

# **Fairness Analysis Report**

# Information about test dataset

## Grouping by sex

category	size	number patients with event
F	454	40
M	1102	97

## Grouping by age\_group

category	size	number patients with event
<50	117	15
50-65	477	44
65-75	544	57
75-85	357	20
>85	61	1

## Grouping by APACHE\_group

category	size	number patients with event
Cardiovascular	913	50
Neurological	254	37
Gastrointestinal	111	9
Respiratory	116	20
Other	65	4
Trauma	76	13
Metabolic	13	2

## Grouping by surgical\_status

category	size	number patients with event
Surgical	934	54
Non-surgical	578	76

## Grouping by admission\_type

category	size	number patients with event
emergency	874	110
elective	652	22

# Model Performance Analysis

**Goal: Comparing the model performance across cohorts of patients**

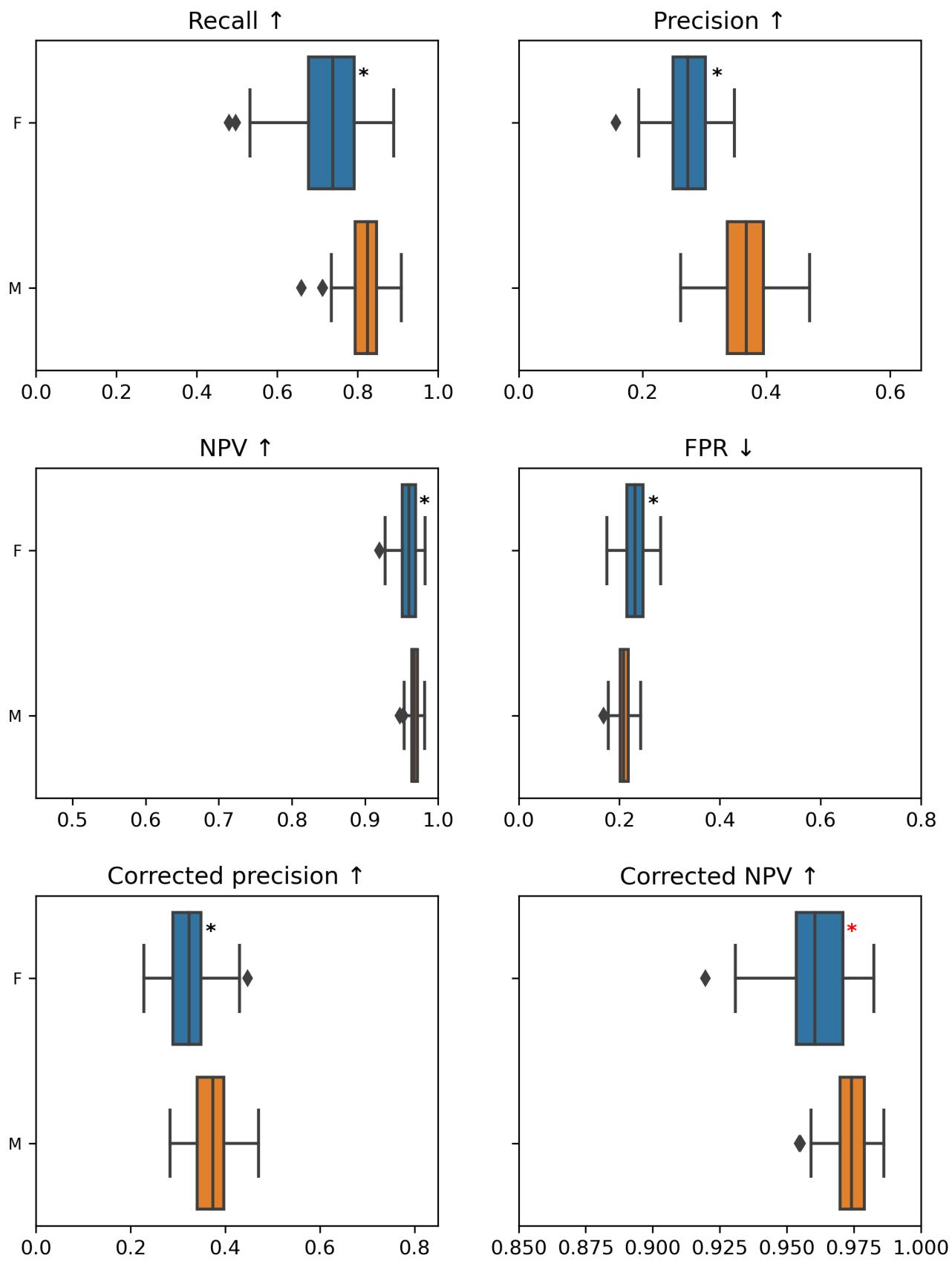
Binary metrics computed with a threshold on score of 0.157.

In the tables presenting the results of the statistical analysis, we present only metrics and groups with a significant p-value (smaller than 0.001/nb\_comparison) and whose delta is bigger than 0.

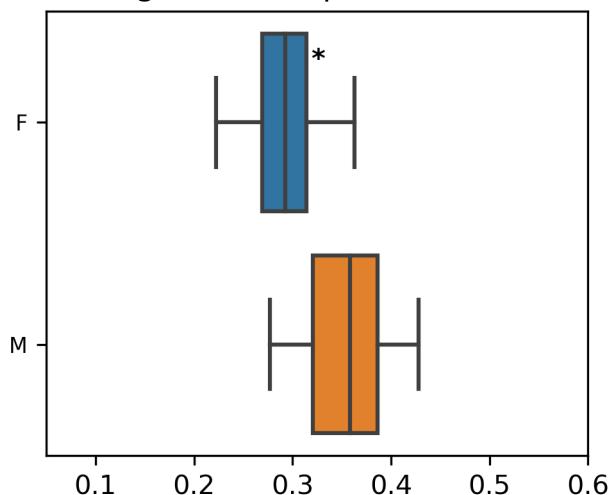
## Top 3 categories with biggest deltas in terms of performance metric discrepancy

Metric	Cat 1	Cat 2	Cat 3
Recall ↑	Metabolic (0.306)	<50 (0.124)	Neurological (0.096)
Precision ↑	75-85 (0.219)	Metabolic (0.141)	>85 (0.13)
NPV ↑	Metabolic (0.08)	Neurological (0.066)	Non-surgical (0.049)
FPR ↓	emergency (0.25)	Non-surgical (0.248)	Respiratory (0.205)
Corrected precision ↑	emergency (0.363)	Non-surgical (0.275)	Metabolic (0.142)
Corrected NPV ↑	Metabolic (0.021)	F (0.014)	Non-surgical (0.007)
Avg. score on positive class	Metabolic (0.176)	Other (0.103)	<50 (0.079)
Avg. score on negative class	emergency (0.093)	Non-surgical (0.085)	Respiratory (0.079)
AUROC ↑	Metabolic (0.167)	emergency (0.137)	Trauma (0.136)
AUPRC ↑	Metabolic (0.294)	75-85 (0.222)	Trauma (0.189)
Corrected AUPRC ↑	Metabolic (0.262)	emergency (0.258)	Non-surgical (0.24)

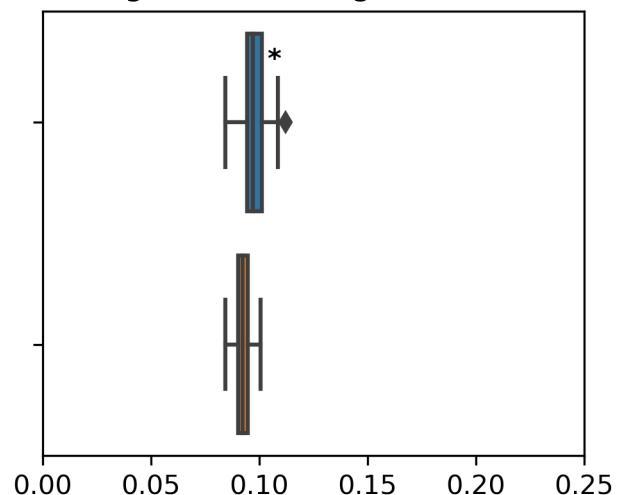
## Grouping by sex



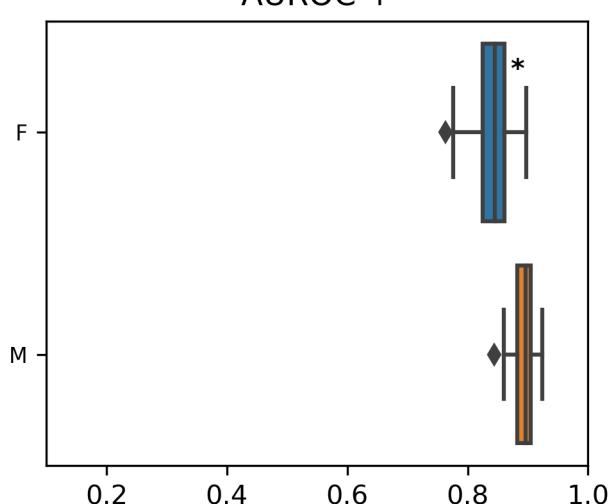
Avg. score on positive class



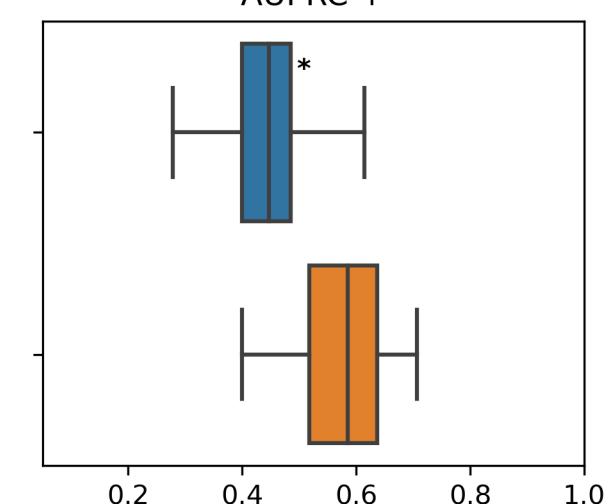
Avg. score on negative class



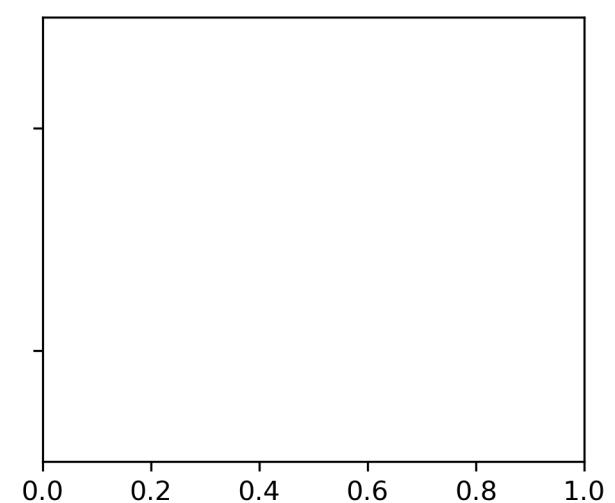
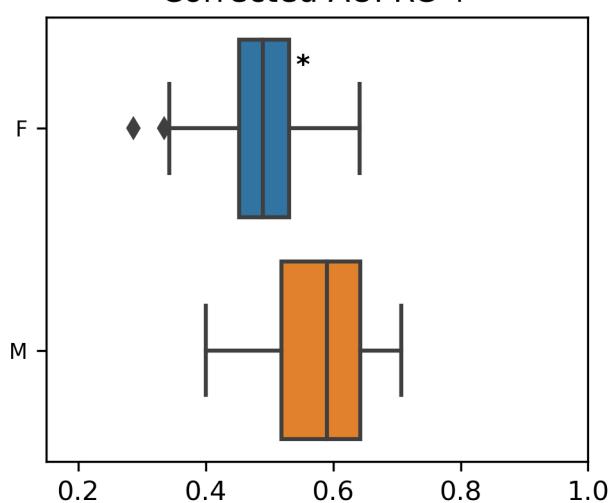
AUROC ↑



