time to Next Admission model evaluation on discovery cohort and results.

Supplementary Figure 7



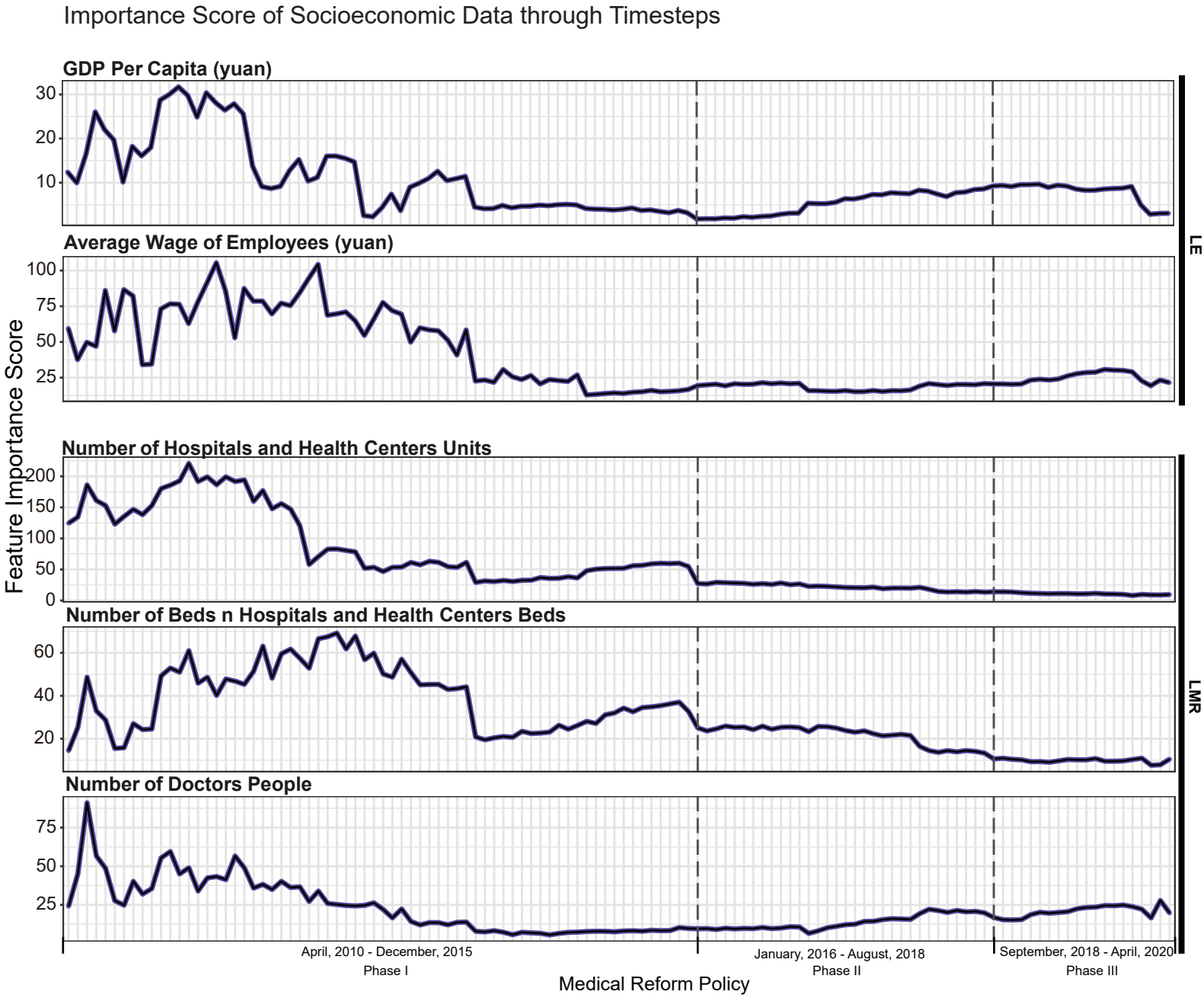
Supplementary Figure 7.
Heatmap of Feature Contribution Scores Used in the Regression of Time to Next Admission task. This heatmap illustrates the contribution of each feature to the model's output, evaluated by importance scores

