

HOSEA Aim I – Results (MICE)

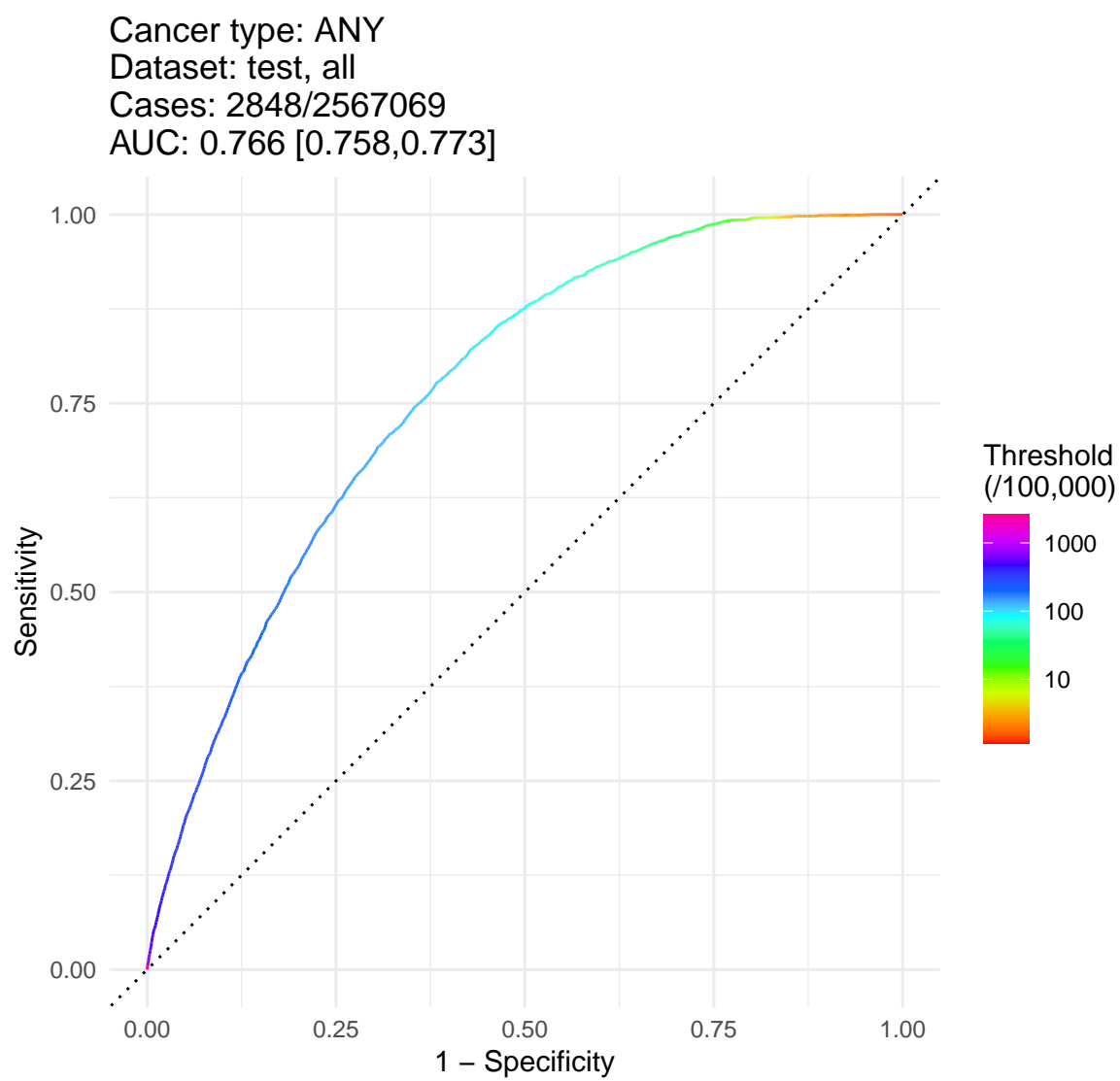
Simon Fontaine

November 8, 2022

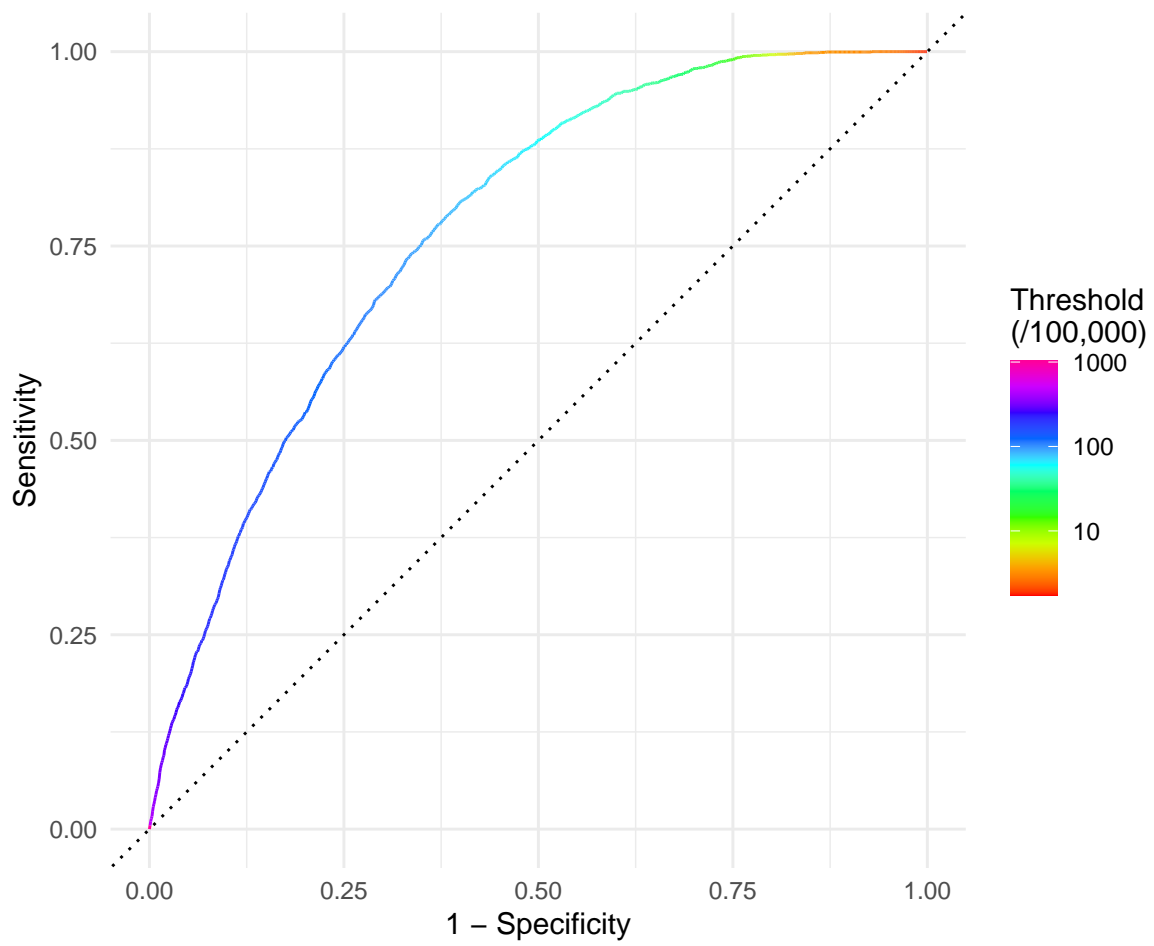
Contents

1	Overall performance	2
2	Comparison to HUNT, Kunzmann and Guidelines	5
2.1	Full test data	5
2.2	Complete test patients	8
2.3	Representative sample	11
3	Calibration	14
4	Thresholds	17
5	Stratification by identity groups	21
5.1	Age	21
5.2	Sex	24
5.3	Race	27
6	Cancer stage	30
7	Variable importance	31
7.1	Gain VI	31
7.2	SHAP VI	34
8	Years prior	37
9	SHAP correlation and imputation model	40

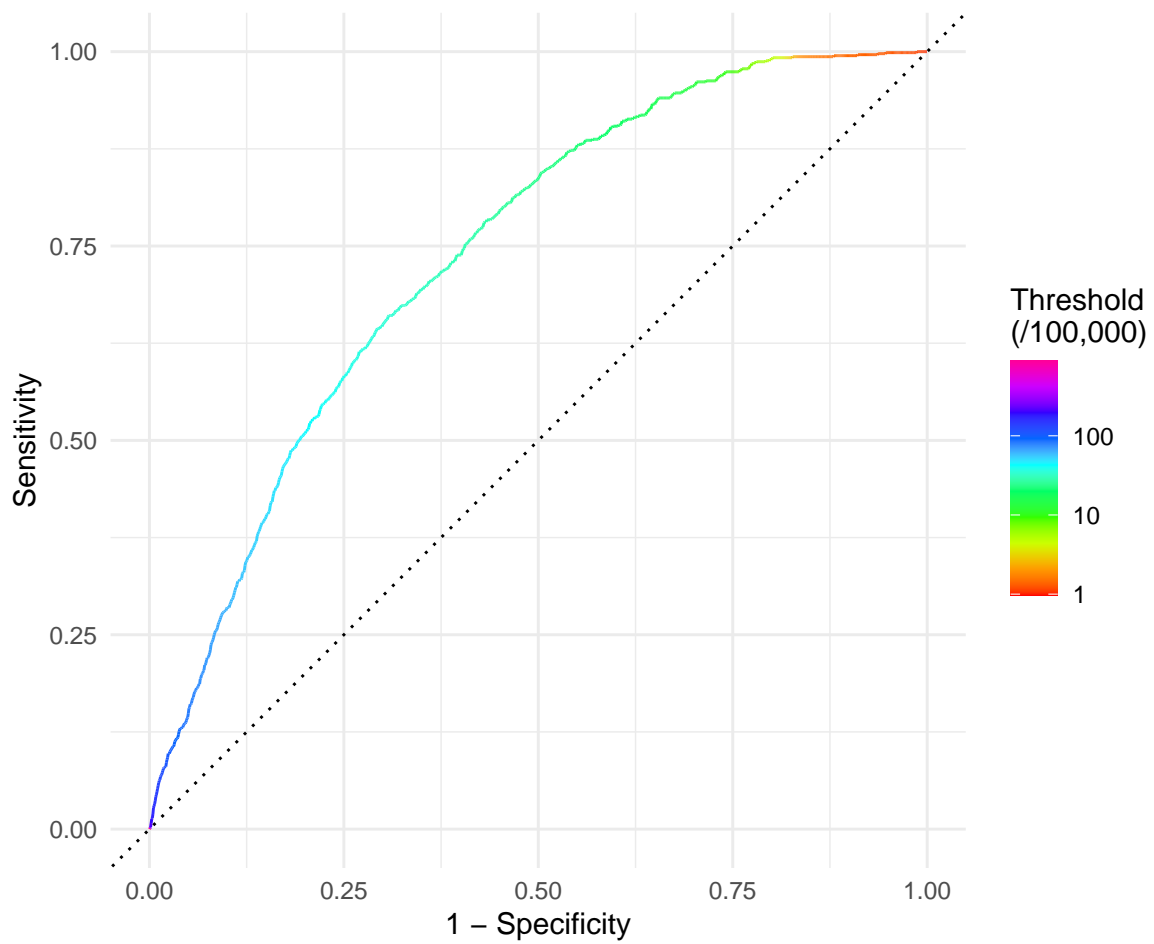
1 Overall performance



Cancer type: EAC
Dataset: test, all
Cases: 2076/2567069
AUC: 0.772 [0.763,0.780]