AUPRC ↑

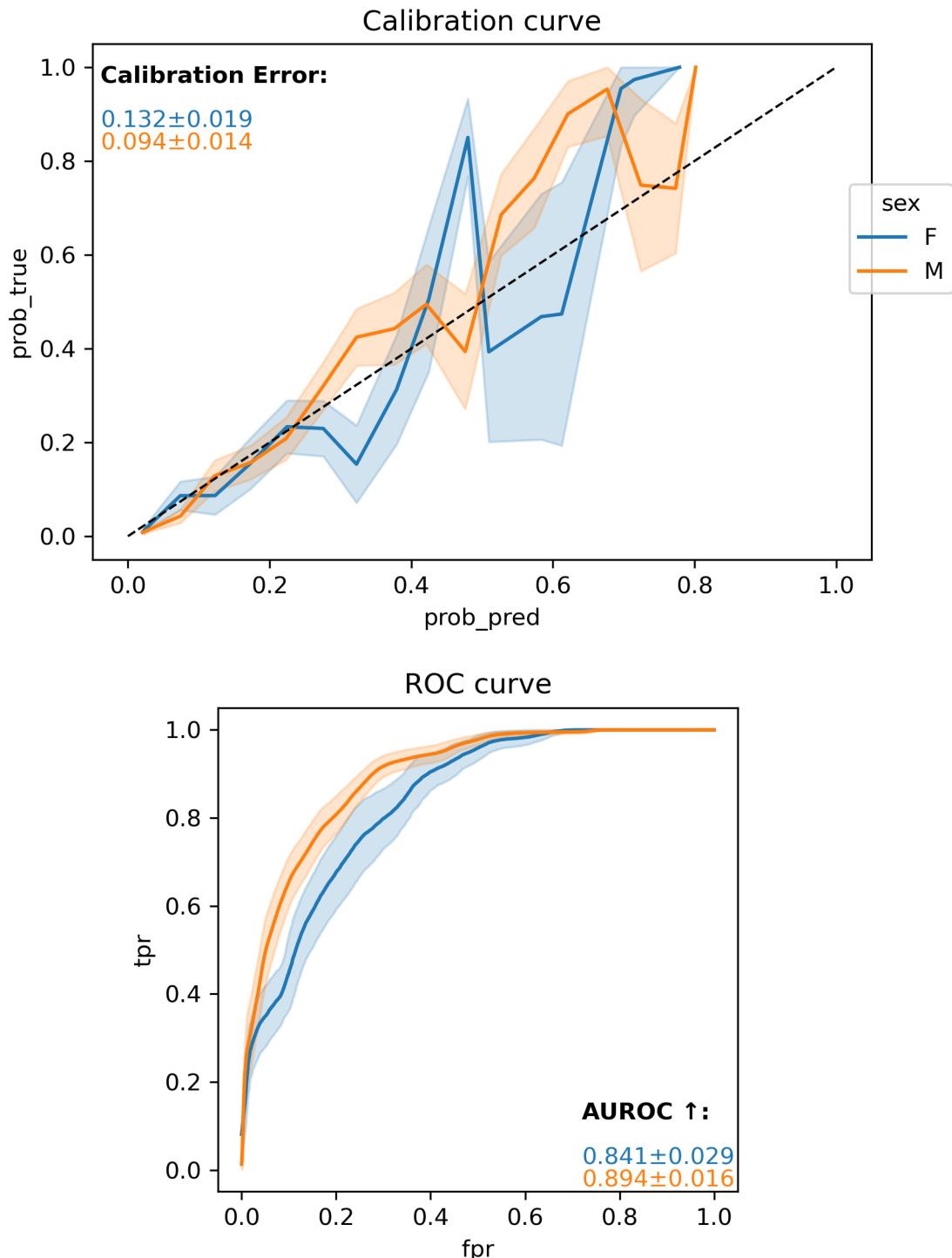


Corrected AUPRC ↑

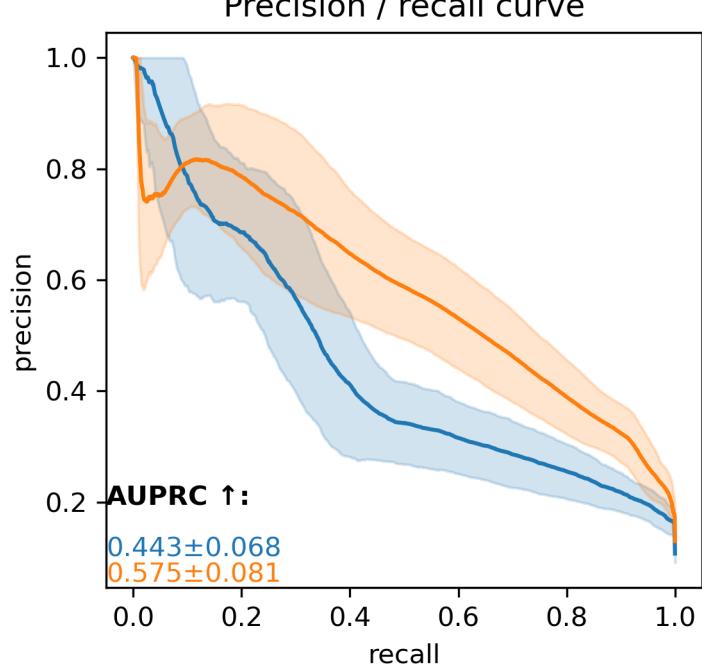


Metric	Group with worst dist.	P-value	Delta
Recall ↑	F	6.73e-18	0.086
Precision ↑	F	8.10e-27	0.095
NPV ↑	F	3.18e-06	0.007
FPR ↓	F	2.36e-12	0.022
Corrected precision ↑	F	5.10e-12	0.05
Corrected NPV ↑	F	2.89e-16	0.014

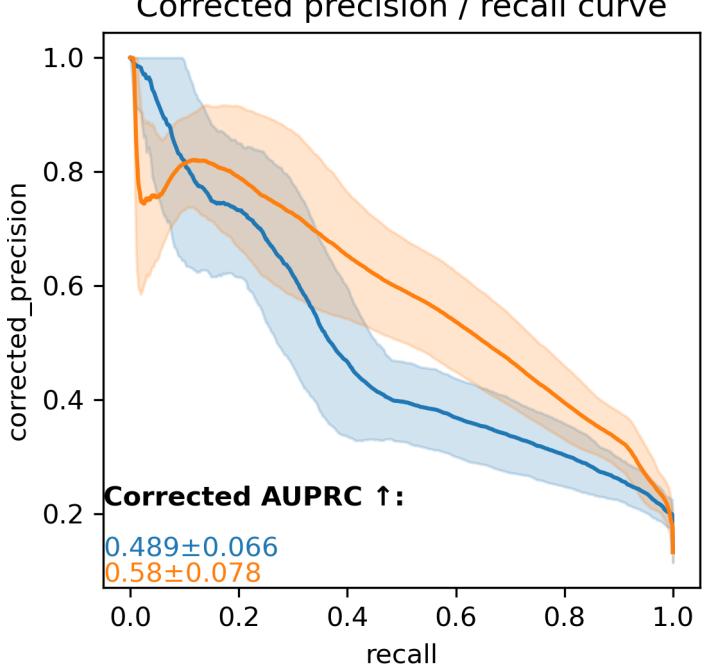
Avg. score on positive class	F	1.78e-20	0.065
Avg. score on negative class	F	9.21e-12	0.005
AUROC ↑	F	1.29e-29	0.051
AUPRC ↑	F	2.25e-21	0.139
Corrected AUPRC ↑	F	6.22e-14	0.1



