

HOSEA Aim I – Report

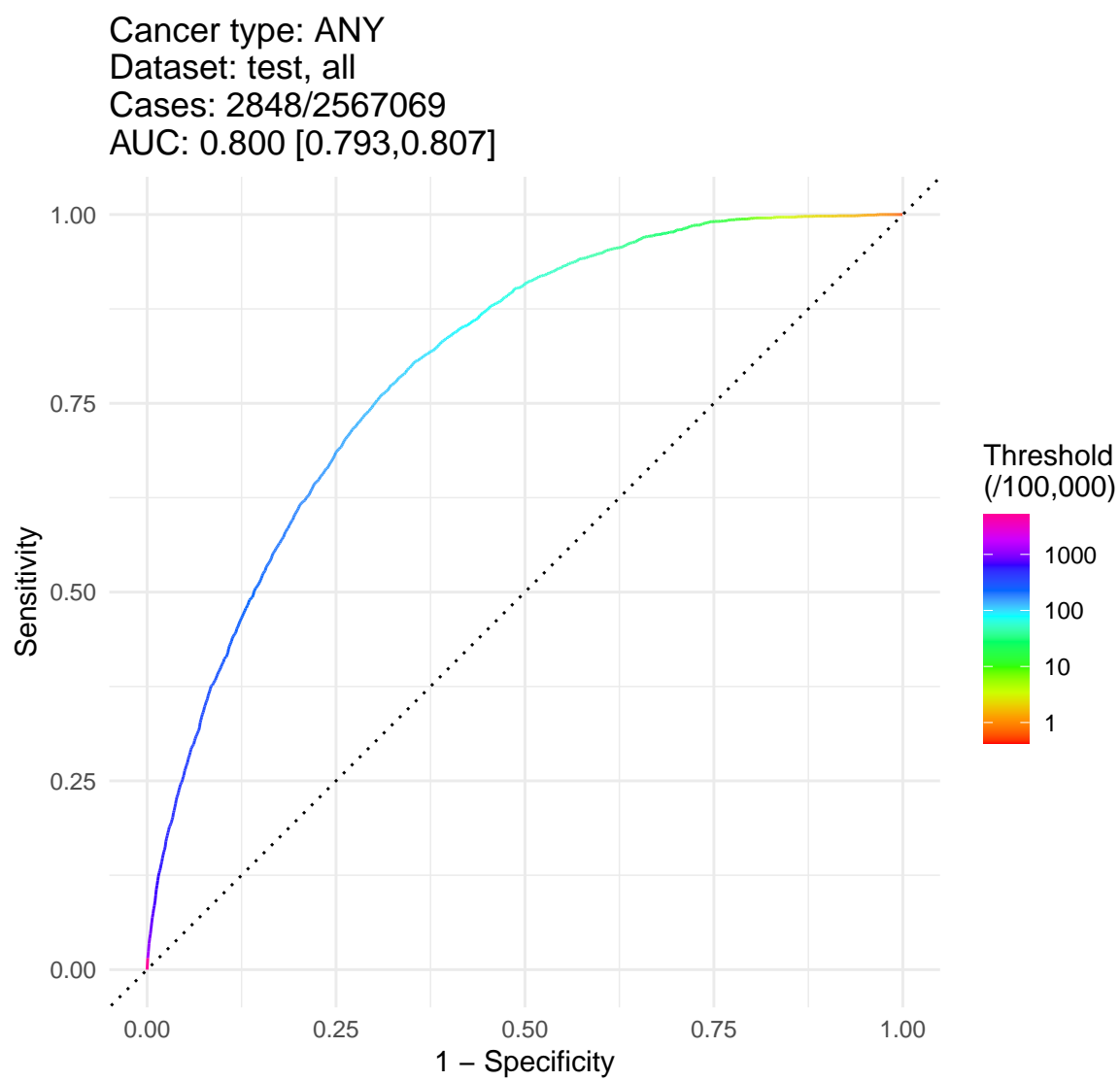
Simon Fontaine

September 19, 2022

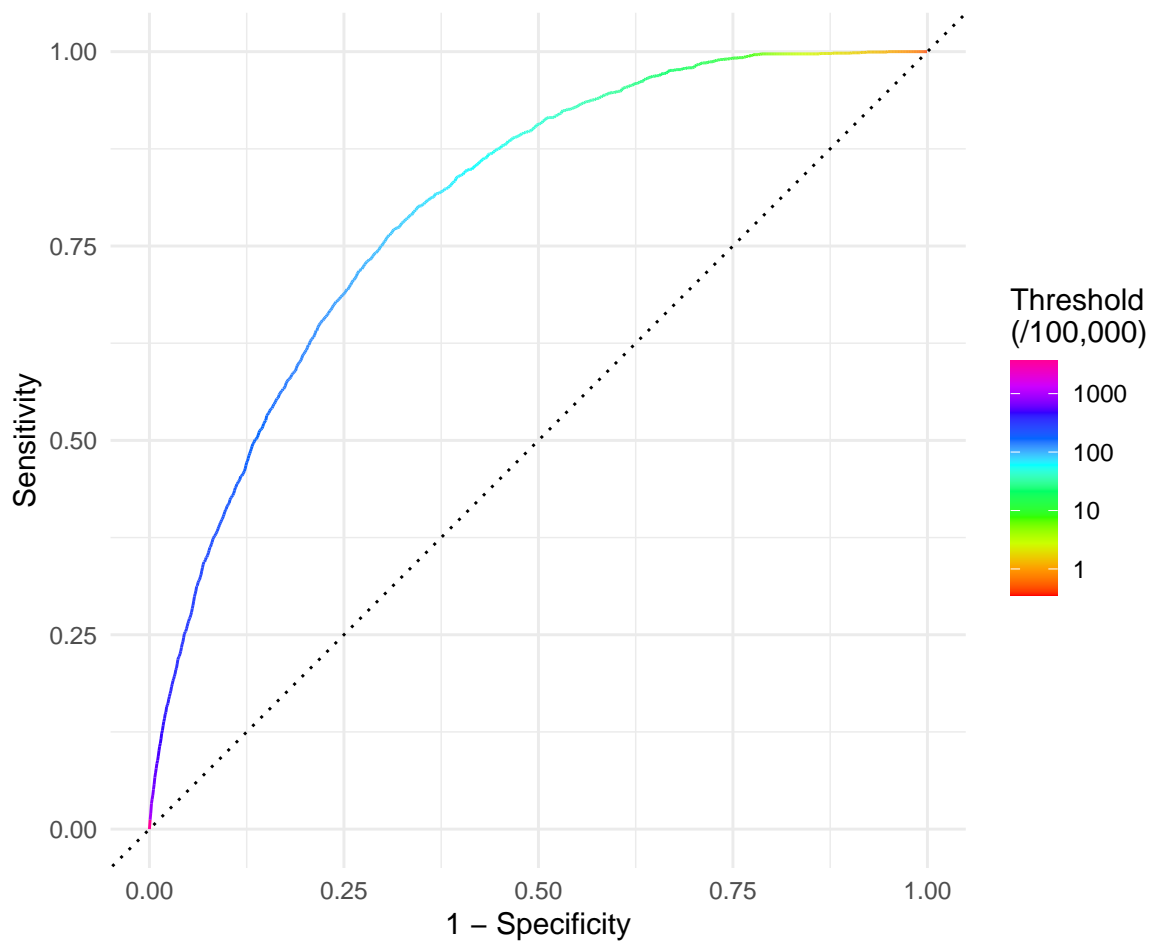
Contents

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	32