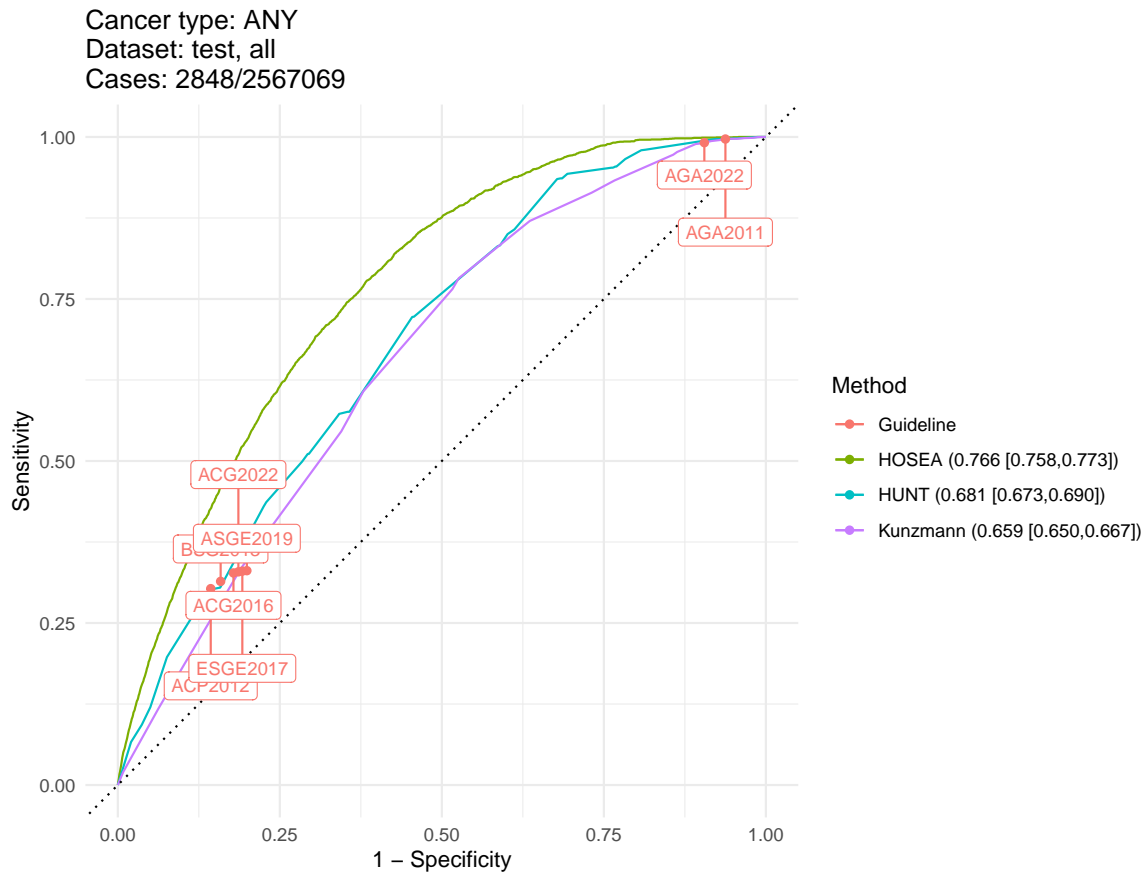


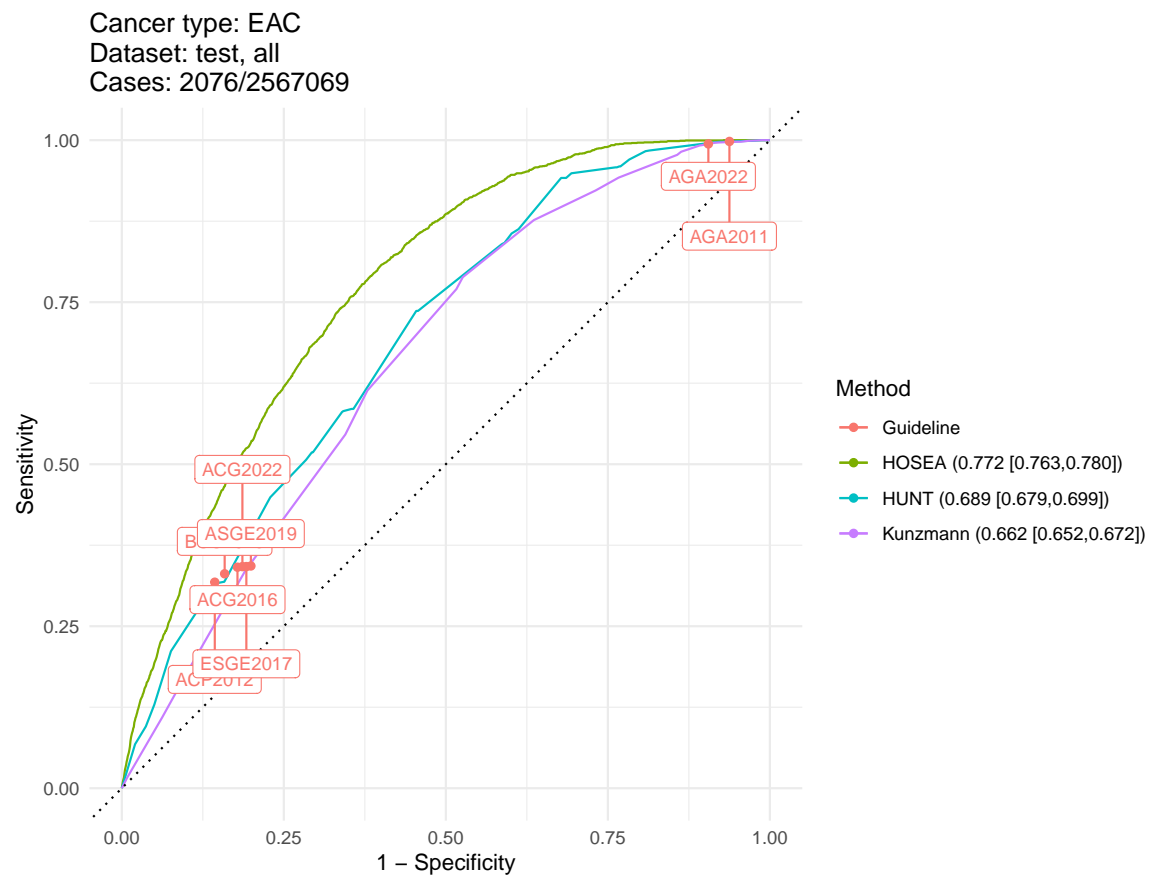
Cancer type: EGJAC
Dataset: test, all
Cases: 772/2567069
AUC: 0.740 [0.725,0.755]



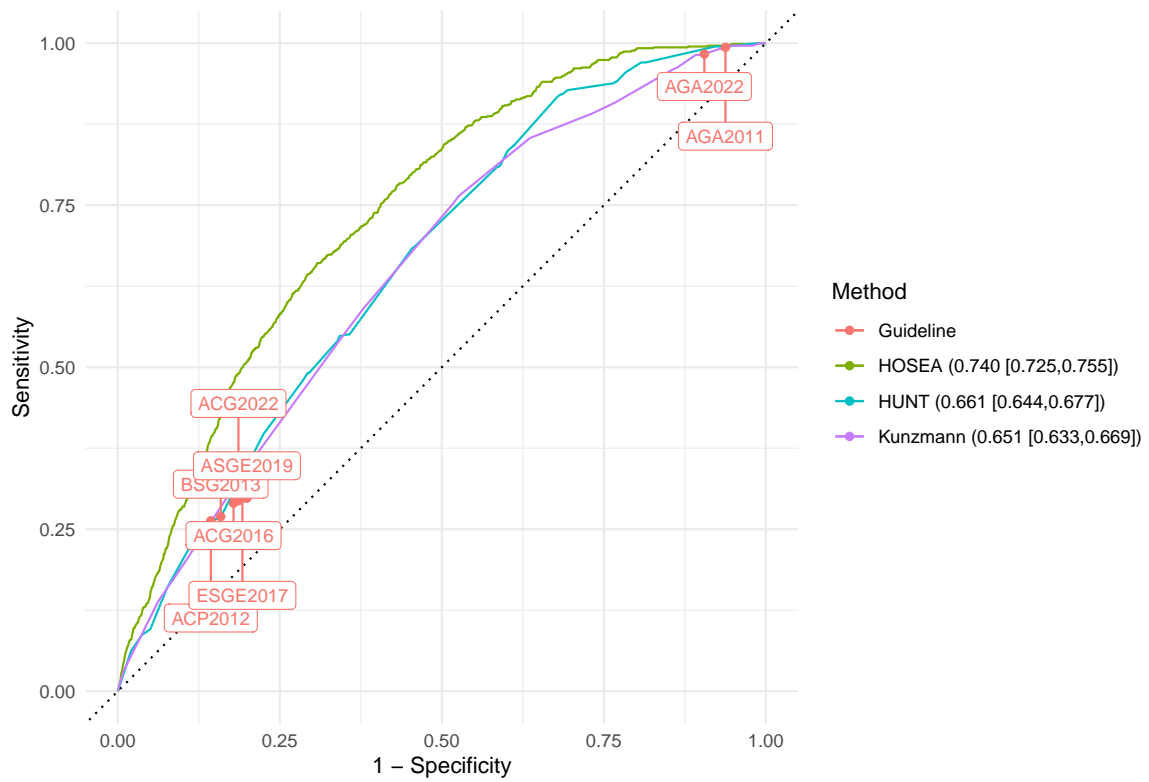
2 Comparison to HUNT, Kunzmann and Guidelines

2.1 Full test data



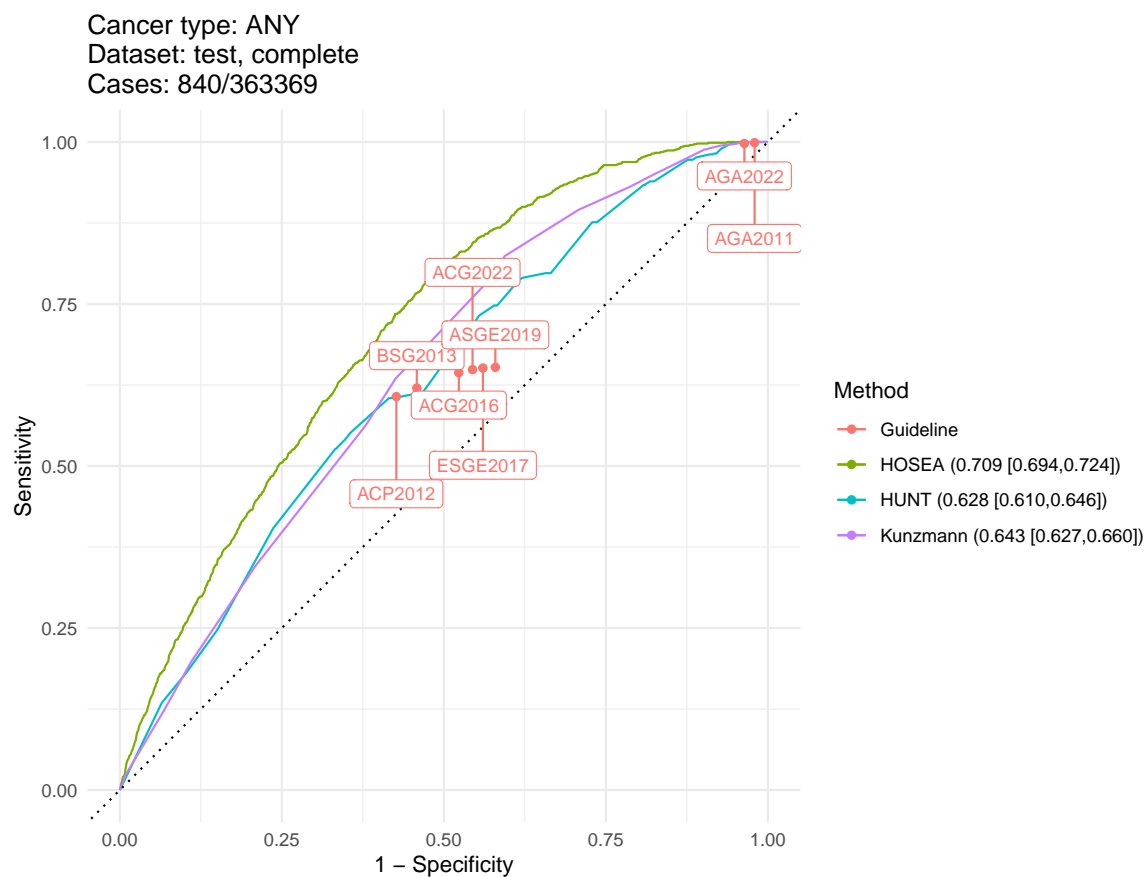


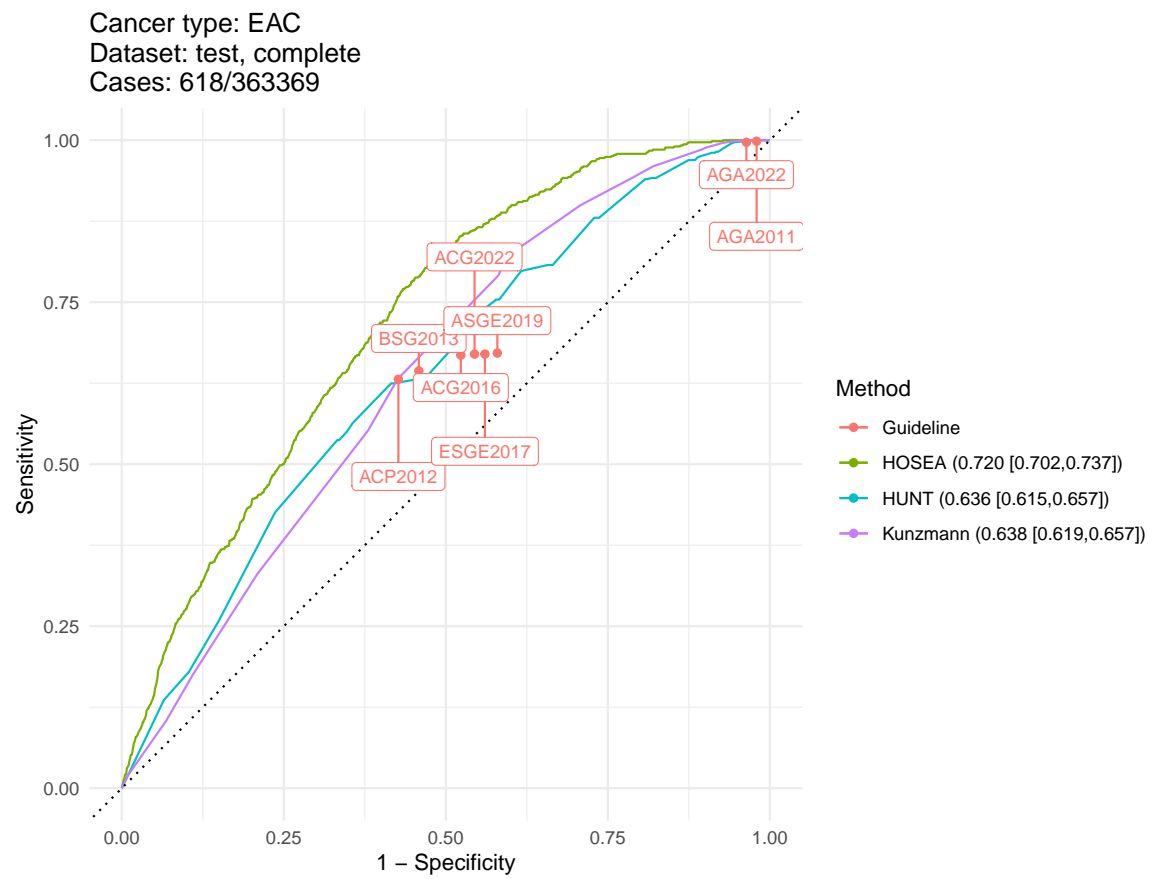
Cancer type: EGJAC
 Dataset: test, all
 Cases: 772/2567069



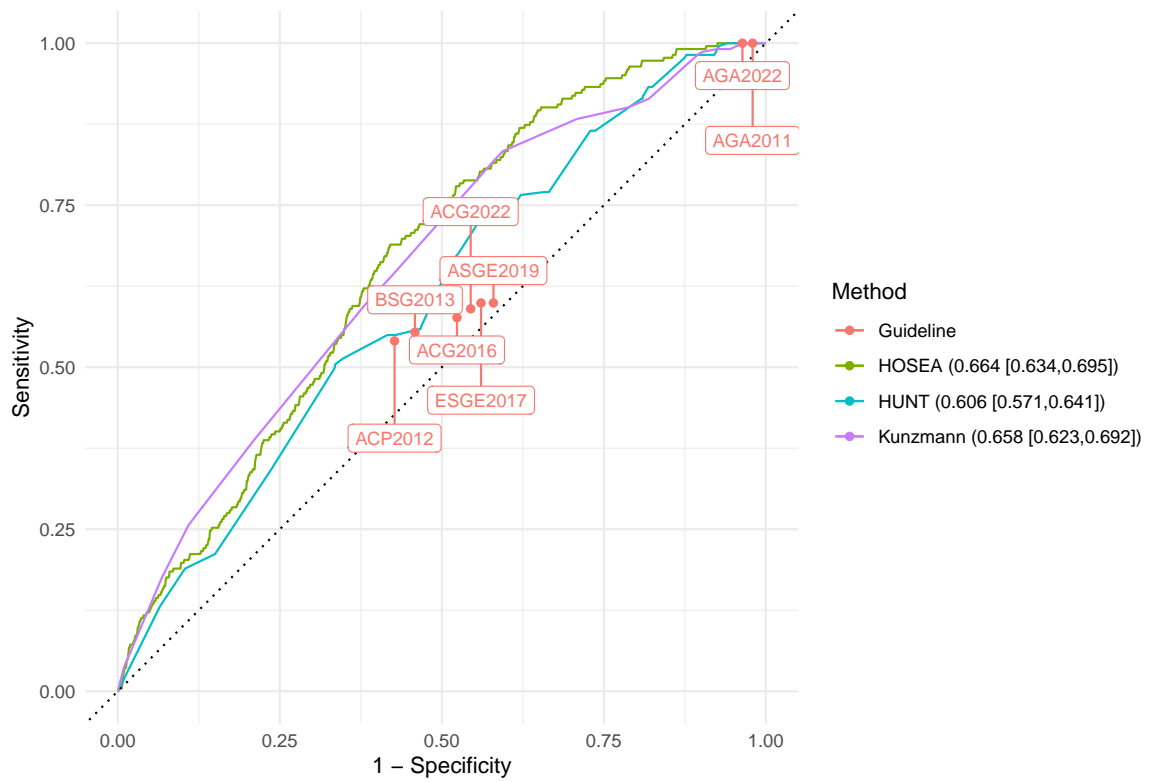
2.2 Complete test patients

w.r.t. HUNT, Kunzmann, & Guidelines (i.e., requires age, sex, GERD, bmi, smoking, etc.)