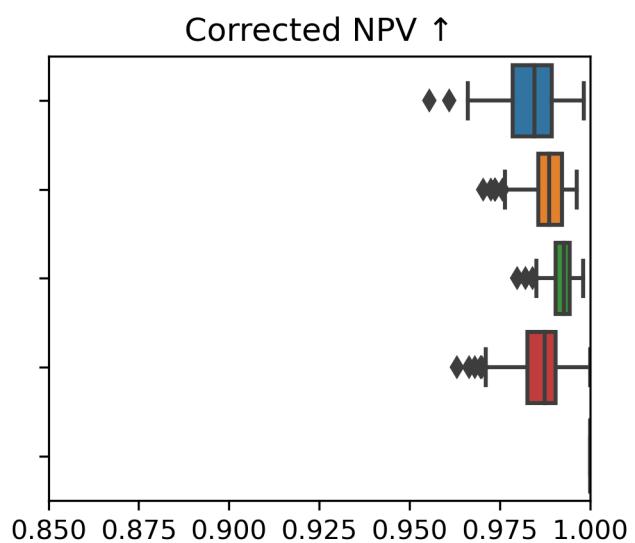
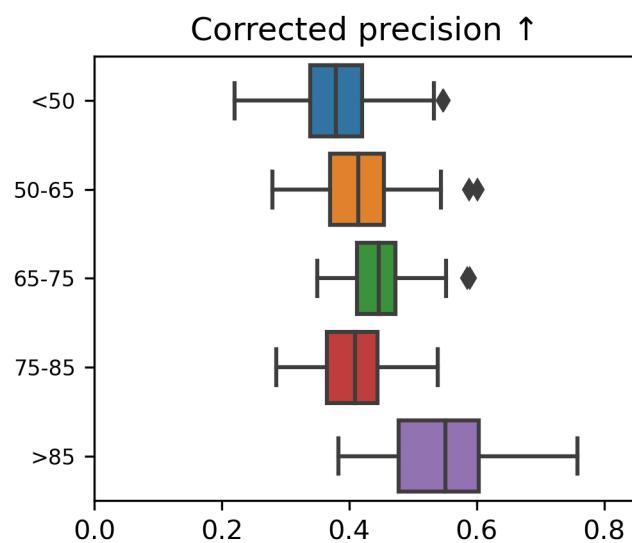
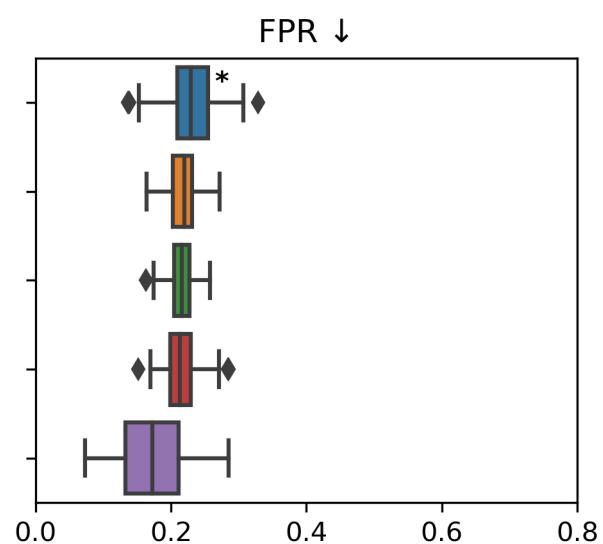
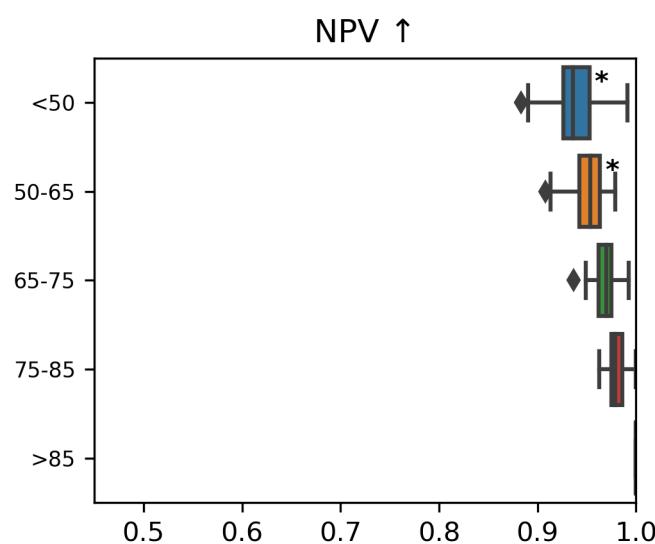
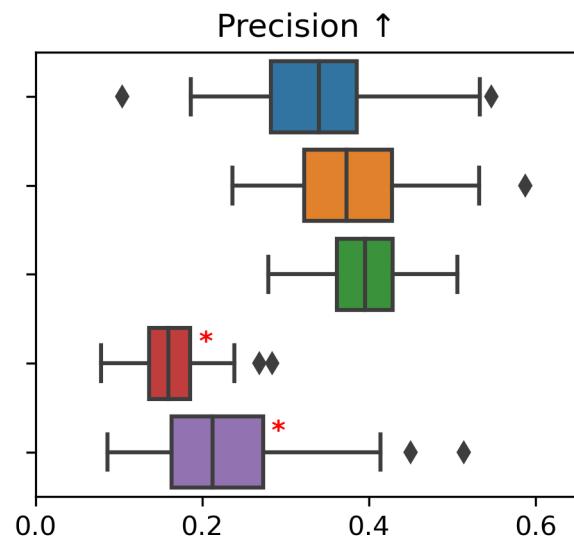
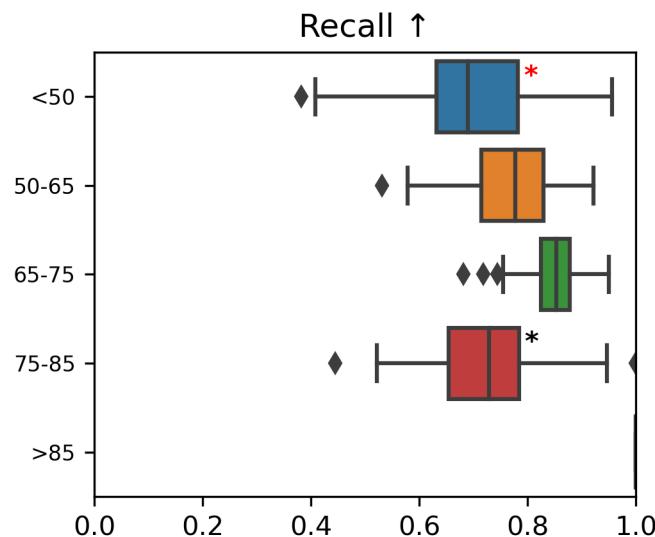
Precision / recall curve

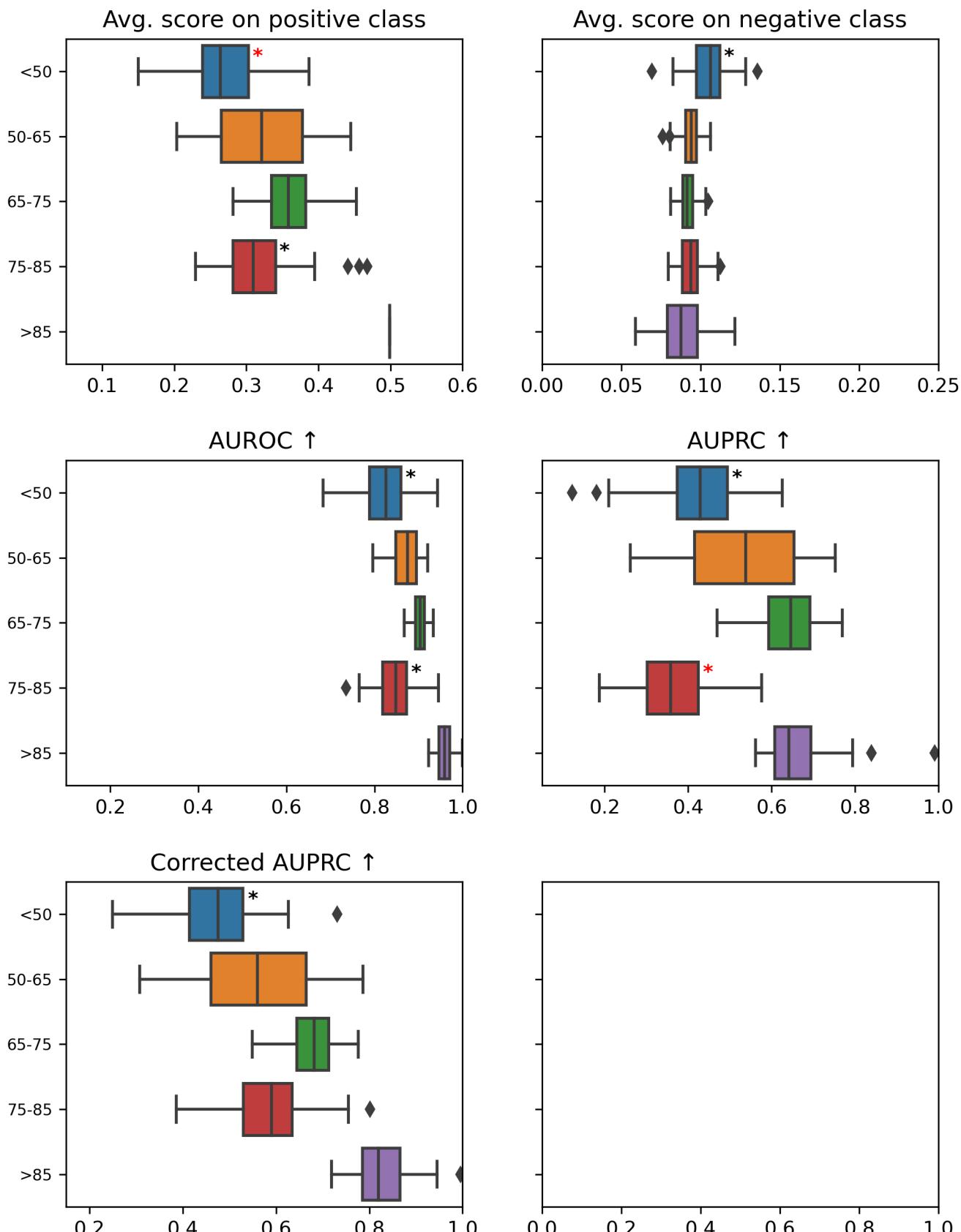


Corrected precision / recall curve



## Grouping by age\_group

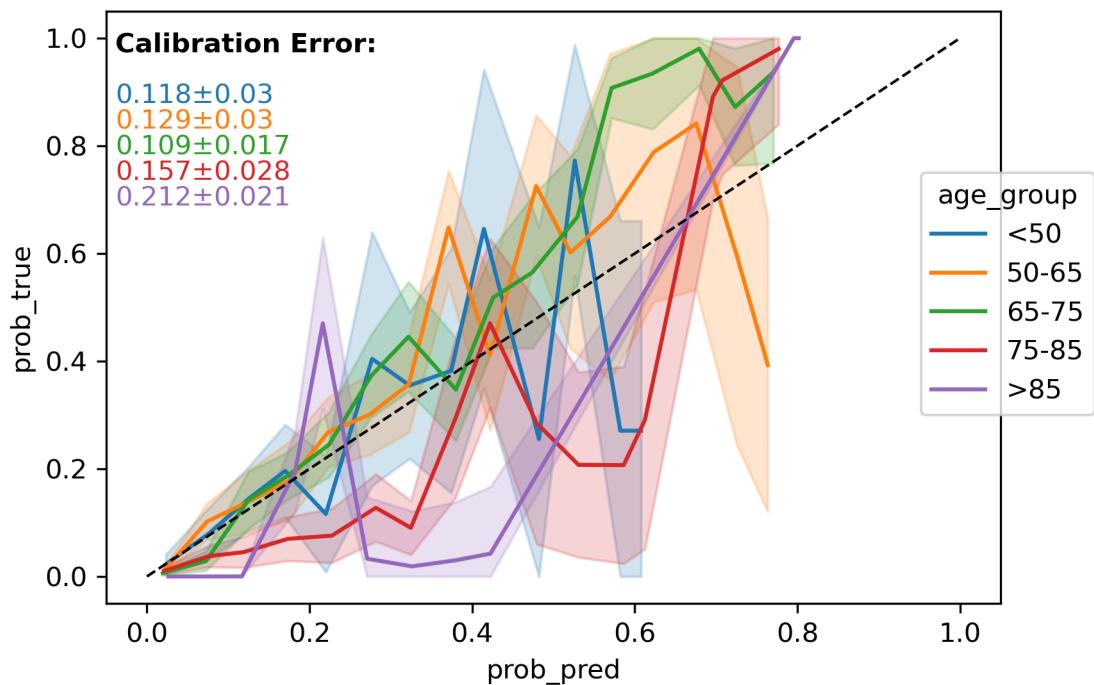




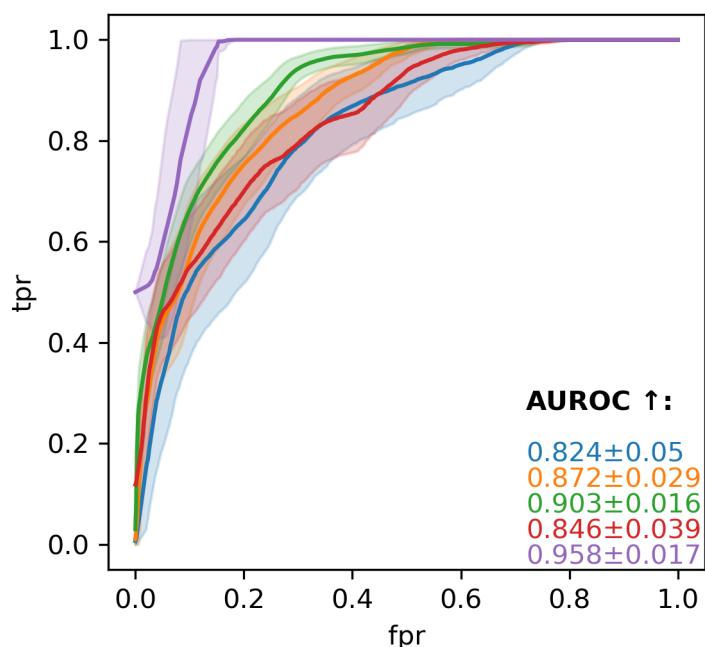
Metric	Group name	Group vs. rest	P-value	Delta
Recall ↑	<50	worse	5.59e-16	0.124
Recall ↑	65-75	better	9.79e-24	0.09
Recall ↑	75-85	worse	1.10e-12	0.08
Recall ↑	>85	better	2.82e-39	0.202
Precision ↑	50-65	better	4.03e-10	0.048
Precision ↑	65-75	better	1.27e-25	0.088