Supplementary Figure 8



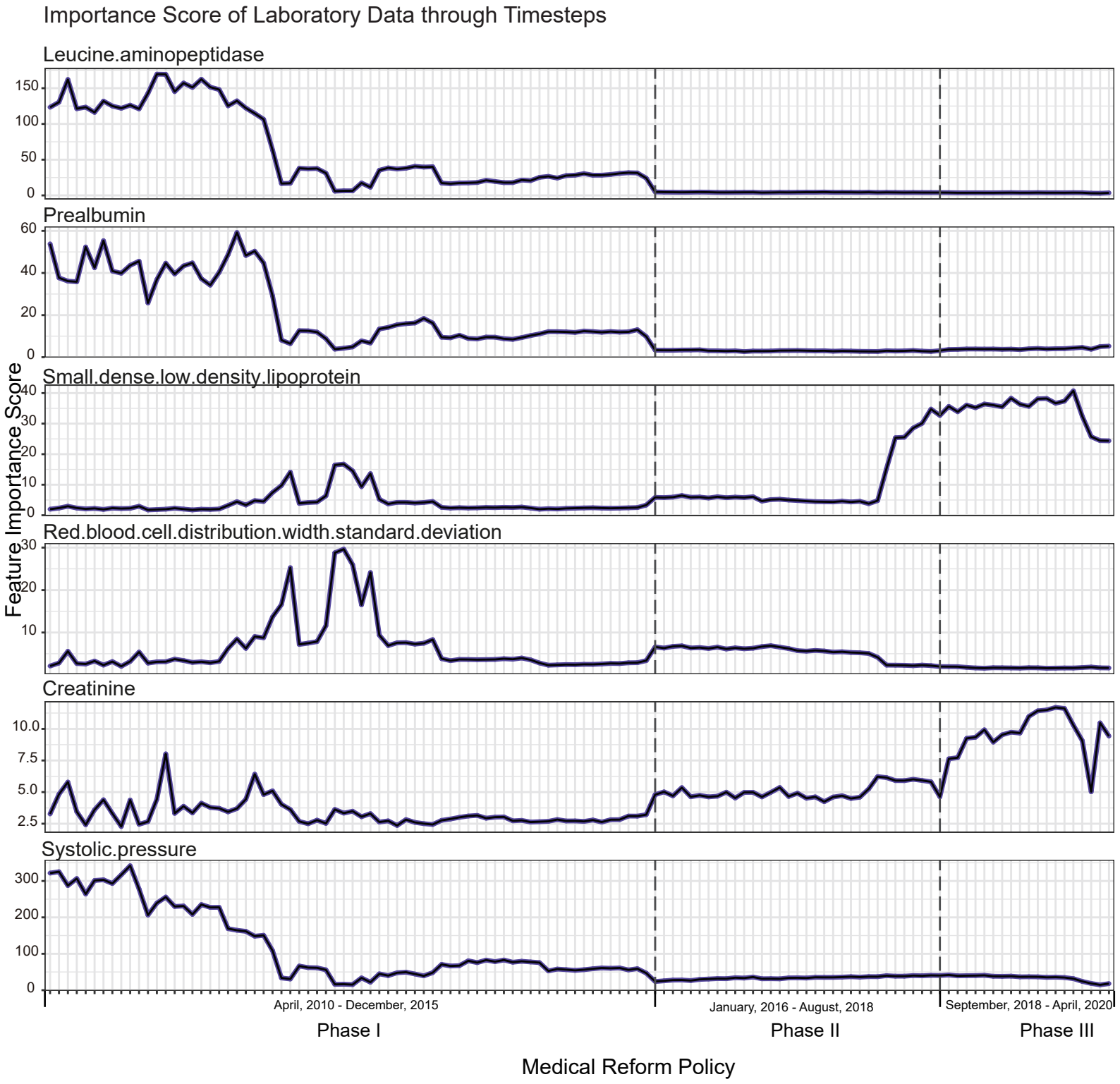
Supplementary Figure 8.
Spearman correlation matrix of model features in the Regression of Time to Next Admission task. Pairwise Spearman correlations were calculated monthly across all features. Red indicates positive correlation and blue indicates negative correlation. Socioeconomic features are labeled in red; green boxes highlight clusters of strongly correlated variables.