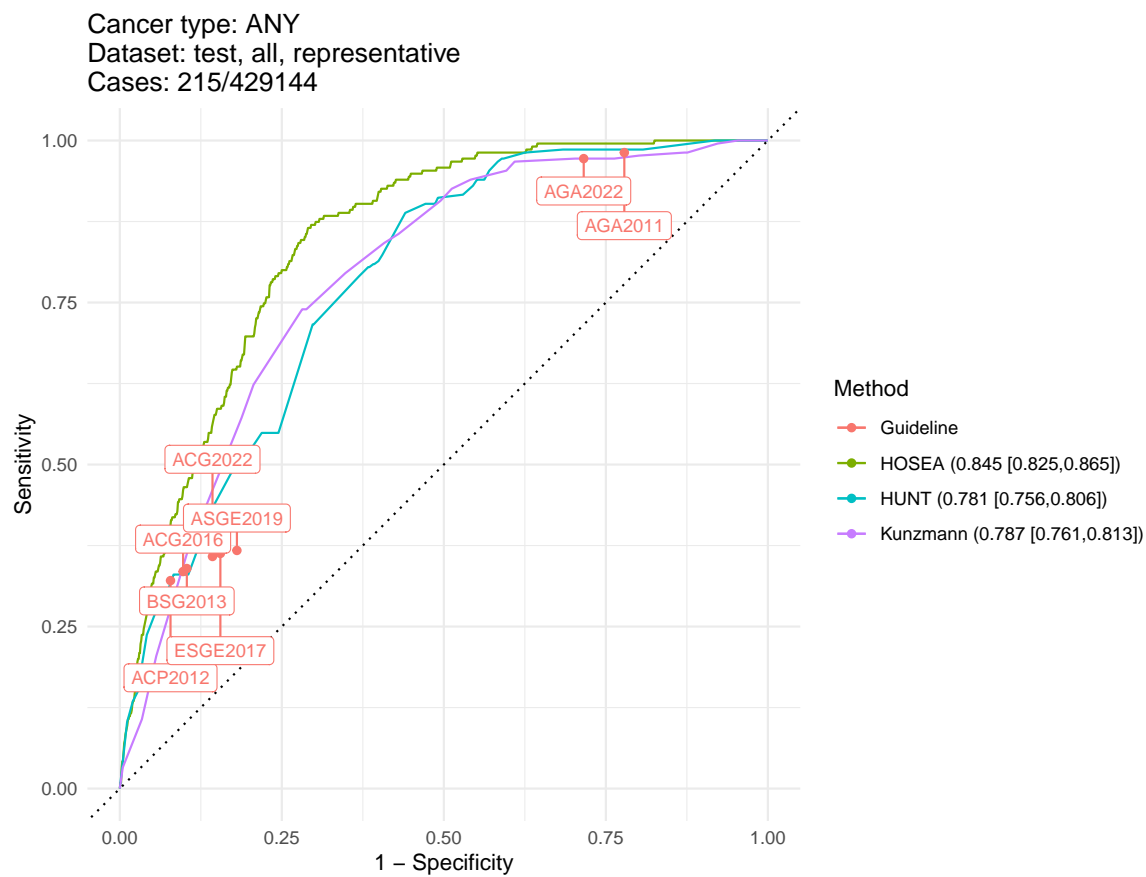


Cancer type: EGJAC
 Dataset: test, complete
 Cases: 222/363369

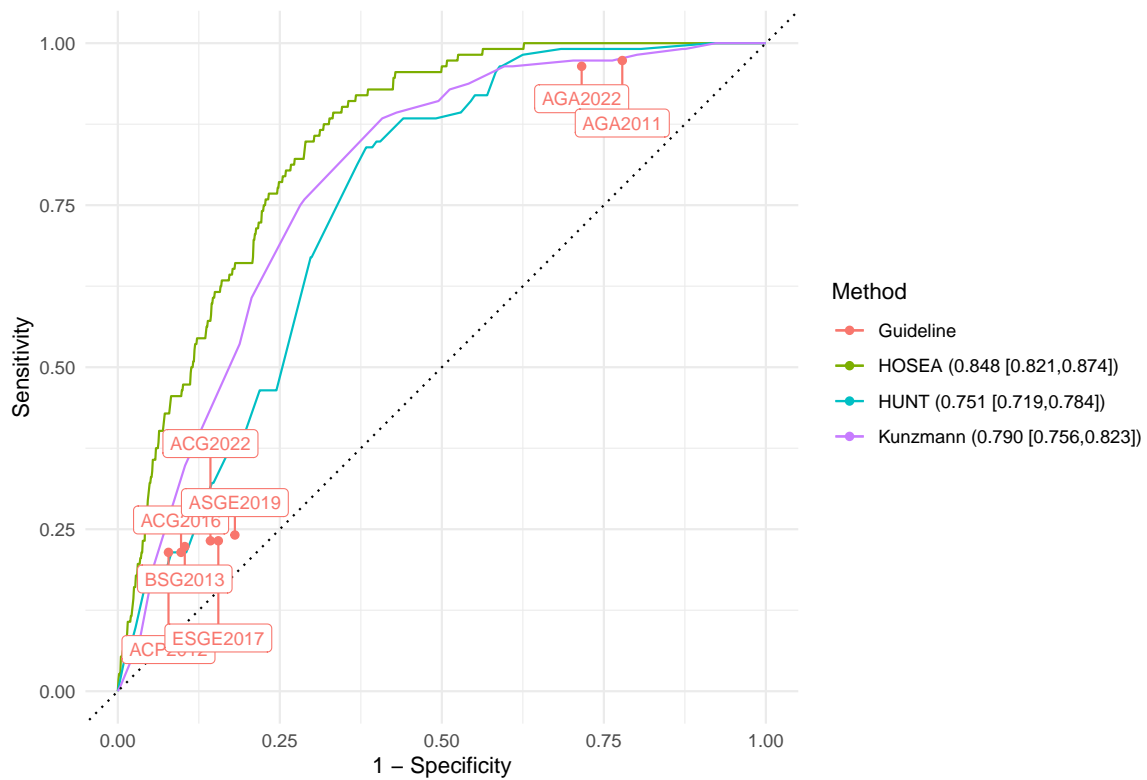


2.3 Representative sample

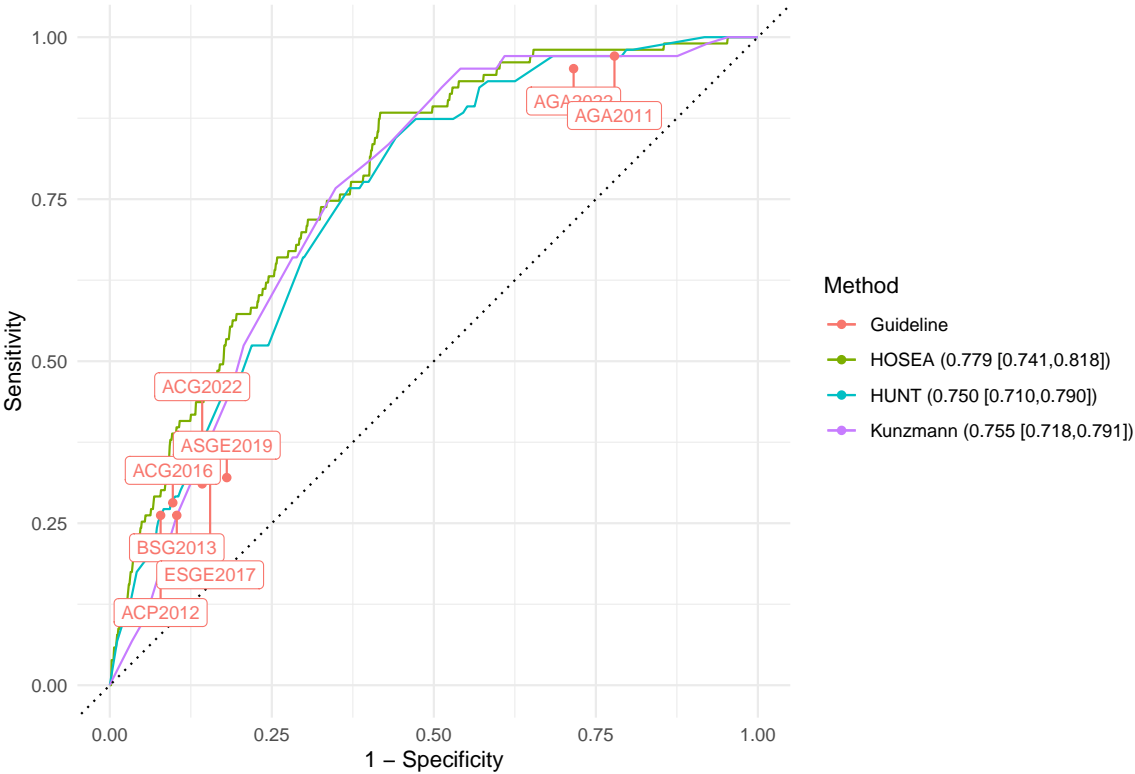
w.r.t. sex and prevalence ratio by sex



Cancer type: EAC
 Dataset: test, all, representative
 Cases: 112/429144

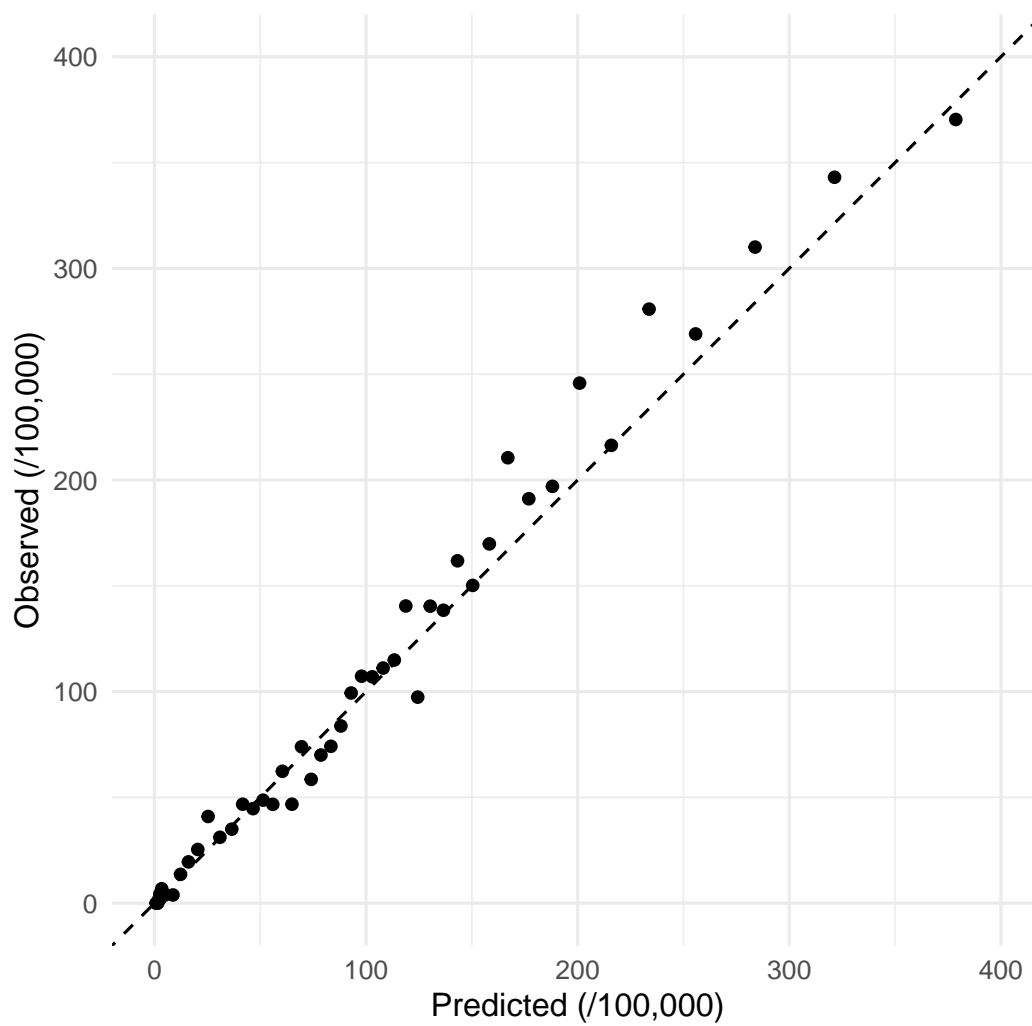


Cancer type: EGJAC
Dataset: test, all, representative
Cases: 103/429144

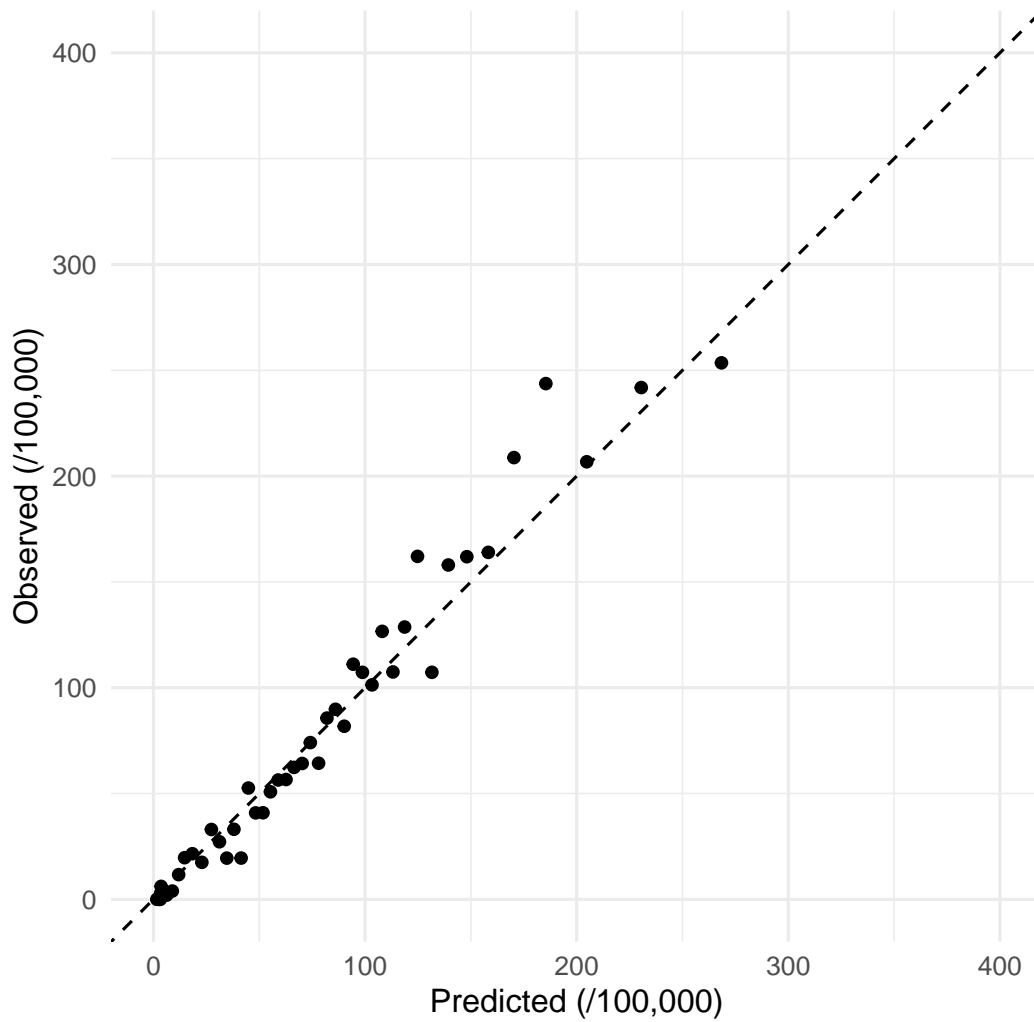


3 Calibration

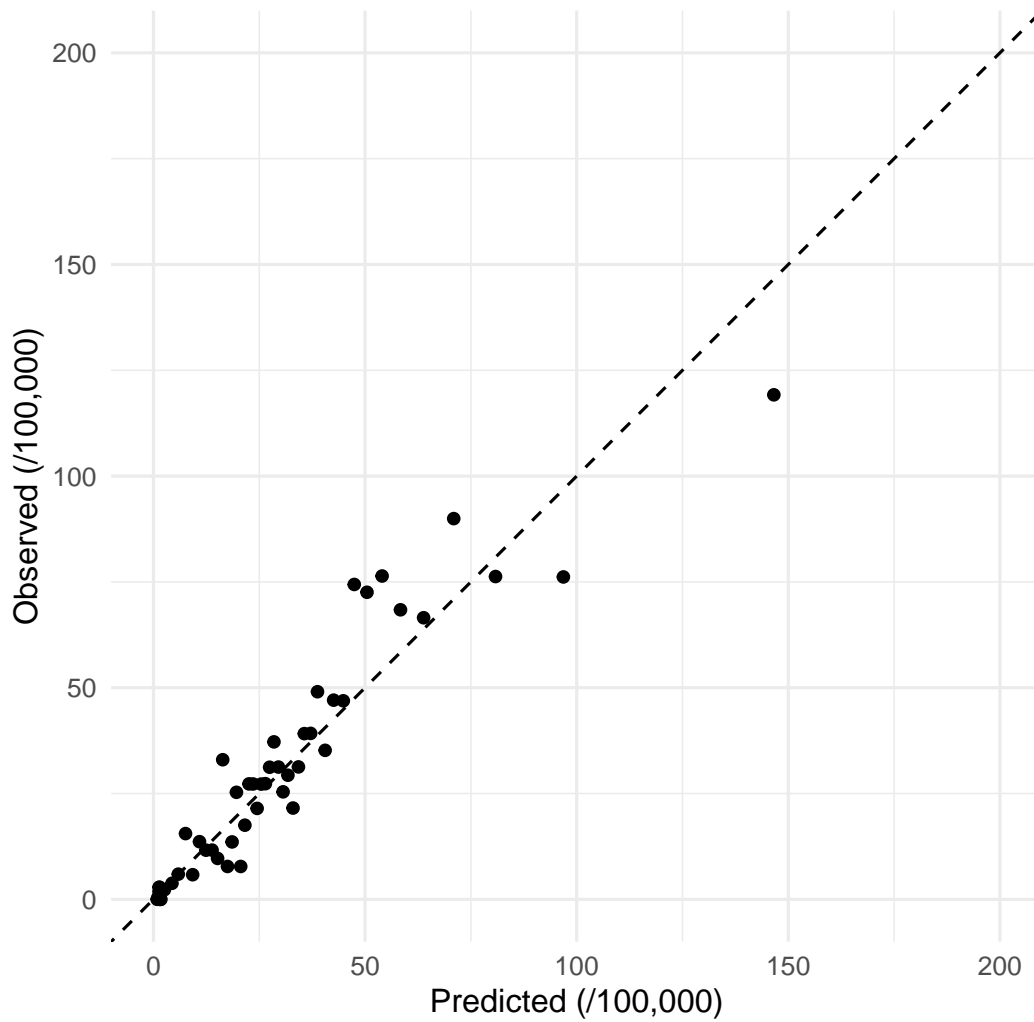
Cancer type: ANY
Dataset: test, all
Cases: 2848/2567069
HL: H=44.79, df=50, p=0.682



Cancer type: EAC
Dataset: test, all
Cases: 2076/2567069
HL: $H=61.72$, $df=50$, $p=0.124$



Cancer type: EGJAC
Dataset: test, all
Cases: 772/2567069
HL: H=62.06, df=50, p=0.118



4 Thresholds

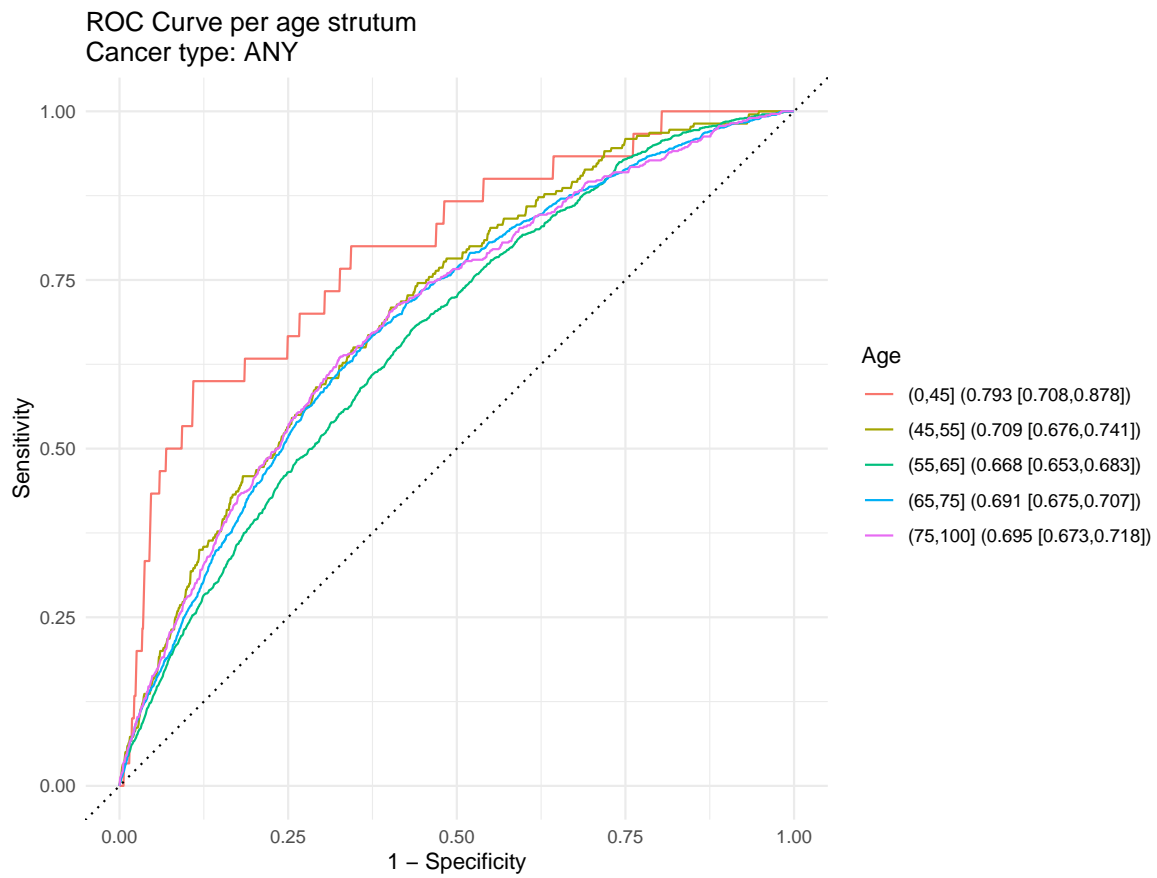
Threshold	TPR	PPV	DetPrevalence
0	100.00	0.11	100.00
5	99.61	0.13	83.06
10	99.54	0.14	80.31
15	99.23	0.14	77.45
20	98.74	0.15	75.16
25	98.10	0.15	73.09
30	97.61	0.15	71.28
35	97.05	0.15	69.54
40	96.38	0.16	67.64
45	95.58	0.16	65.62
50	94.59	0.17	63.53
55	93.82	0.17	61.36
60	92.77	0.17	59.15
65	91.75	0.18	56.95
70	90.48	0.18	54.73
75	89.19	0.19	52.54
80	87.99	0.19	50.40
85	86.34	0.20	48.27
90	84.97	0.20	46.17
95	83.08	0.21	44.13
100	81.04	0.21	42.12
105	79.21	0.22	40.14
110	77.49	0.22	38.22
115	75.21	0.23	36.34
120	73.24	0.24	34.52
125	71.38	0.24	32.76
130	69.63	0.25	31.06
135	67.38	0.25	29.45
140	65.59	0.26	27.91
145	63.59	0.27	26.46
150	61.59	0.27	25.07
155	59.73	0.28	23.77
160	58.01	0.29	22.55
165	55.90	0.29	21.40
170	53.97	0.29	20.34
175	52.35	0.30	19.33
180	50.63	0.31	18.39
185	48.63	0.31	17.50
190	47.26	0.31	16.65
195	46.10	0.32	15.86
200	44.24	0.32	15.11
220	39.04	0.35	12.48
240	33.60	0.36	10.35
260	29.71	0.38	8.63
280	25.56	0.39	7.22
300	22.40	0.41	6.01
320	19.70	0.43	5.03
340	16.89	0.45	4.19
360	14.82	0.47	3.50
380	12.75	0.49	2.91
400	11.13	0.51	2.41
420	9.55	0.53	2.00
440	8.18	0.55	1.65
460	6.99	0.57	1.33

Threshold	TPR	PPV	DetPrevalence
0	100.00	0.08	100.00
5	99.66	0.10	82.06
10	99.52	0.10	78.27
15	98.94	0.11	74.74