Precision ↑	75-85	worse	1.28e-34	0.219
Precision ↑	>85	worse	1.68e-24	0.13
NPV ↑	<50	worse	1.72e-23	0.033
NPV ↑	50-65	worse	2.23e-22	0.017
NPV ↑	65-75	better	3.29e-06	0.006
NPV ↑	75-85	better	6.24e-30	0.017
NPV ↑	>85	better	2.82e-39	0.035
FPR ↓	<50	worse	6.28e-06	0.016
FPR ↓	>85	better	3.81e-13	0.043
Corrected precision ↑	50-65	better	6.57e-06	0.033
Corrected precision ↑	65-75	better	1.25e-21	0.082
Corrected precision ↑	>85	better	9.95e-33	0.178
Corrected NPV ↑	<50	better	7.65e-24	0.011
Corrected NPV ↑	50-65	better	1.15e-29	0.015
Corrected NPV ↑	65-75	better	1.28e-34	0.026
Corrected NPV ↑	75-85	better	3.92e-24	0.014
Corrected NPV ↑	>85	better	2.82e-39	0.029
Avg. score on positive class	<50	worse	2.48e-23	0.079
Avg. score on positive class	65-75	better	1.21e-10	0.041
Avg. score on positive class	75-85	worse	5.29e-07	0.028
Avg. score on positive class	>85	better	2.82e-39	0.166
Avg. score on negative class	<50	worse	3.63e-19	0.013
Avg. score on negative class	65-75	better	1.88e-06	0.003
AUROC ↑	<50	worse	2.46e-21	0.064
AUROC ↑	65-75	better	6.84e-26	0.036
AUROC ↑	75-85	worse	2.72e-16	0.04
AUROC ↑	>85	better	1.28e-34	0.076
AUPRC ↑	<50	worse	6.44e-20	0.13
AUPRC ↑	65-75	better	1.60e-23	0.166
AUPRC ↑	75-85	worse	1.88e-30	0.222
AUPRC ↑	>85	better	7.83e-22	0.092
Corrected AUPRC ↑	<50	worse	4.4e-20	0.12
Corrected AUPRC ↑	65-75	better	2.67e-23	0.134
Corrected AUPRC ↑	>85	better	1.28e-34	0.239