Supplementary Figure 9



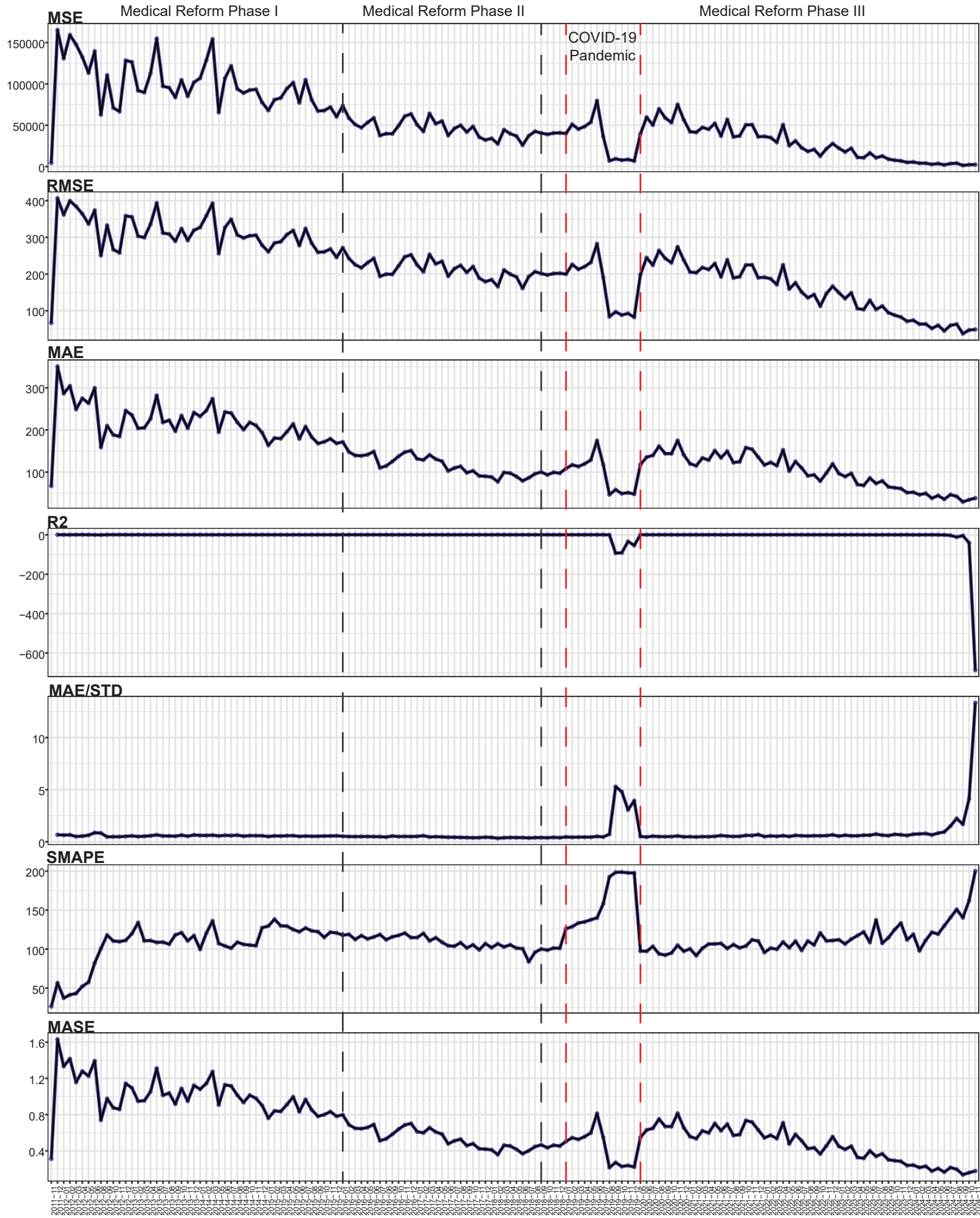
Supplementary Figure 9.
Importance scores of features related to LE and LMR indicators. The plot also shows how their importance score changes over time, reflecting the influence of non-clinical factors on patient behavior and model prediction.