20	98.22	0.11	72.19
25	97.83	0.11	70.06
30	96.87	0.12	67.62
35	95.95	0.12	64.73
40	94.94	0.12	61.79
45	93.74	0.13	58.86
50	92.20	0.13	55.95
55	90.80	0.14	53.10
60	88.68	0.14	50.37
65	86.95	0.15	47.70
70	84.78	0.15	45.09
75	82.42	0.16	42.50
80	80.59	0.16	39.95
85	77.84	0.17	37.44
90	75.29	0.17	35.02
95	72.45	0.18	32.68
100	69.36	0.18	30.40
105	66.57	0.19	28.23
110	63.54	0.20	26.19
115	60.93	0.20	24.27
120	58.48	0.21	22.51
125	55.25	0.21	20.89
130	52.50	0.22	19.40
135	50.87	0.23	18.06
140	48.41	0.23	16.81
145	46.10	0.24	15.64
150	43.88	0.24	14.57
155	41.96	0.25	13.56
160	40.32	0.26	12.65
165	38.54	0.26	11.81
170	36.51	0.27	11.02
175	34.54	0.27	10.30
180	32.47	0.27	9.64
185	30.68	0.28	9.02
190	28.85	0.28	8.44
195	27.36	0.28	7.91
200	25.92	0.28	7.41
220	21.53	0.31	5.71
240	17.53	0.32	4.37
260	14.64	0.36	3.32
280	12.09	0.39	2.49
300	9.44	0.41	1.87
320	7.66	0.45	1.37
340	5.25	0.42	1.00
360	4.05	0.45	0.72
380	2.99	0.47	0.52
400	1.93	0.43	0.37
420	1.45	0.45	0.26
440	1.06	0.48	0.18
460	0.87	0.44	0.13

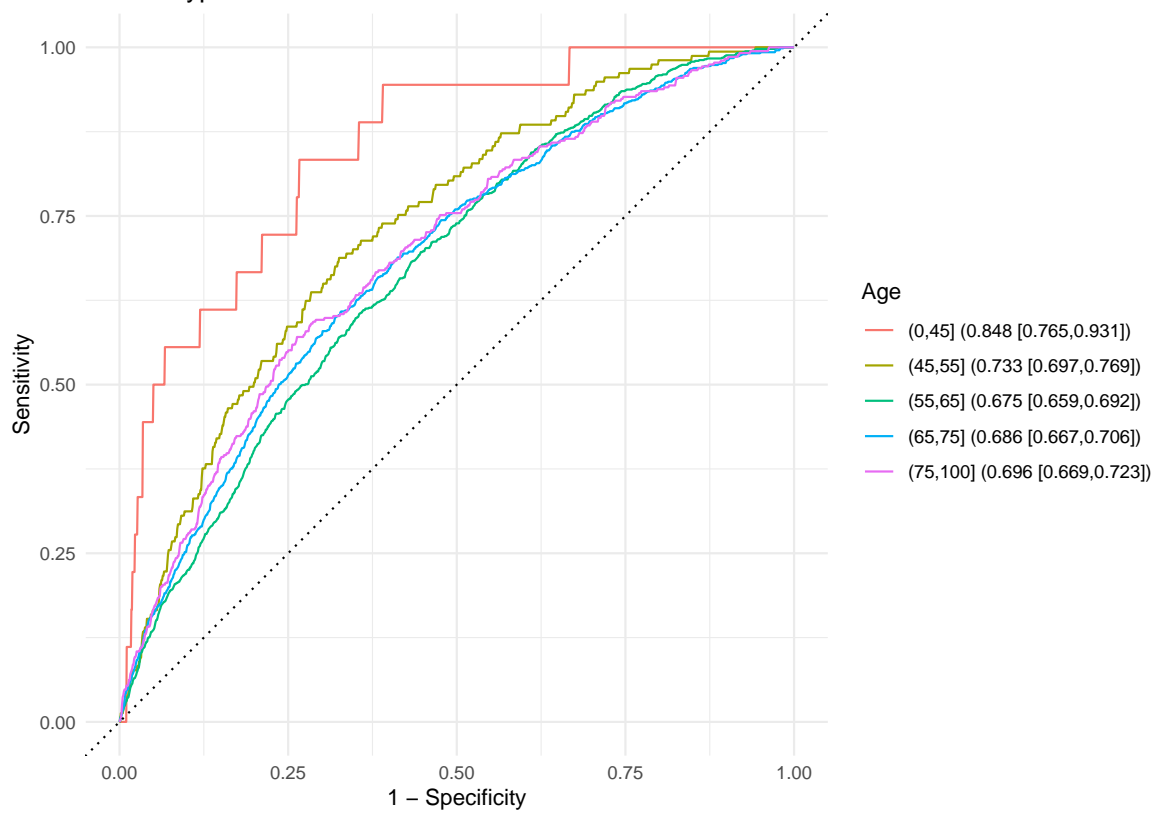
Threshold	TPR	PPV	DetPrevalence
0	100.00	0.03	100.00
5	99.09	0.04	80.04
10	97.28	0.04	73.97
15	94.17	0.04	67.09
20	89.12	0.05	58.04
25	81.99	0.05	47.75
30	71.89	0.06	37.95
35	64.64	0.07	29.76
40	55.83	0.07	23.48
45	49.09	0.08	18.79
50	40.54	0.08	15.20
55	34.46	0.08	12.45
60	28.63	0.08	10.29
65	25.52	0.09	8.58
70	21.11	0.09	7.18
75	18.01	0.09	6.04
80	15.54	0.09	5.08
85	13.08	0.09	4.30
90	11.79	0.10	3.65
95	10.62	0.10	3.11
100	9.97	0.11	2.65
105	8.94	0.12	2.27
110	7.90	0.12	1.95
115	7.25	0.13	1.67
120	6.87	0.14	1.44
125	6.22	0.15	1.25
130	5.31	0.15	1.09
135	4.66	0.15	0.94
140	4.27	0.16	0.82
145	3.63	0.15	0.72
150	3.24	0.15	0.63
155	2.72	0.15	0.55
160	2.46	0.15	0.48
165	1.94	0.14	0.43
170	1.55	0.12	0.38
175	1.42	0.13	0.33
180	1.30	0.13	0.29
185	1.04	0.12	0.26
190	0.78	0.10	0.23
195	0.52	0.08	0.21
200	0.39	0.06	0.18
220	0.26	0.07	0.12
240	0.26	0.10	0.08
260	0.13	0.08	0.05
280	0.00	0.00	0.04
300	0.00	0.00	0.02
320	0.00	0.00	0.02
340	0.00	0.00	0.01
360	0.00	0.00	0.01
380	0.00	0.00	0.01
400	0.00	0.00	0.00
420	0.00	0.00	0.00
440	0.00	0.00	0.00
460	0.00	0.00	0.00