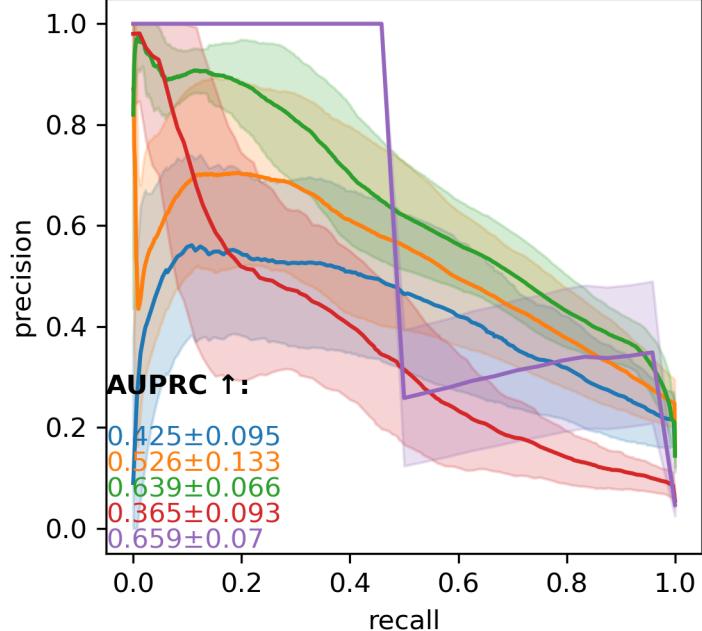
Calibration curve



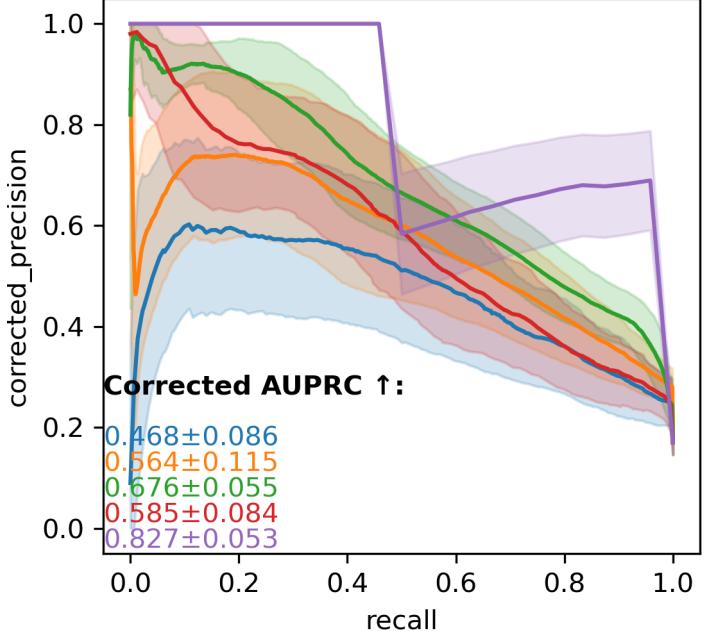
ROC curve



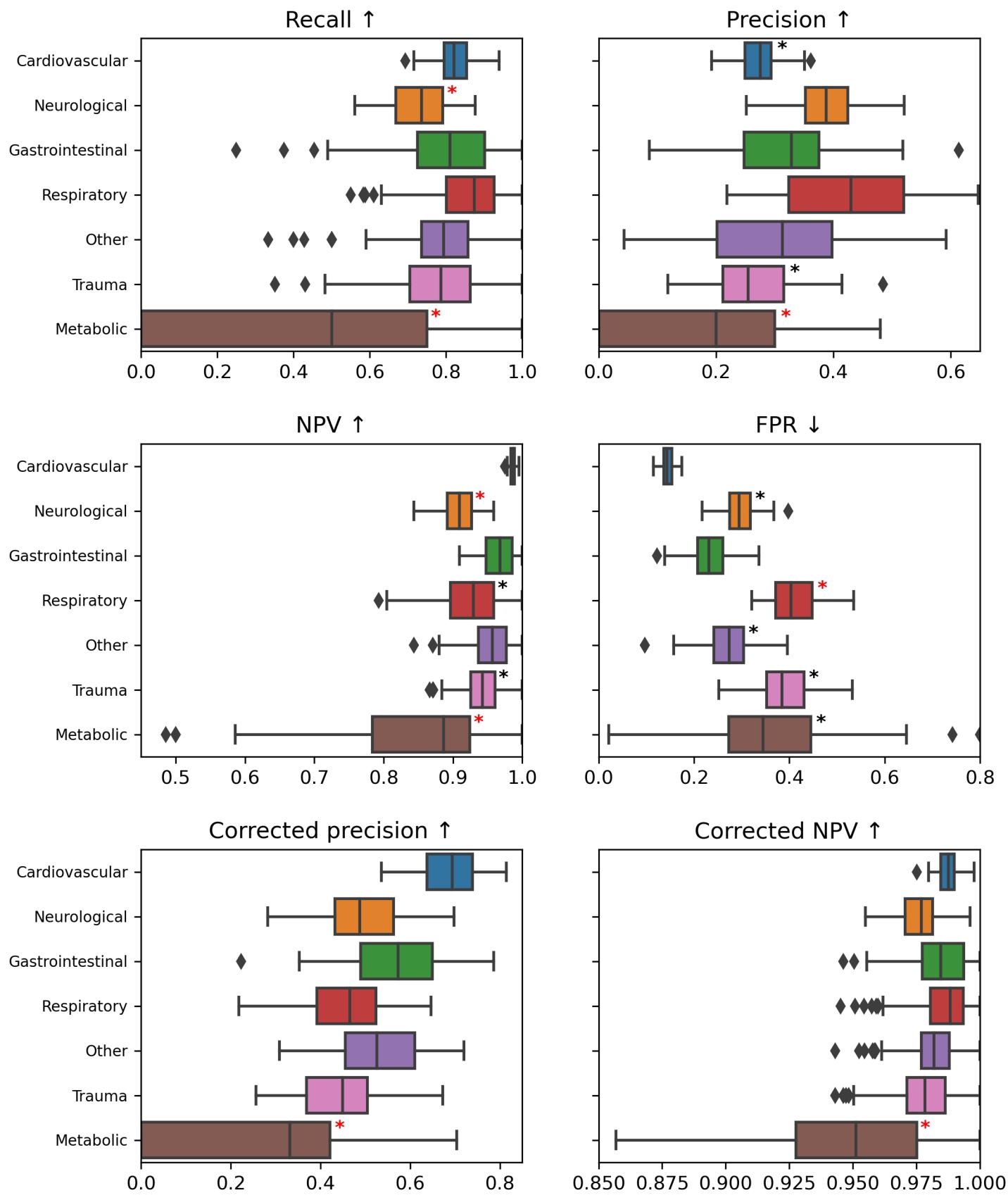
Precision / recall curve



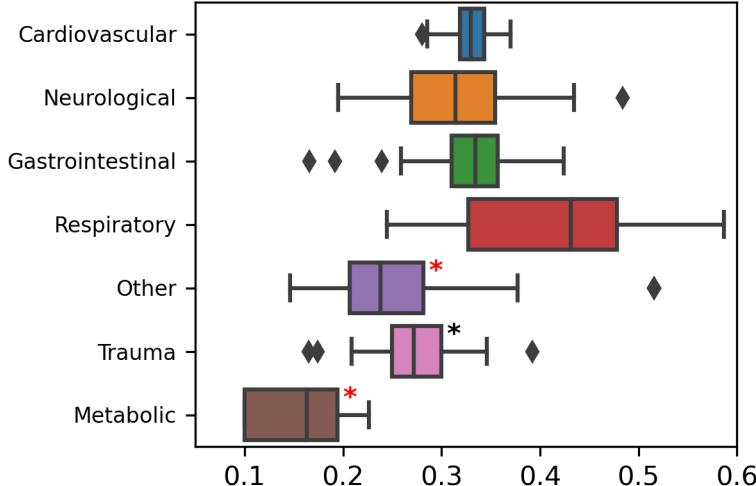
Corrected precision / recall curve



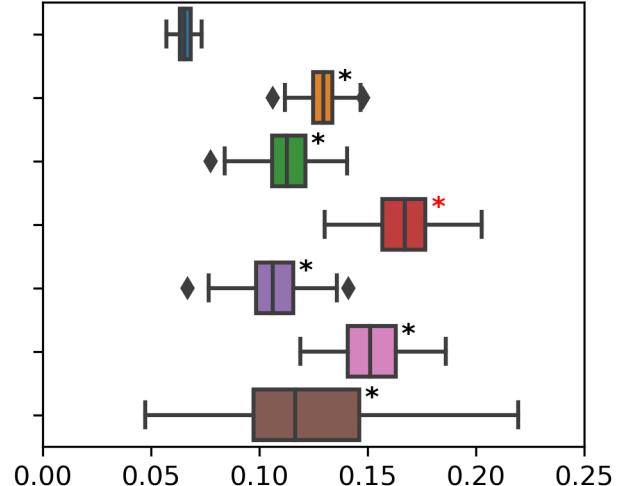
## Grouping by APACHE\_group



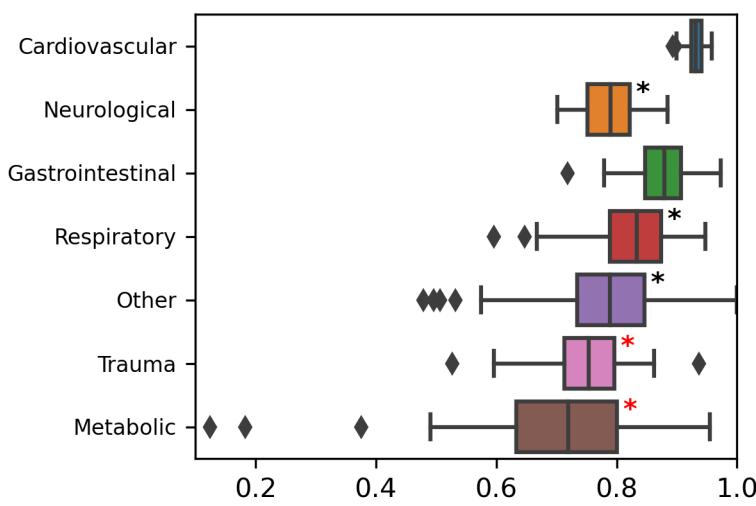
Avg. score on positive class



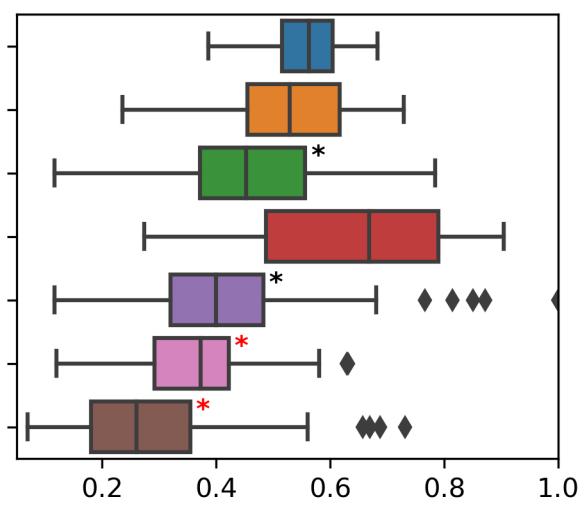
Avg. score on negative class



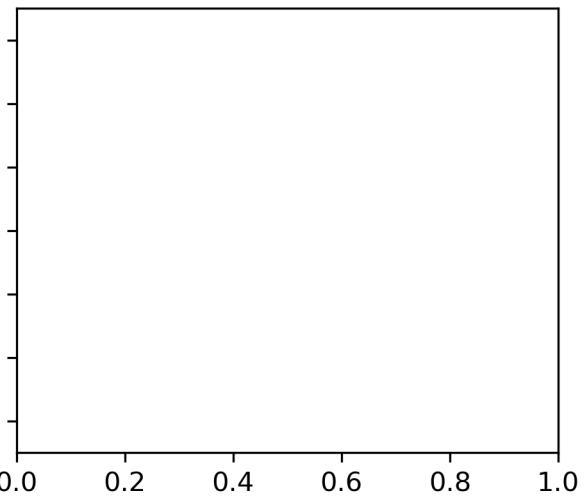
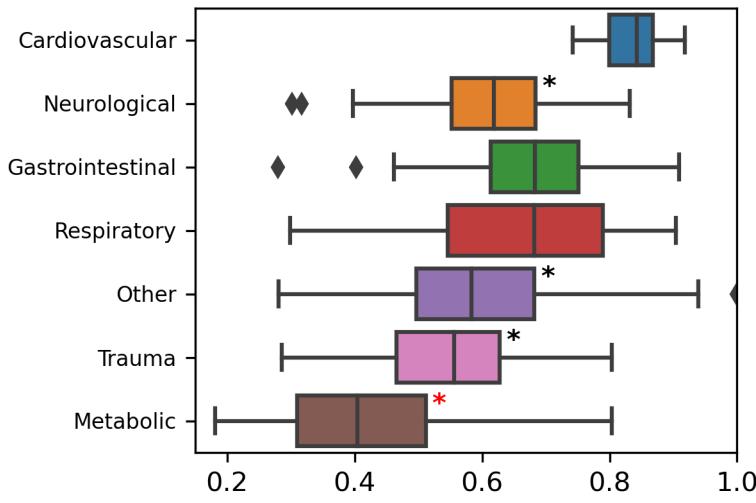
AUROC ↑



AUPRC ↑



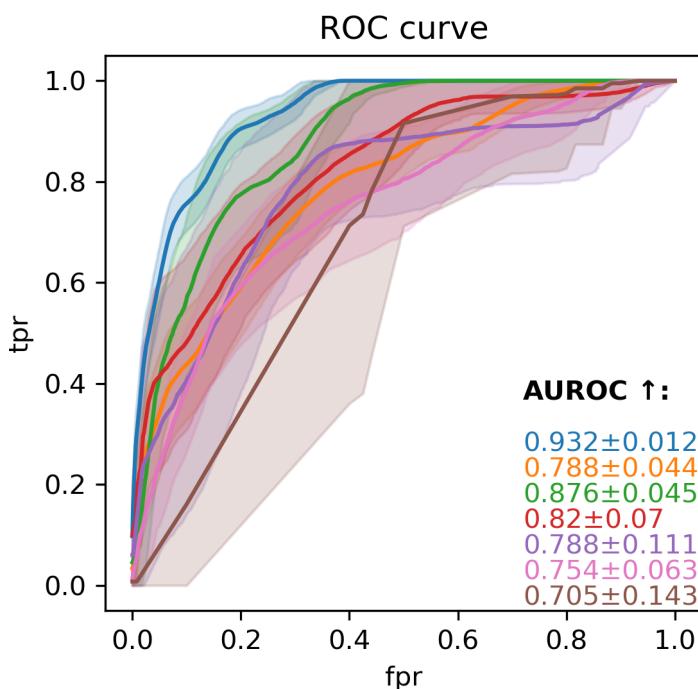
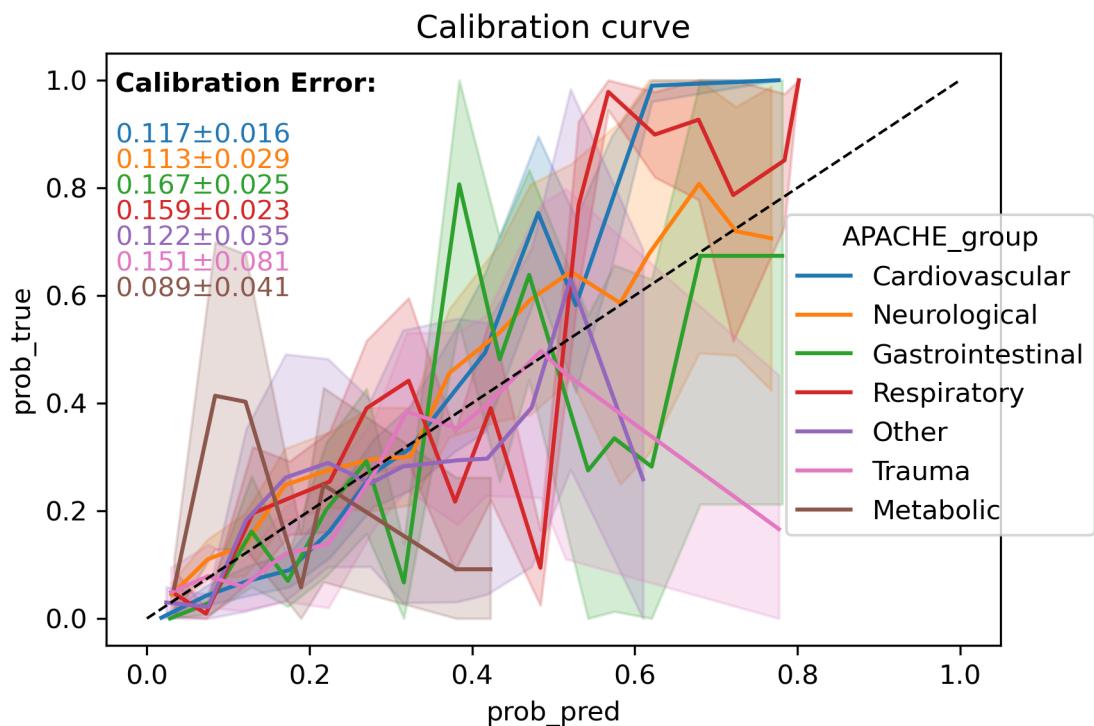
Corrected AUPRC ↑



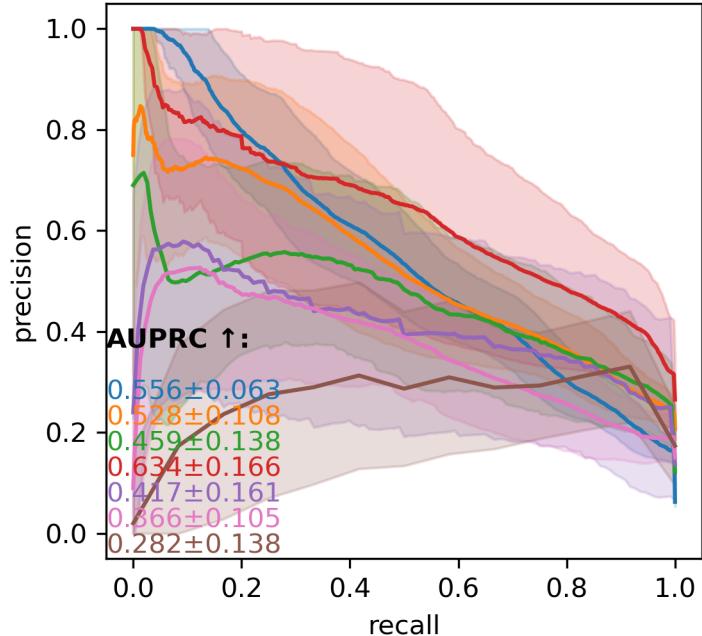
Metric	Group name	Group vs. rest	P-value	Delta
Recall ↑	Cardiovascular	better	3.68e-07	0.038
Recall ↑	Neurological	worse	2.87e-20	0.096
Recall ↑	Respiratory	better	7.15e-12	0.09
Recall ↑	Metabolic	worse	4.16e-12	0.306
Precision ↑	Cardiovascular	worse	4.47e-30	0.102
Precision ↑	Neurological	better	4.87e-16	0.068