Supplementary Figure 10



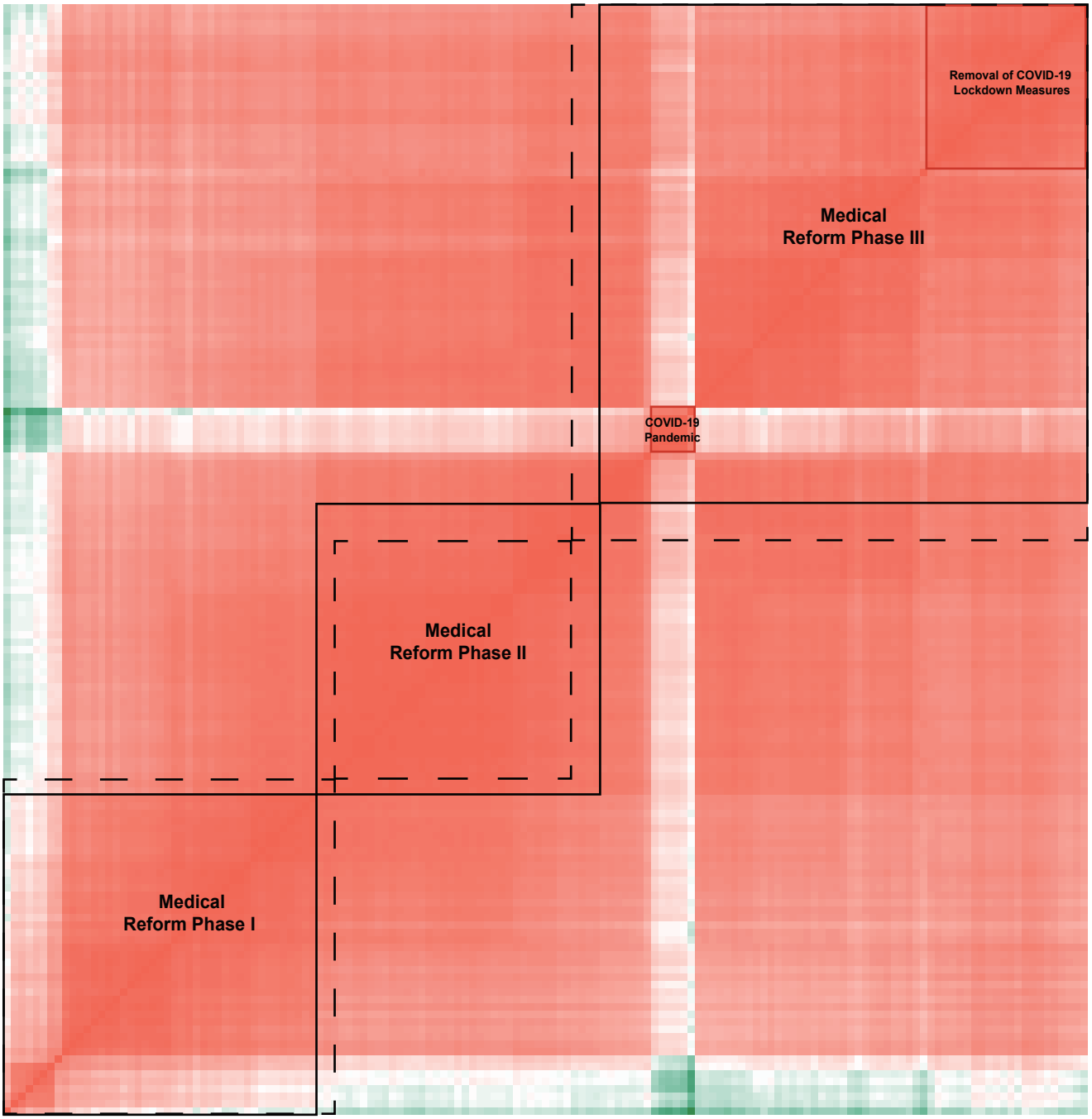
Supplementary Figure 10.
The laboratory data variables demonstrated high importance in both the Regression of Time to Next Admission models. Six time-sensitive variables were selected to illustrate their temporal patterns and influence on the model predictions.

Supplementary Figure 11



Supplementary Figure 11.
Model performance was evaluated monthly using the validation cohort from Wuhan Union Hospital. Metrics include MSE, RMSE, MAE, R^2, MAE/STD, SMAPE, and MASE.

Supplementary Figure 12



Supplementary Figure 12.
Spearman correlation was performed on Validation Cohort from Wuhan Union Hospital between feature values and time using monthly feature evaluations. Dashed lines indicate the time period of real-world external events, while solid lines represent the model-inferred periods of significant deviation.