5 Stratification by identity groups

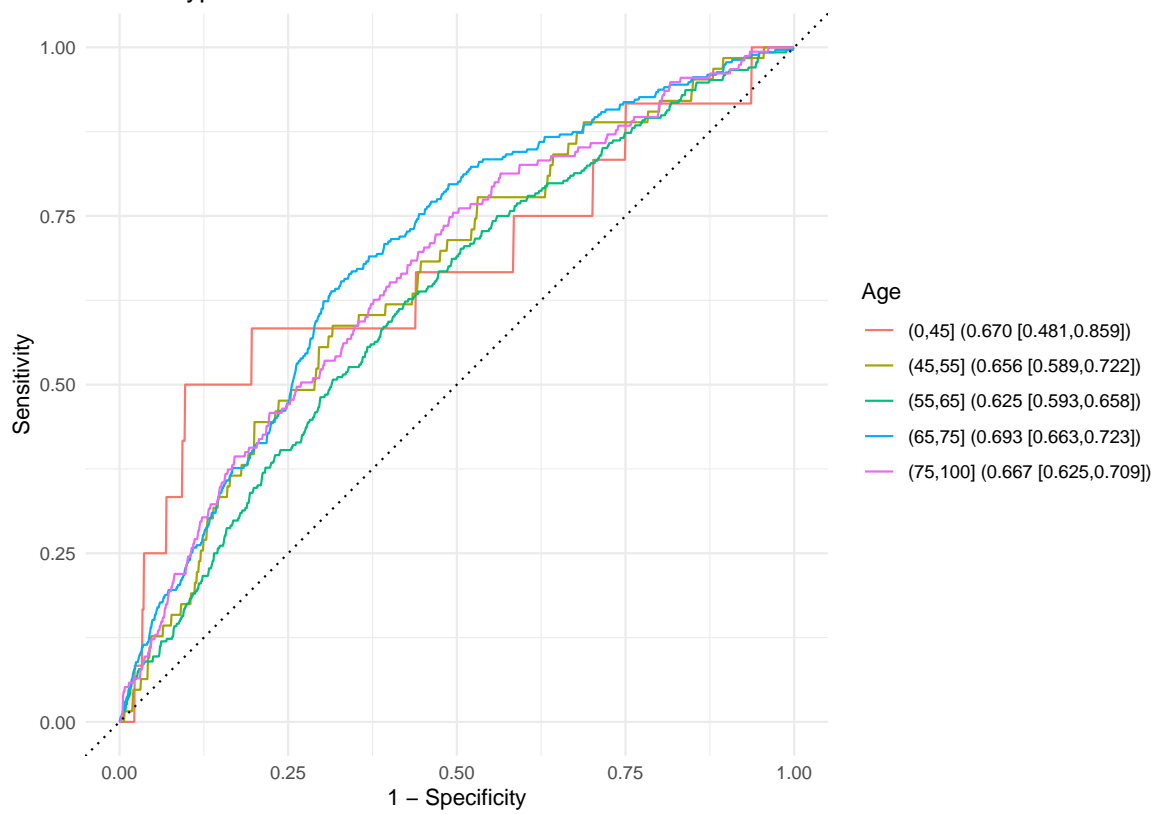
5.1 Age



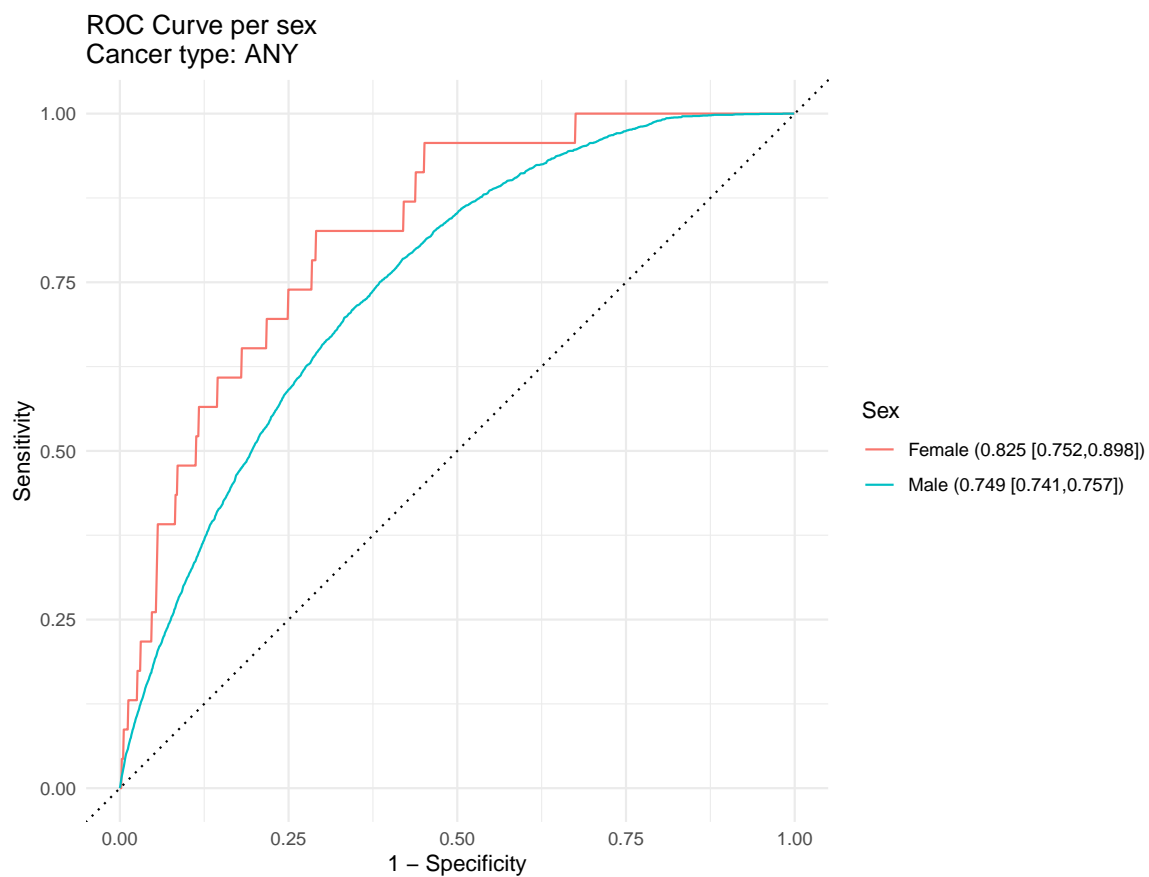
ROC Curve per age stratum
Cancer type: EAC

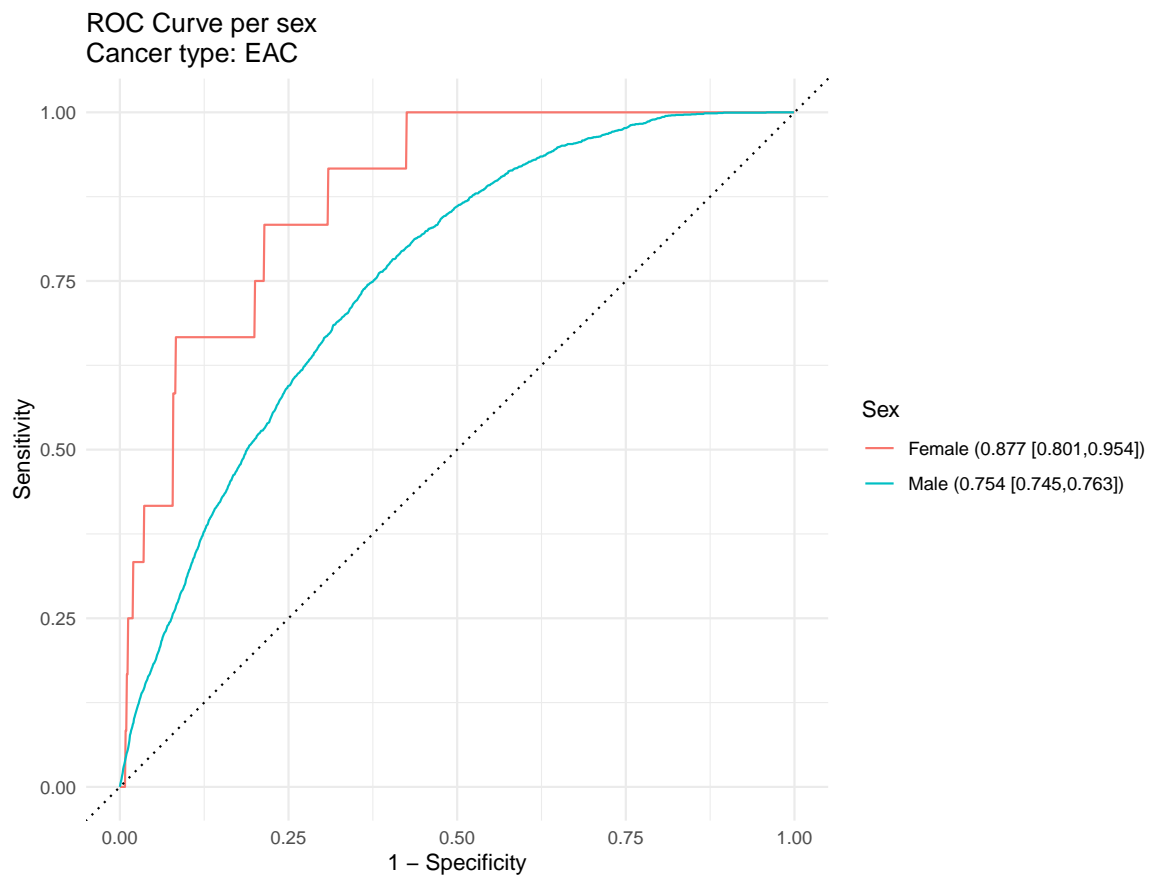


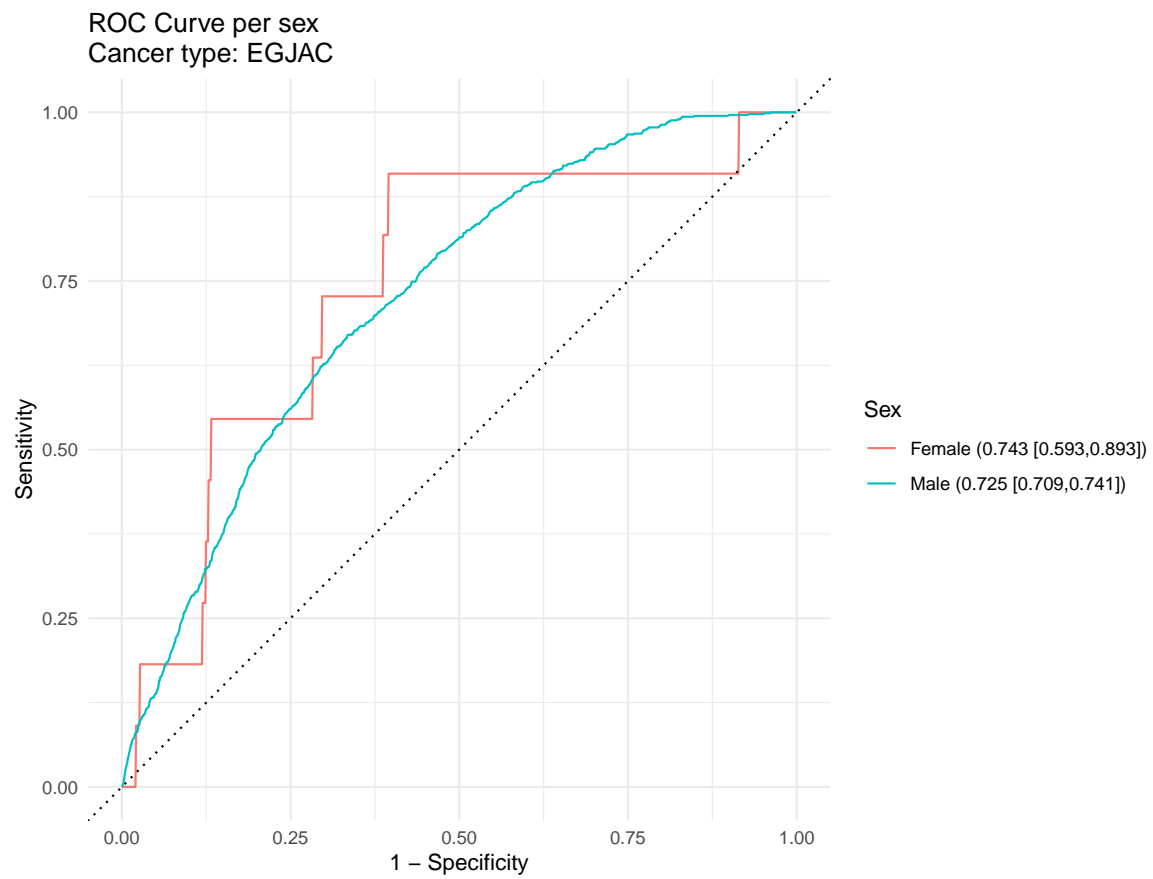
ROC Curve per age stratum
Cancer type: EGJAC