Precision ↑	Respiratory	better	1.23e-10	0.109
Precision ↑	Trauma	worse	6.70e-19	0.093
Precision ↑	Metabolic	worse	4.69e-19	0.141
NPV ↑	Cardiovascular	better	1.28e-34	0.054
NPV ↑	Neurological	worse	1.28e-34	0.066
NPV ↑	Respiratory	worse	3.84e-14	0.038
NPV ↑	Trauma	worse	9.66e-14	0.024
NPV ↑	Metabolic	worse	6.76e-13	0.08
FPR ↓	Cardiovascular	better	1.28e-34	0.171
FPR ↓	Neurological	worse	2.26e-34	0.096
FPR ↓	Respiratory	worse	1.28e-34	0.205
FPR ↓	Other	worse	2.e-21	0.062
FPR ↓	Trauma	worse	1.28e-34	0.18
FPR ↓	Metabolic	worse	1.57e-19	0.132
Corrected precision ↑	Cardiovascular	better	1.28e-34	0.316
Corrected precision ↑	Gastrointestinal	better	1.93e-14	0.101
Corrected precision ↑	Other	better	4.53e-06	0.05
Corrected precision ↑	Metabolic	worse	3.71e-16	0.142
Corrected NPV ↑	Cardiovascular	better	1.95e-34	0.021
Corrected NPV ↑	Gastrointestinal	better	2.02e-17	0.012
Corrected NPV ↑	Respiratory	better	1.68e-17	0.018
Corrected NPV ↑	Other	better	4.49e-15	0.01
Corrected NPV ↑	Trauma	better	2.86e-06	0.005
Corrected NPV ↑	Metabolic	worse	1.48e-07	0.021
Avg. score on positive class	Respiratory	better	1.38e-12	0.116
Avg. score on positive class	Other	worse	1.7e-18	0.103
Avg. score on positive class	Trauma	worse	1.1e-24	0.069
Avg. score on positive class	Metabolic	worse	8.91e-35	0.176
Avg. score on negative class	Cardiovascular	better	1.28e-34	0.068
Avg. score on negative class	Neurological	worse	1.28e-34	0.043
Avg. score on negative class	Gastrointestinal	worse	1.18e-29	0.021
Avg. score on negative class	Respiratory	worse	1.28e-34	0.079
Avg. score on negative class	Other	worse	1.74e-18	0.013
Avg. score on negative class	Trauma	worse	1.28e-34	0.061
Avg. score on negative class	Metabolic	worse	3.01e-12	0.023
AUROC ↑	Cardiovascular	better	1.28e-34	0.114
AUROC ↑	Neurological	worse	2.48e-34	0.113
AUROC ↑	Respiratory	worse	2.24e-14	0.049
AUROC ↑	Other	worse	5.04e-17	0.097
AUROC ↑	Trauma	worse	2.91e-33	0.136
AUROC ↑	Metabolic	worse	4.76e-26	0.167
AUPRC ↑	Gastrointestinal	worse	3.79e-09	0.098
AUPRC ↑	Respiratory	better	3.01e-07	0.16

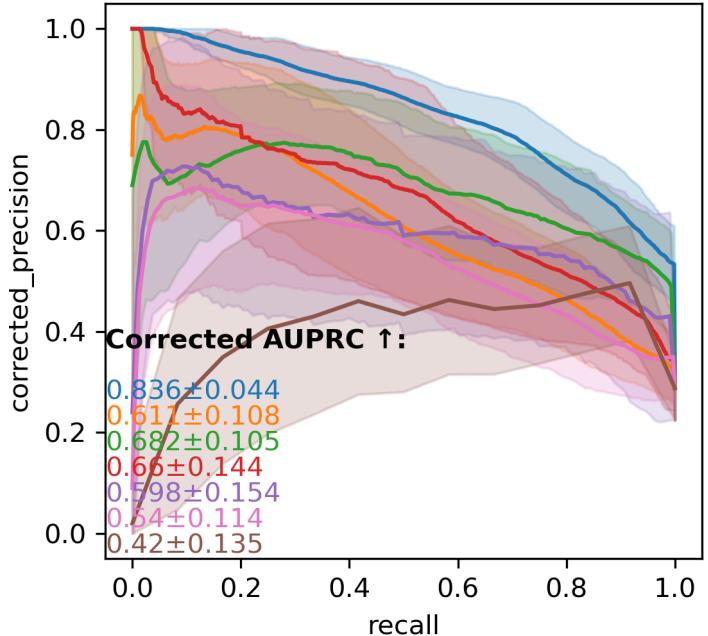
AUPRC ↑	Other	worse	1.56e-15	0.159
AUPRC ↑	Trauma	worse	1.56e-26	0.189
AUPRC ↑	Metabolic	worse	1.24e-27	0.294
Corrected AUPRC ↑	Cardiovascular	better	1.28e-34	0.292
Corrected AUPRC ↑	Neurological	worse	4.92e-09	0.087
Corrected AUPRC ↑	Other	worse	1.23e-06	0.085
Corrected AUPRC ↑	Trauma	worse	3.91e-16	0.118
Corrected AUPRC ↑	Metabolic	worse	2.77e-26	0.262



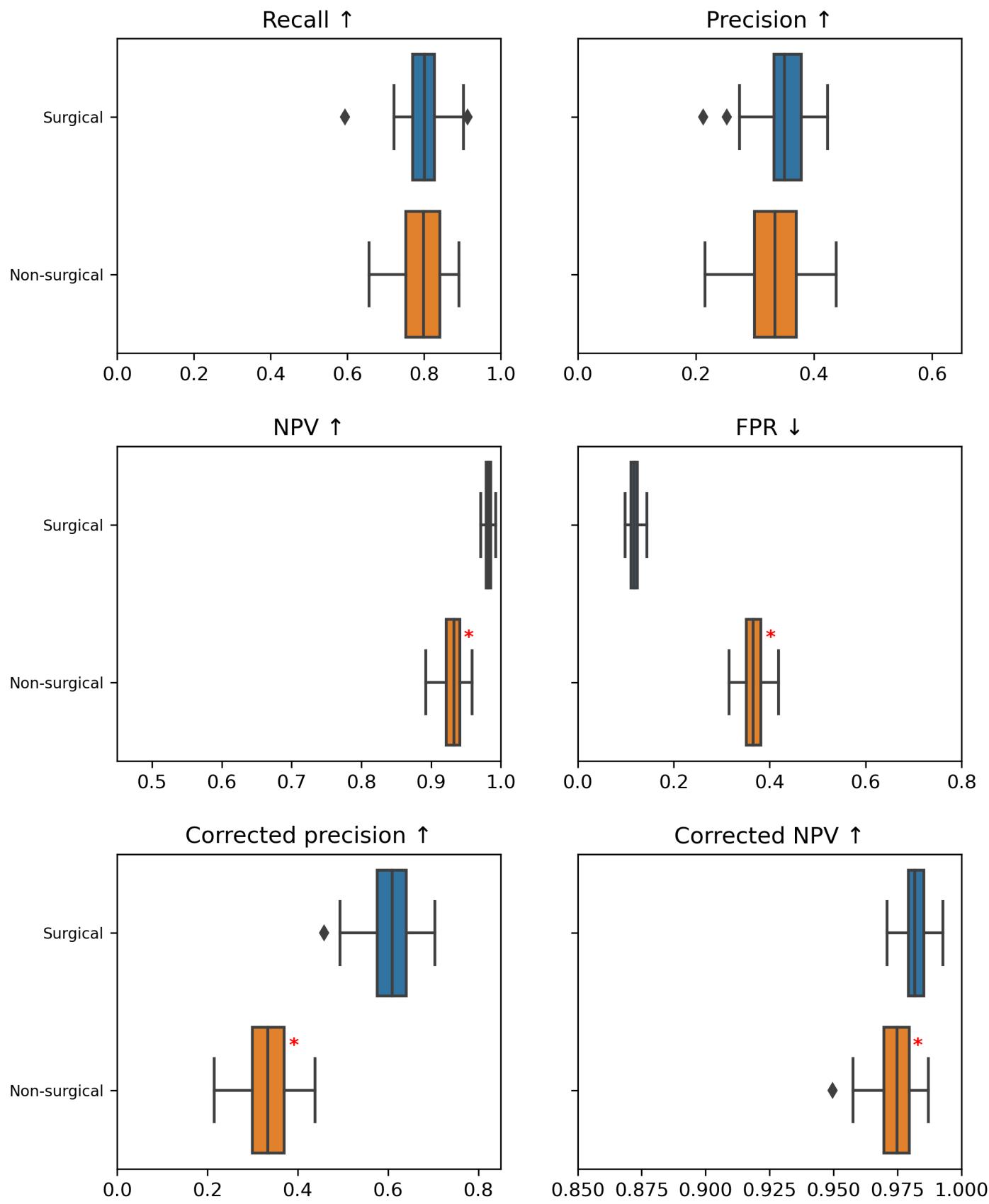
Precision / recall curve

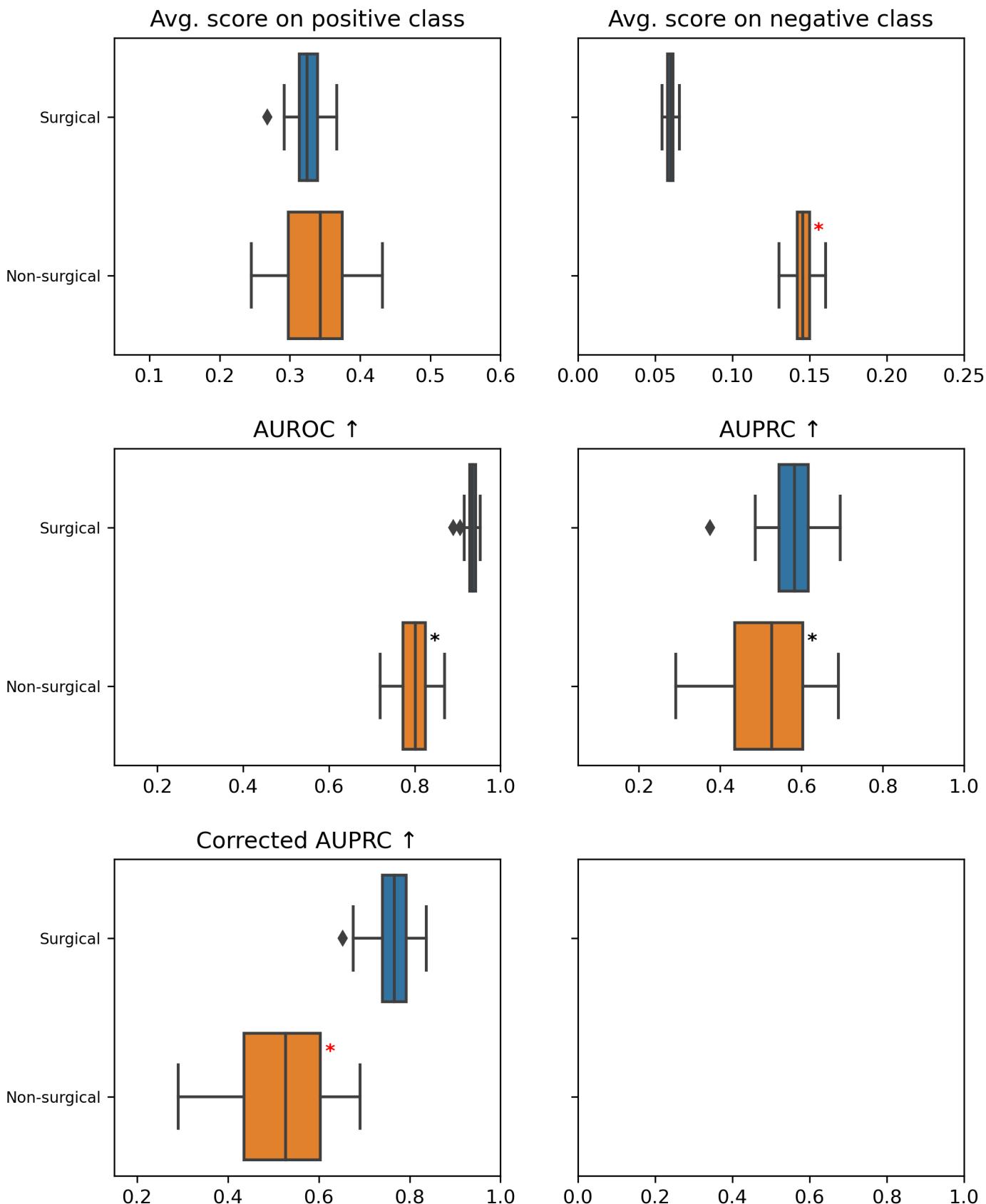


Corrected precision / recall curve



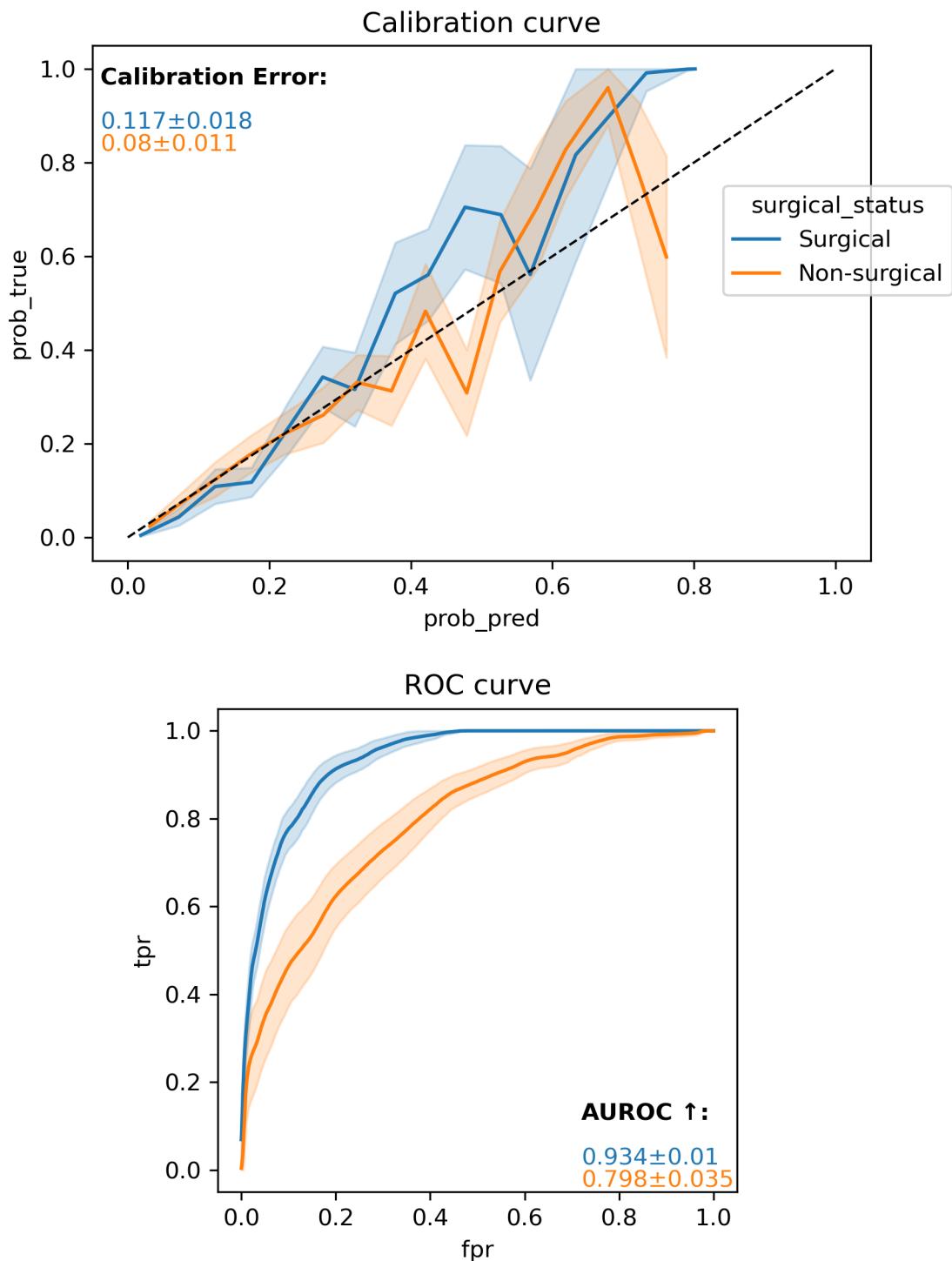
## Grouping by surgical\_status



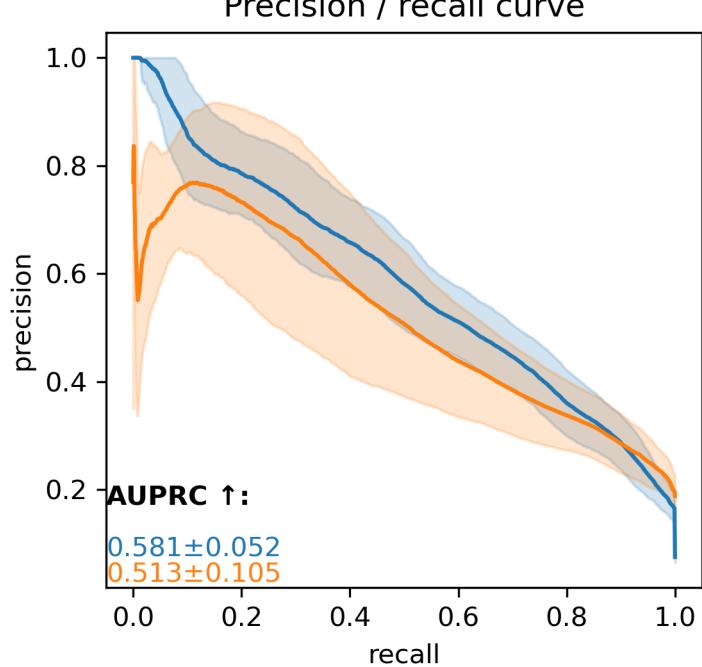


Metric	Group with worst dist.	P-value	Delta
NPV ↑	Non-surgical	1.28e-34	0.049
FPR ↓	Non-surgical	1.28e-34	0.248
Corrected precision ↑	Non-surgical	1.28e-34	0.275
Corrected NPV ↑	Non-surgical	1.03e-14	0.007
Avg. score on negative class	Non-surgical	1.28e-34	0.085
AUROC ↑	Non-surgical	1.28e-34	0.132

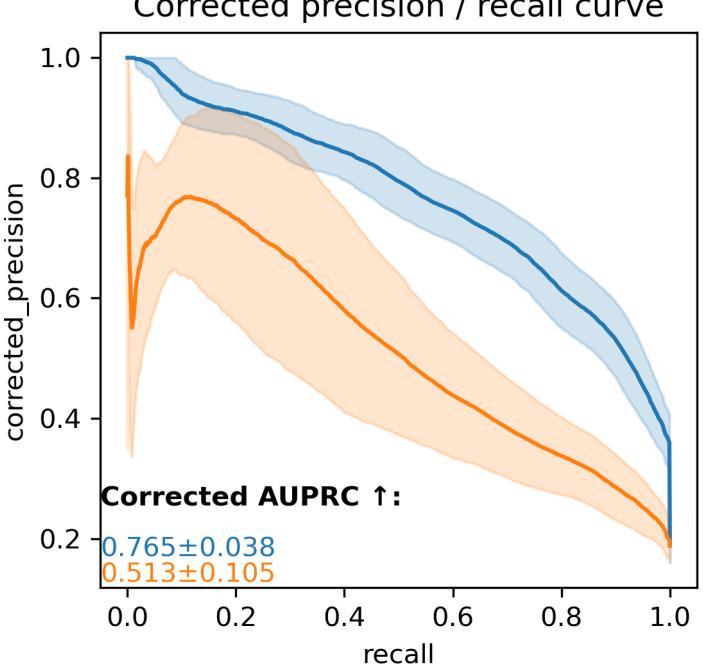
AUPRC ↑	Non-surgical	1.52e-06	0.056
Corrected AUPRC ↑	Non-surgical	2.13e-34	0.24