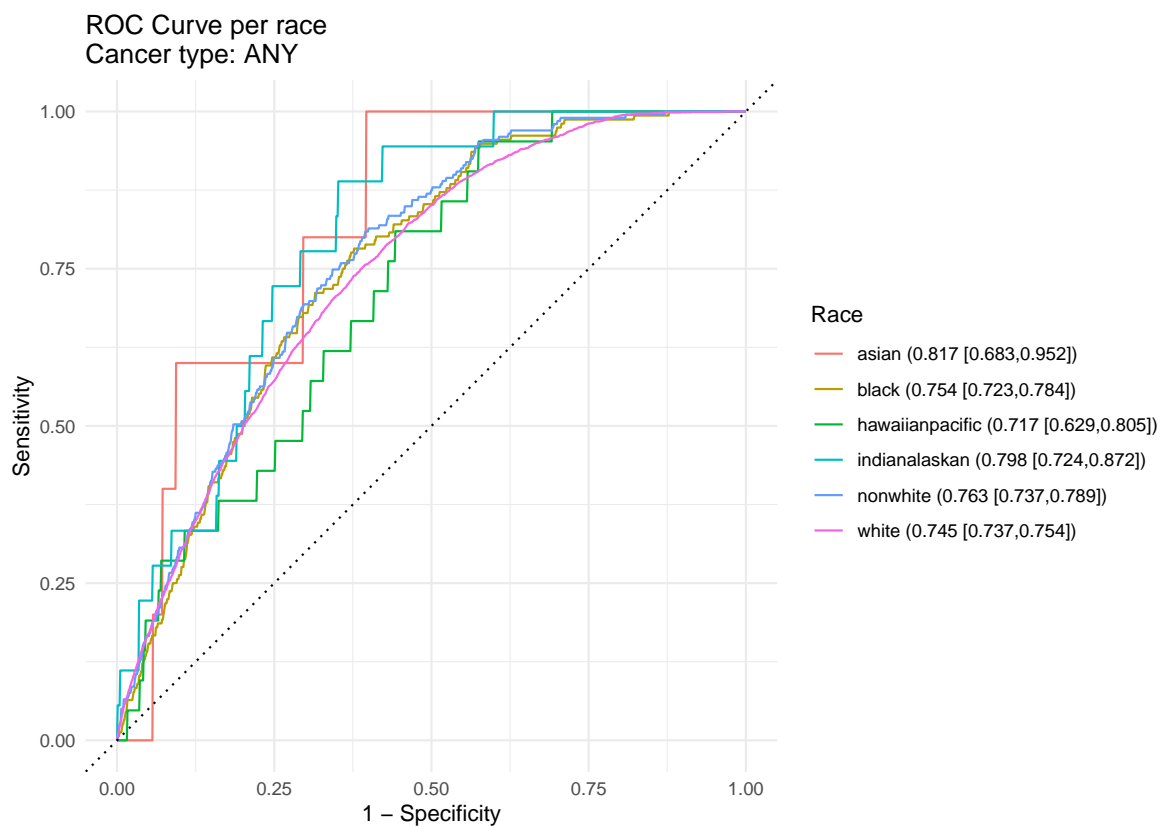
5.2 Sex

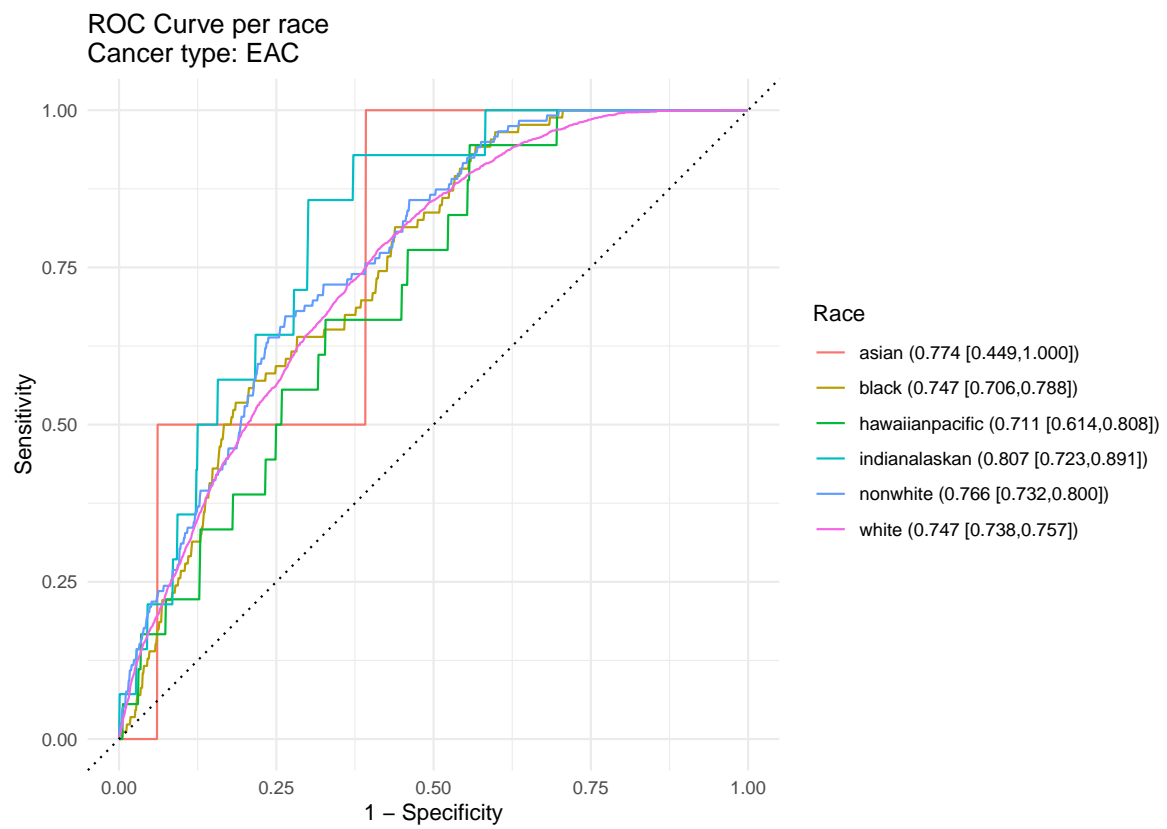


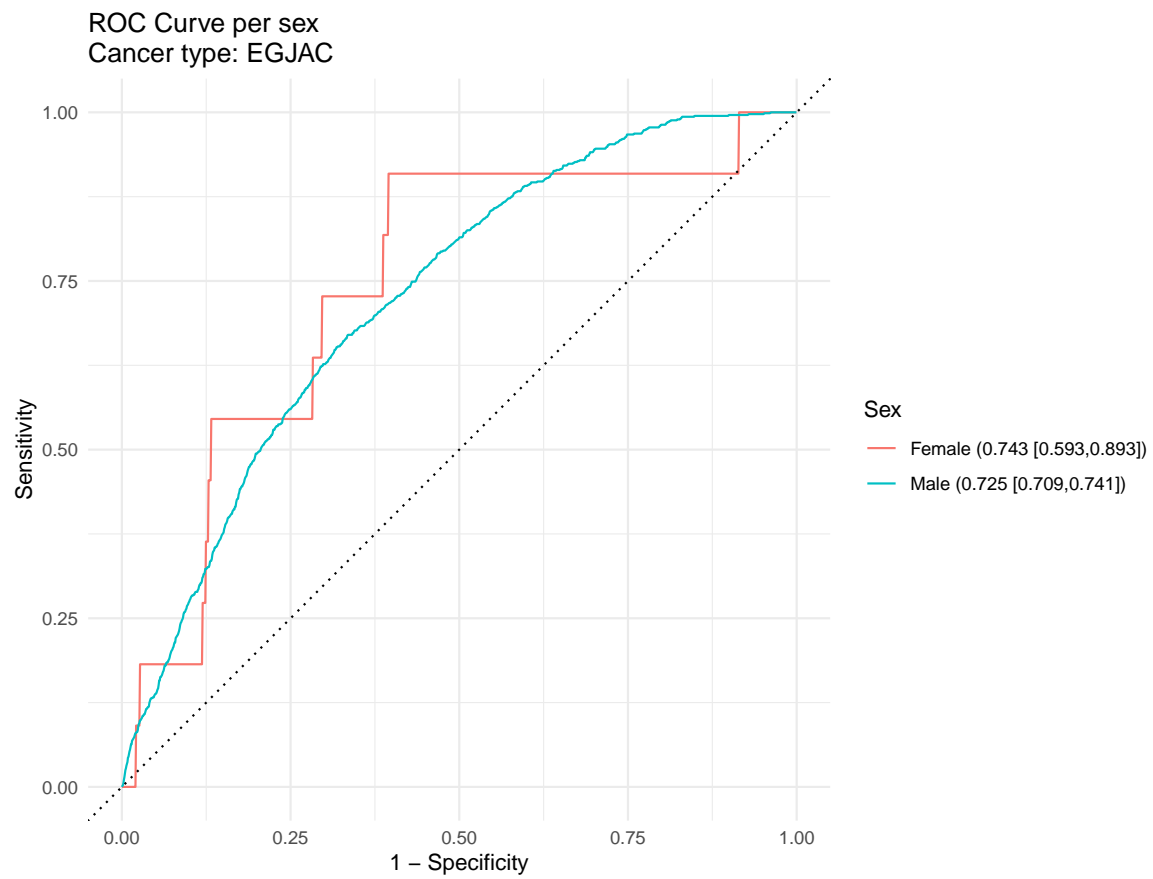




5.3 Race







6 Cancer stage

NB: the discrepancy in number of cases between I+ and Any comes from the fact that some were classified as unknown stage according to the provided staging.

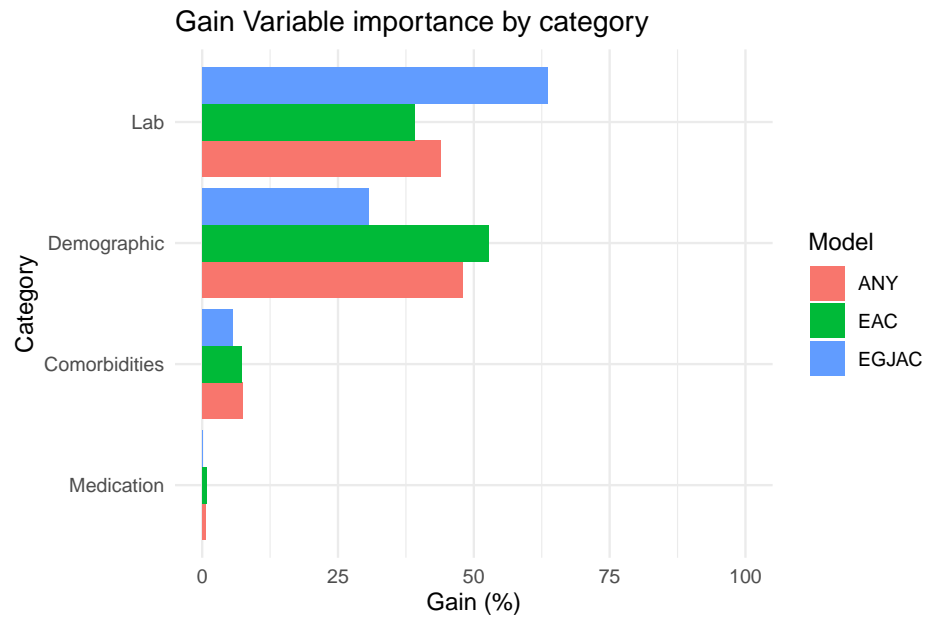
Stage	Test. AUC	Nb. cases
Any	0.765 [0.758,0.773]	2818
I	0.801 [0.784,0.818]	495
II	0.779 [0.760,0.798]	435
III	0.769 [0.756,0.782]	878
IV	0.759 [0.748,0.770]	1244
I+	0.768 [0.760,0.776]	2448
II+	0.762 [0.753,0.770]	2169
III+	0.760 [0.751,0.769]	1950
IV+	0.759 [0.748,0.770]	1244

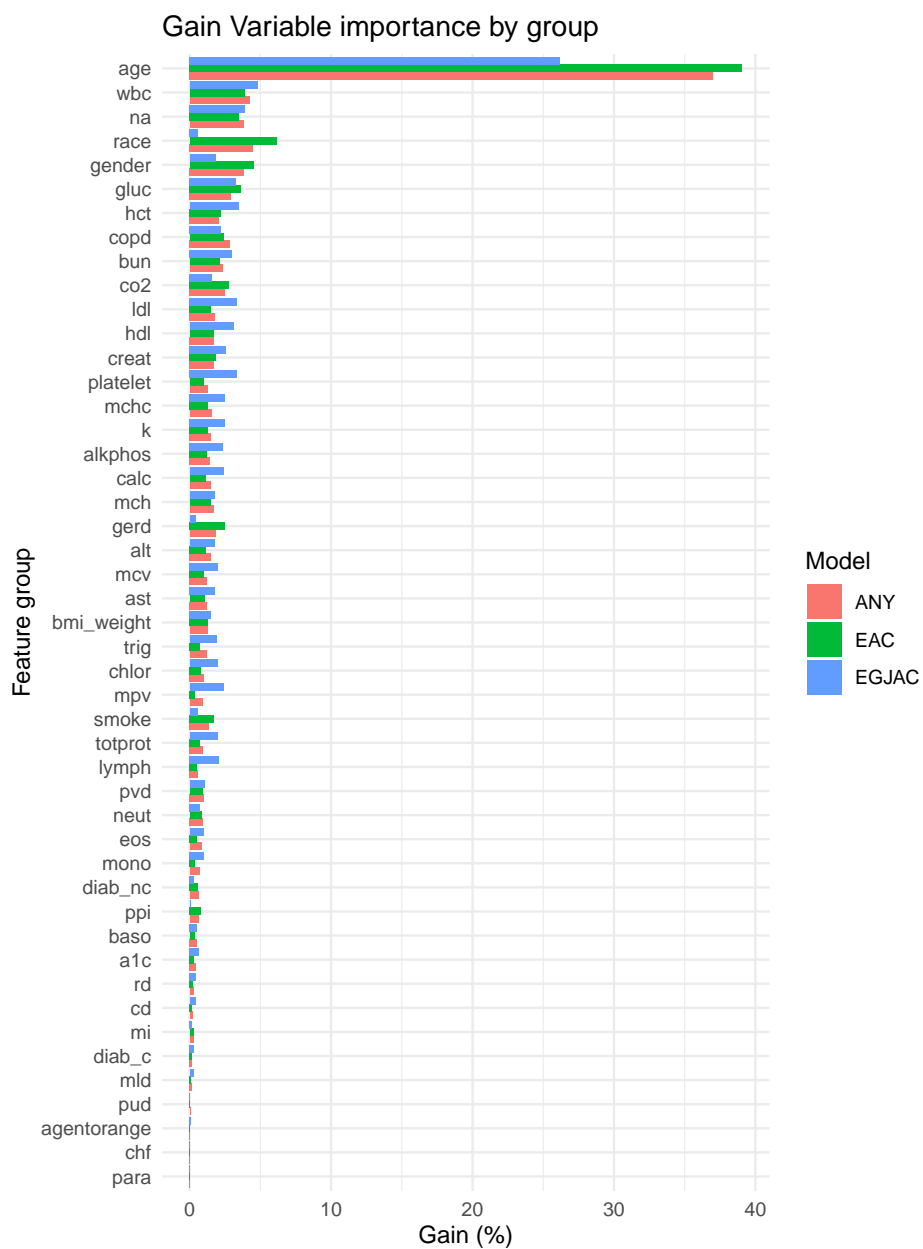
Stage	Test. AUC	Nb. cases
Any	0.771 [0.763,0.780]	2054
I	0.810 [0.791,0.829]	350
II	0.786 [0.765,0.807]	302
III	0.770 [0.756,0.785]	650
IV	0.764 [0.751,0.776]	913
I+	0.772 [0.763,0.780]	1813
II+	0.765 [0.756,0.775]	1608
III+	0.764 [0.754,0.774]	1451
IV+	0.764 [0.751,0.776]	913

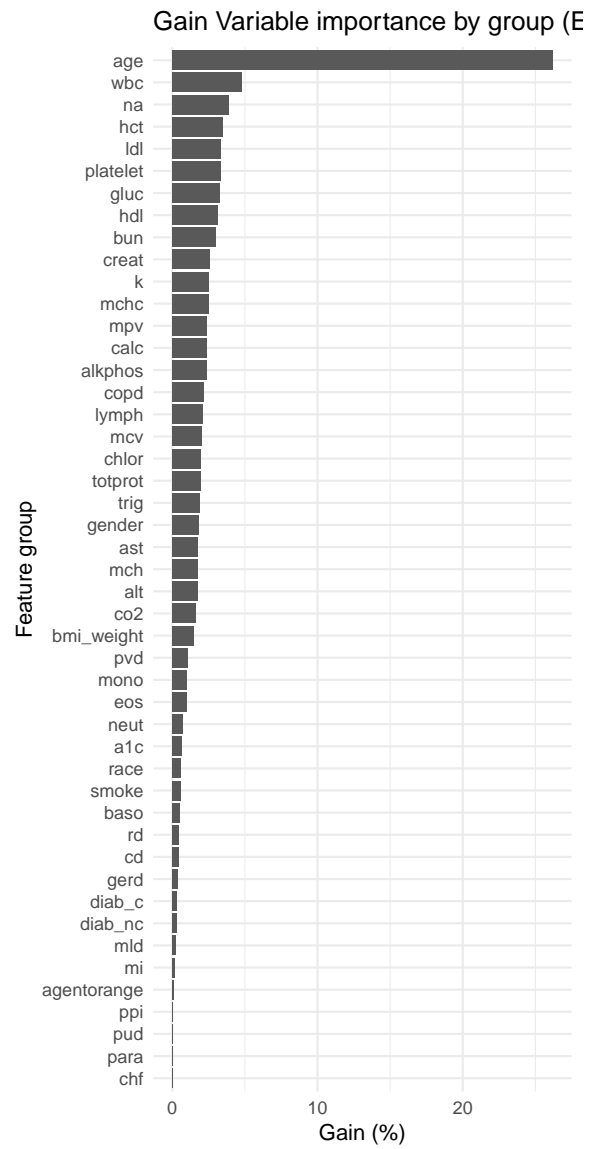
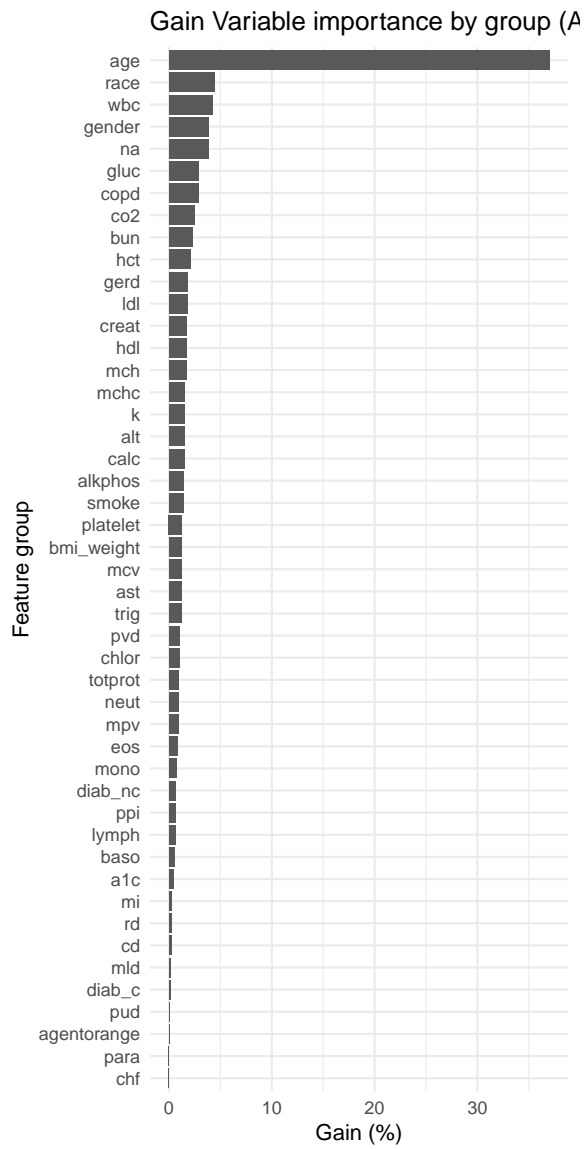
Stage	Test. AUC	Nb. cases
Any	0.739 [0.724,0.754]	764
I	0.766 [0.729,0.802]	145
II	0.750 [0.712,0.788]	133
III	0.750 [0.723,0.777]	228
IV	0.731 [0.707,0.755]	331
I+	0.743 [0.727,0.760]	635
II+	0.738 [0.720,0.755]	561
III+	0.735 [0.717,0.754]	499
IV+	0.731 [0.707,0.755]	331

7 Variable importance

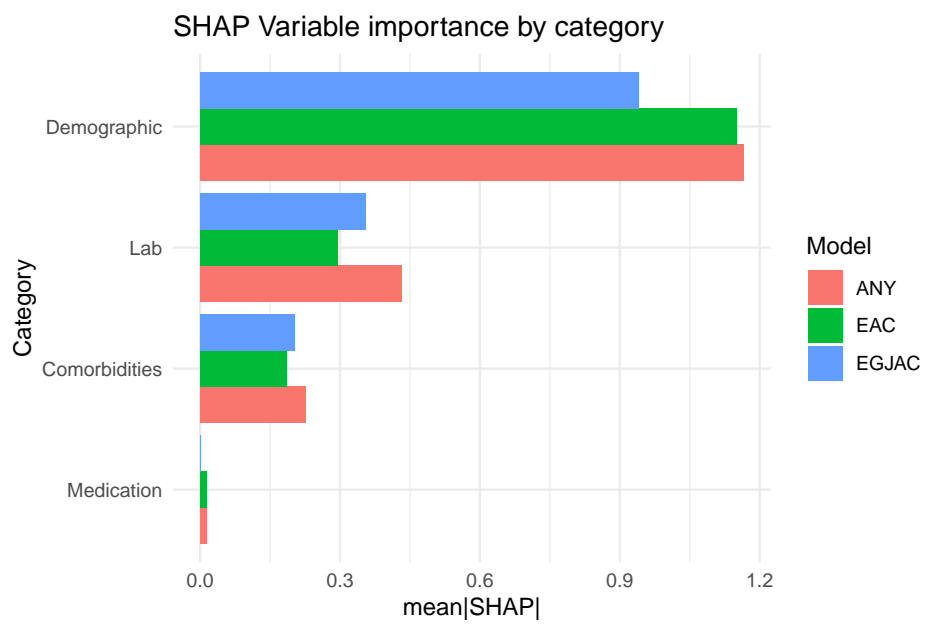
7.1 Gain VI

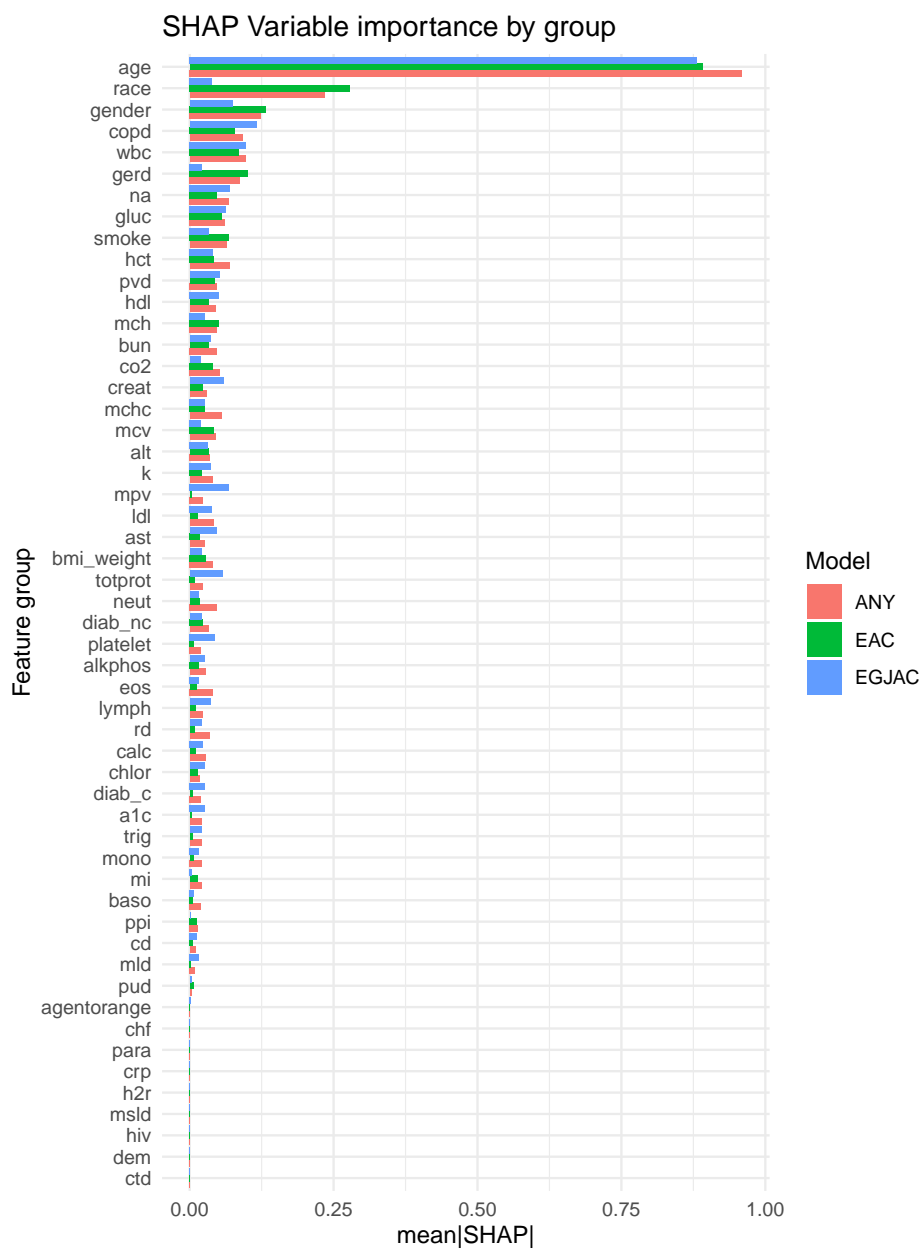


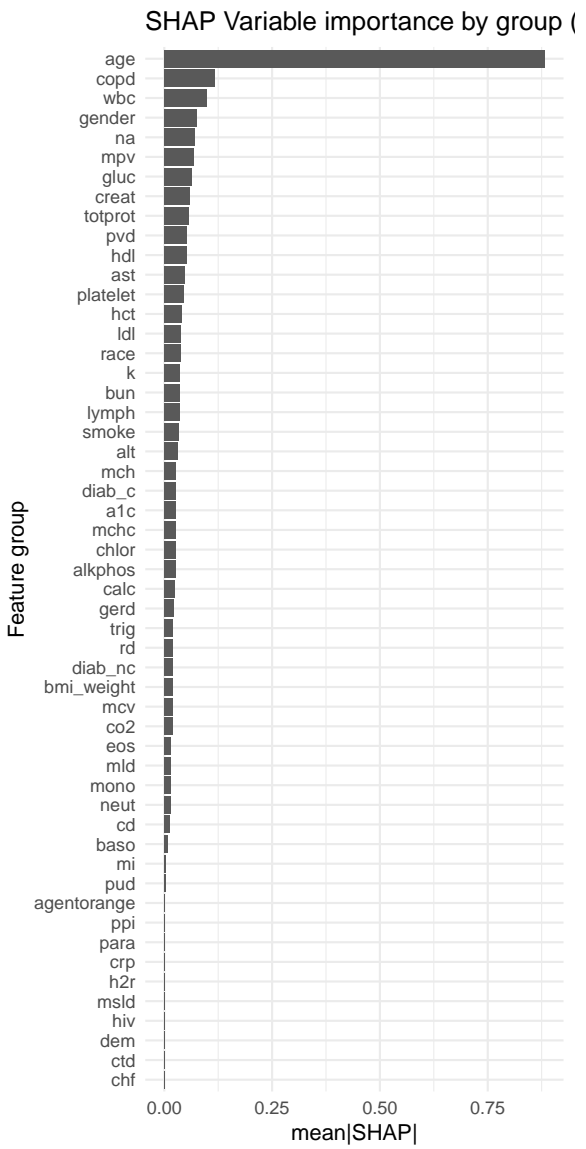
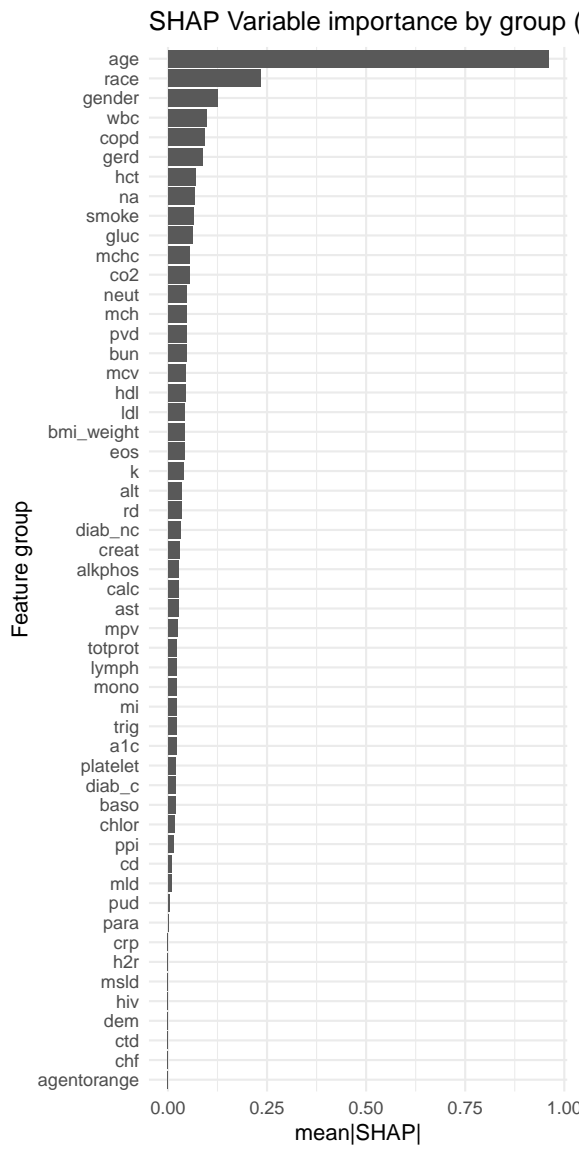