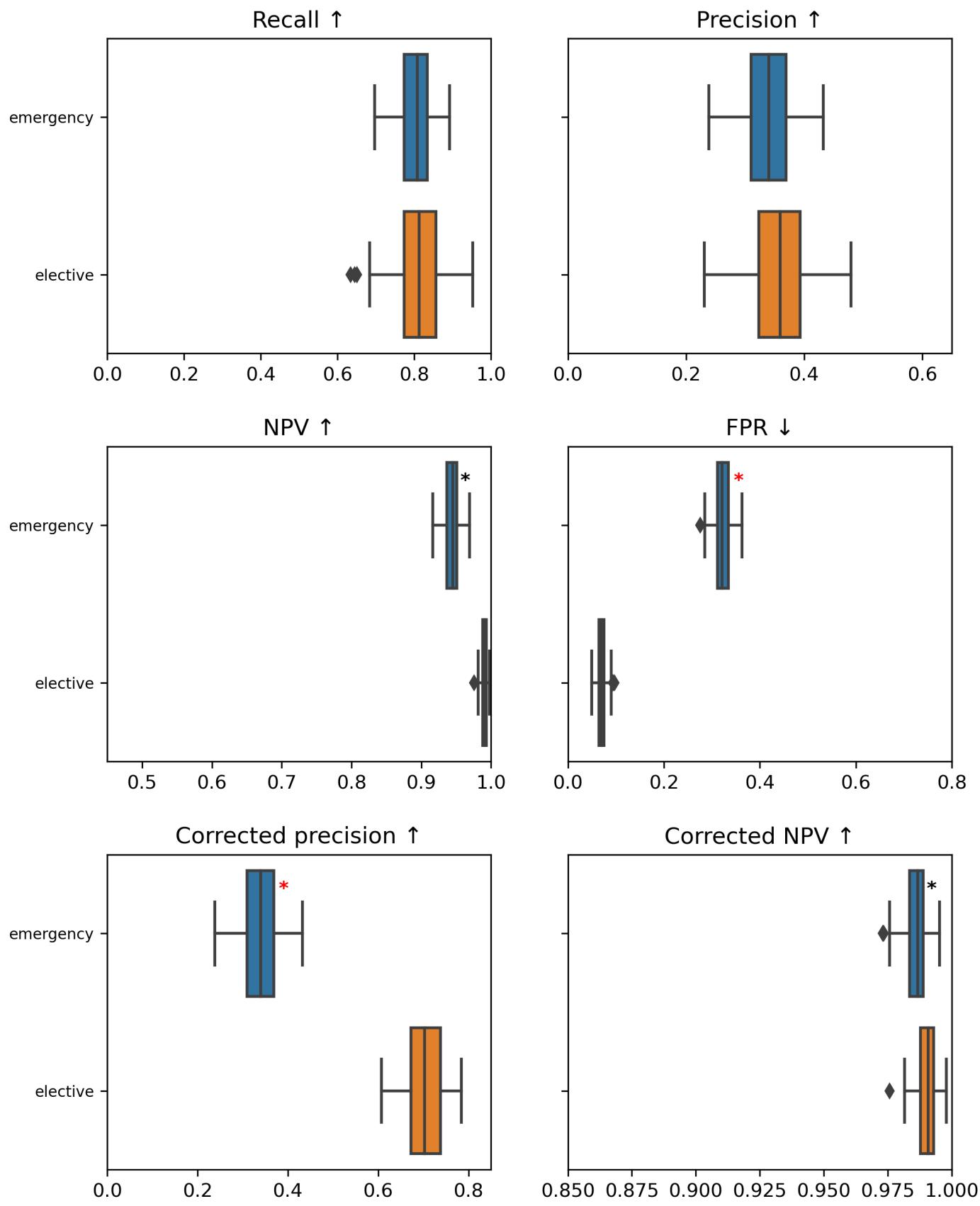
Precision / recall curve

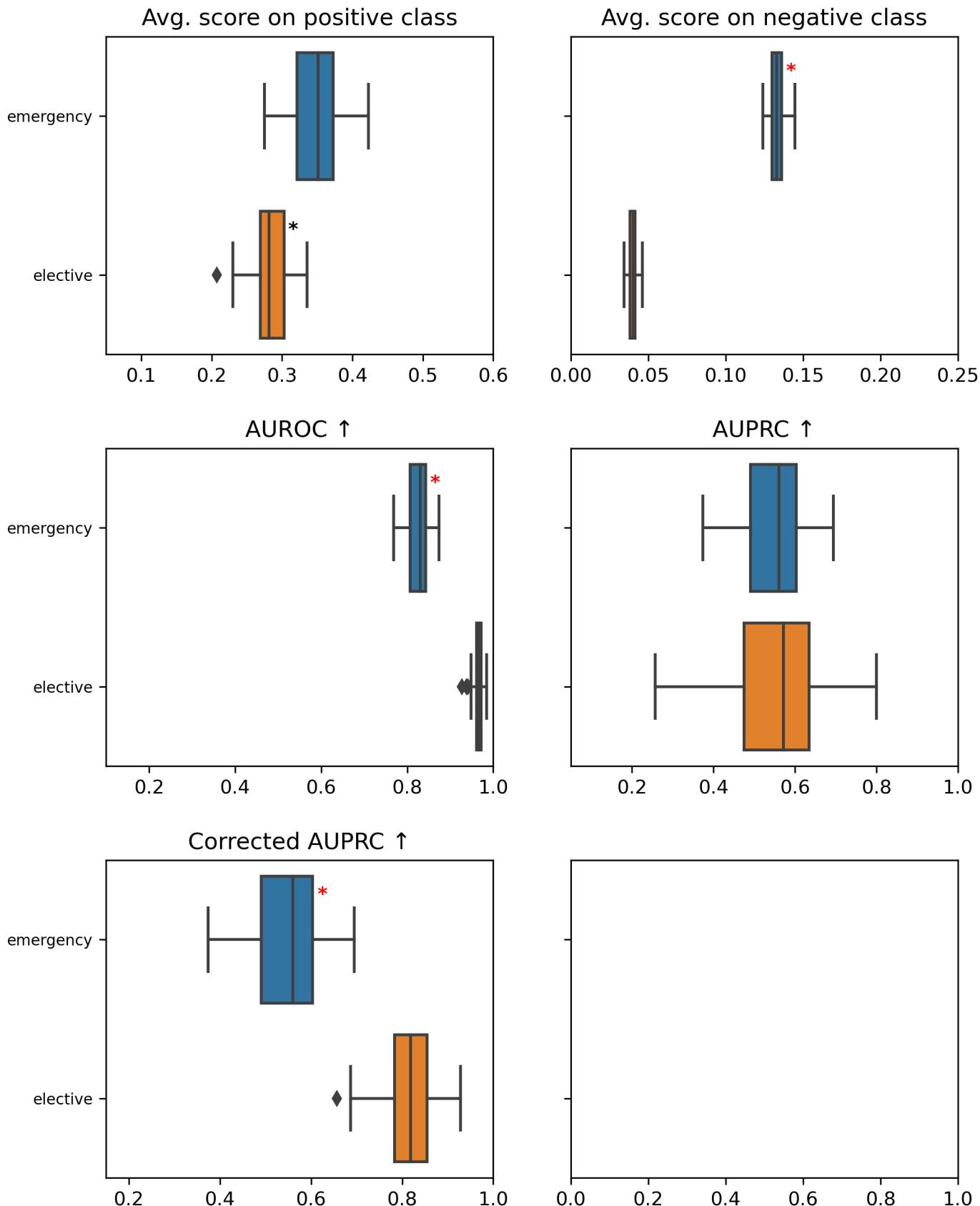


Corrected precision / recall curve



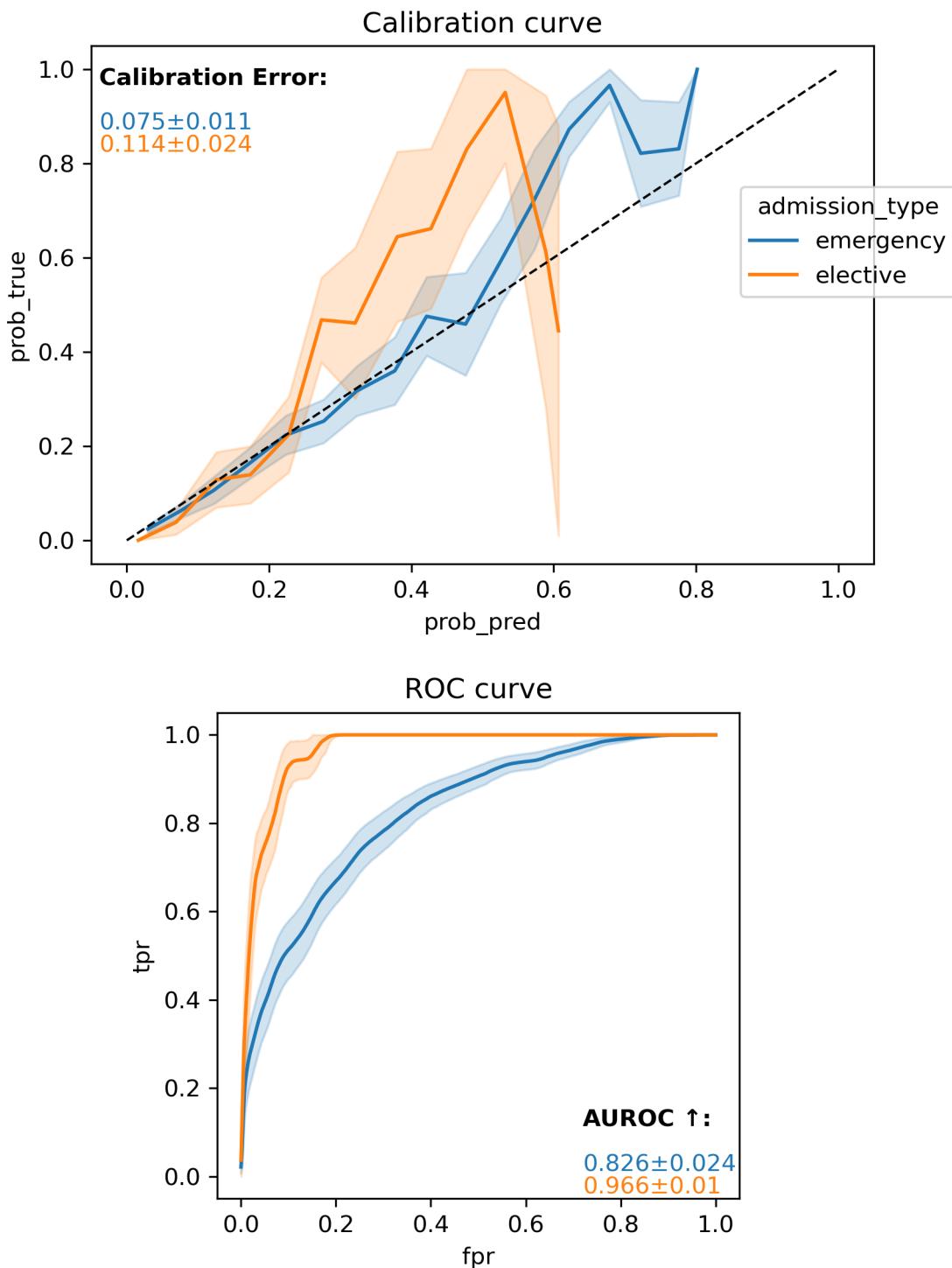
## Grouping by admission\_type



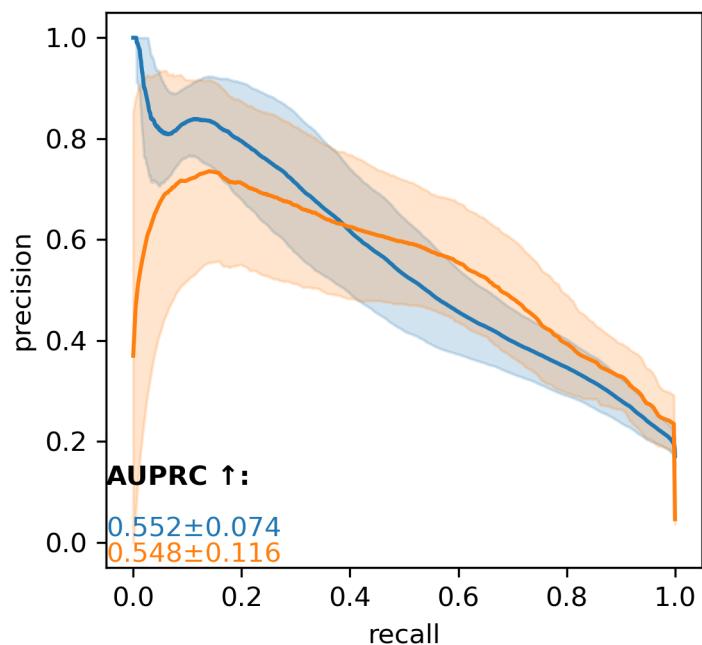


Metric	Group with worst dist.	P-value	Delta
NPV ↑	emergency	1.28e-34	0.046
FPR ↓	emergency	1.28e-34	0.25
Corrected precision ↑	emergency	1.28e-34	0.363
Corrected NPV ↑	emergency	3.34e-11	0.004
Avg. score on positive class	elective	9.24e-27	0.069
Avg. score on negative class	emergency	1.28e-34	0.093

AUROC ↑	emergency	1.28e-34	0.137
Corrected AUPRC ↑	emergency	1.73e-34	0.258



Precision / recall curve



Corrected precision / recall curve