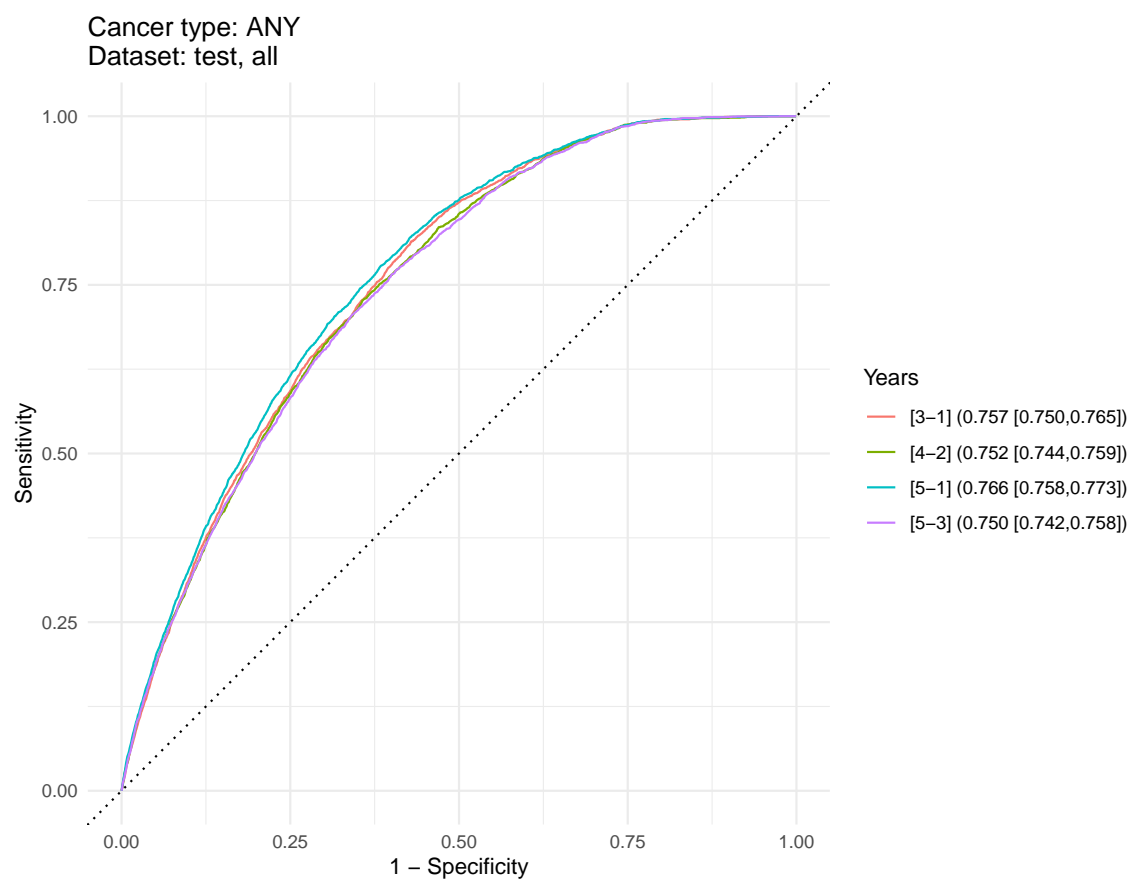
7.2 SHAP VI

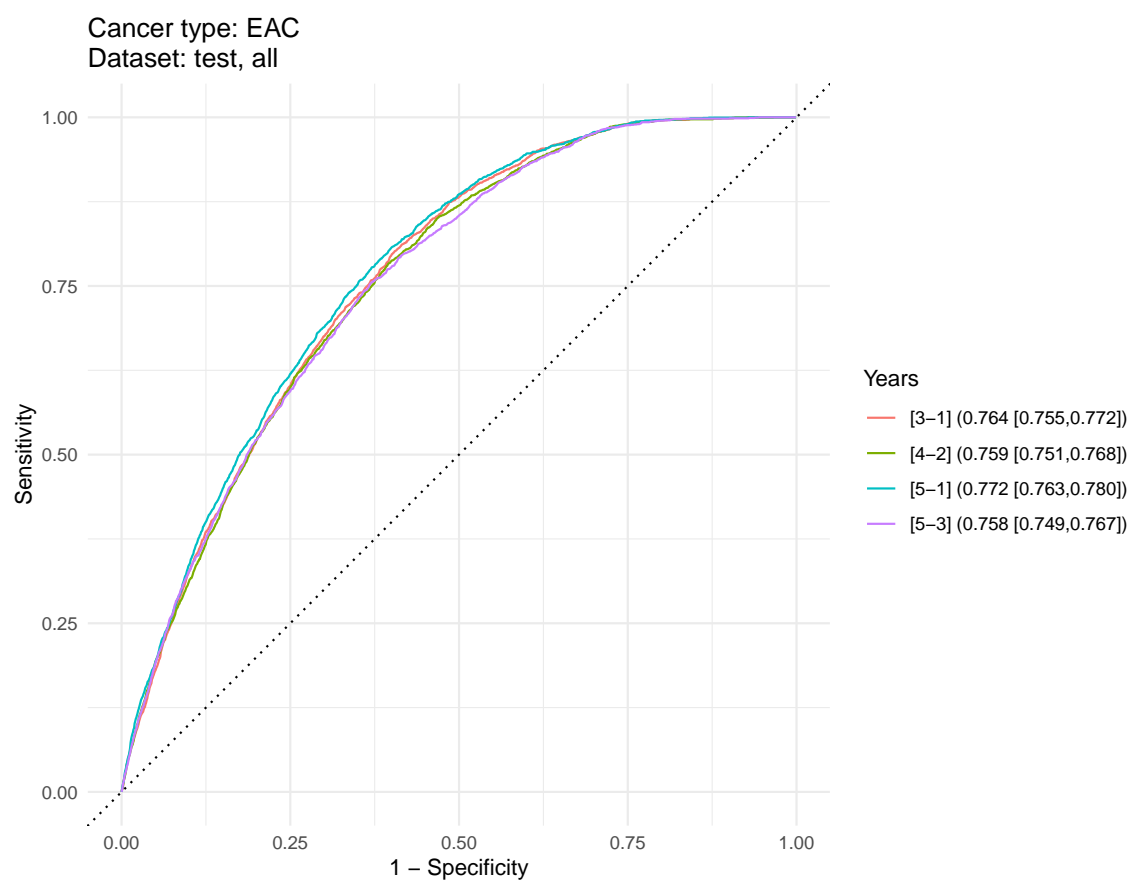


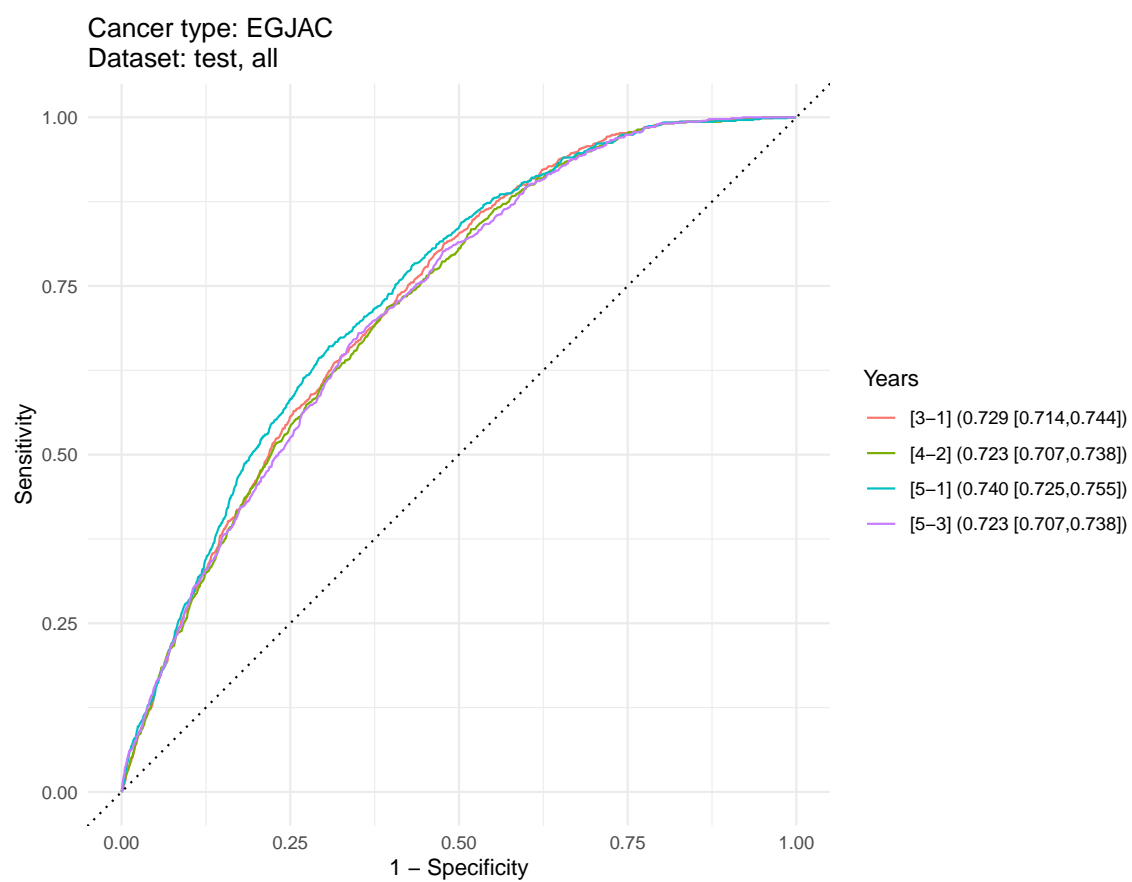




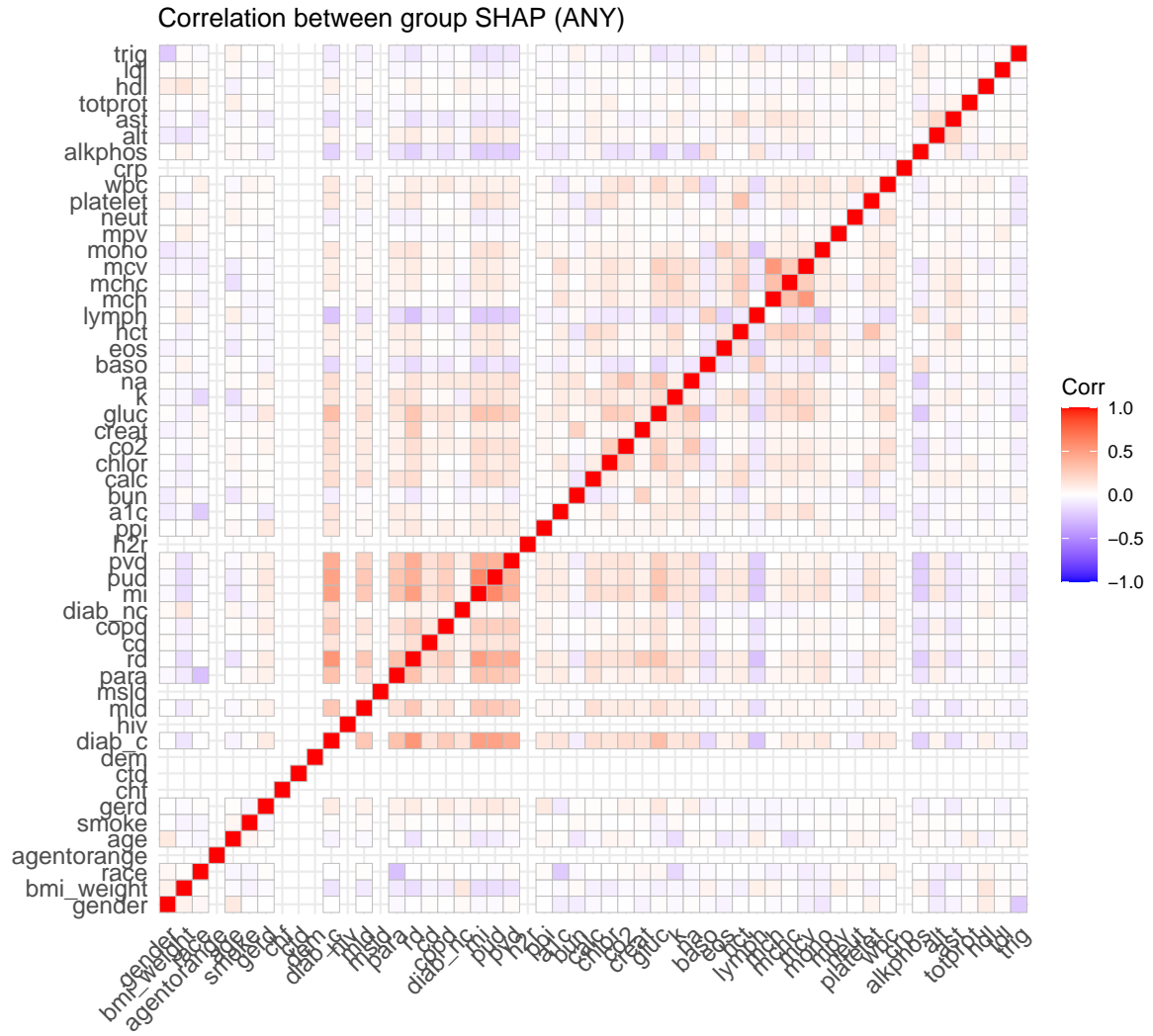
8 Years prior

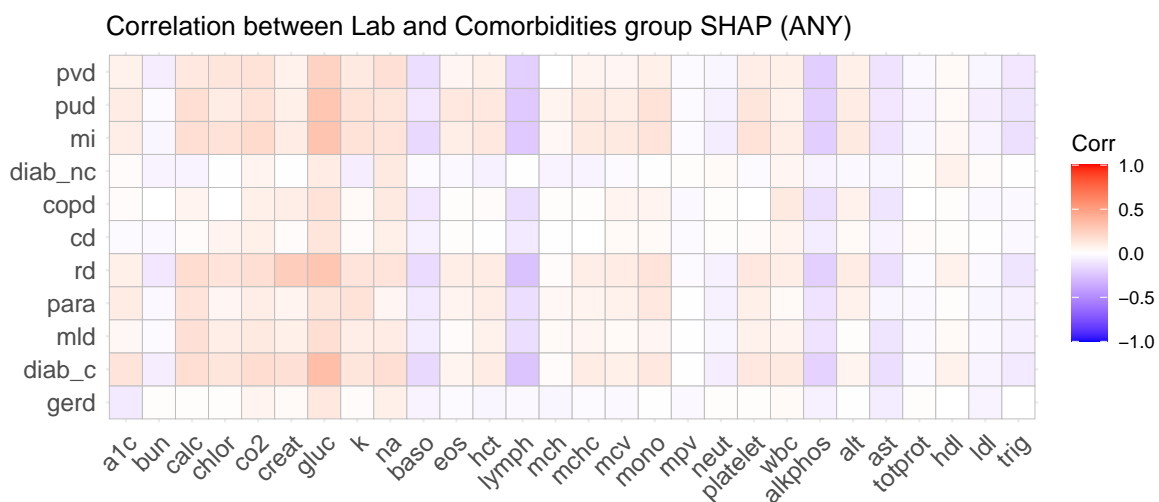
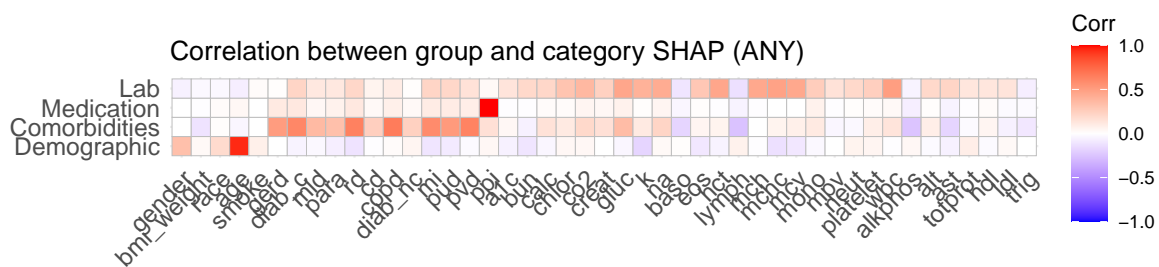






9 SHAP correlation and imputation model





MICE Lab Coefficients for
imputing Comorbidities