7.1	Gain VI	32
7.2	SHAP VI	33
8	Years prior	34

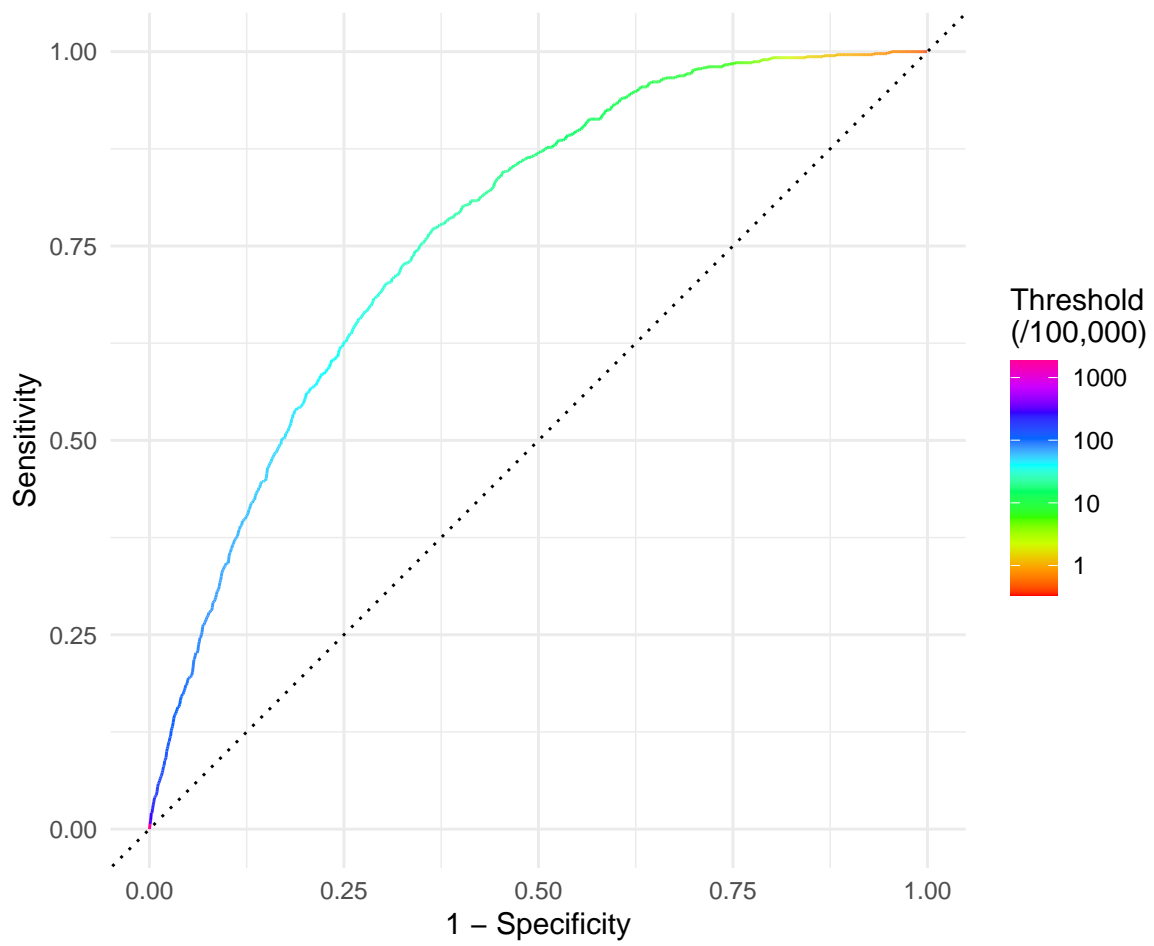
1 Overall performance



Cancer type: EAC
Dataset: test, all
Cases: 2076/2567069
AUC: 0.802 [0.794,0.811]

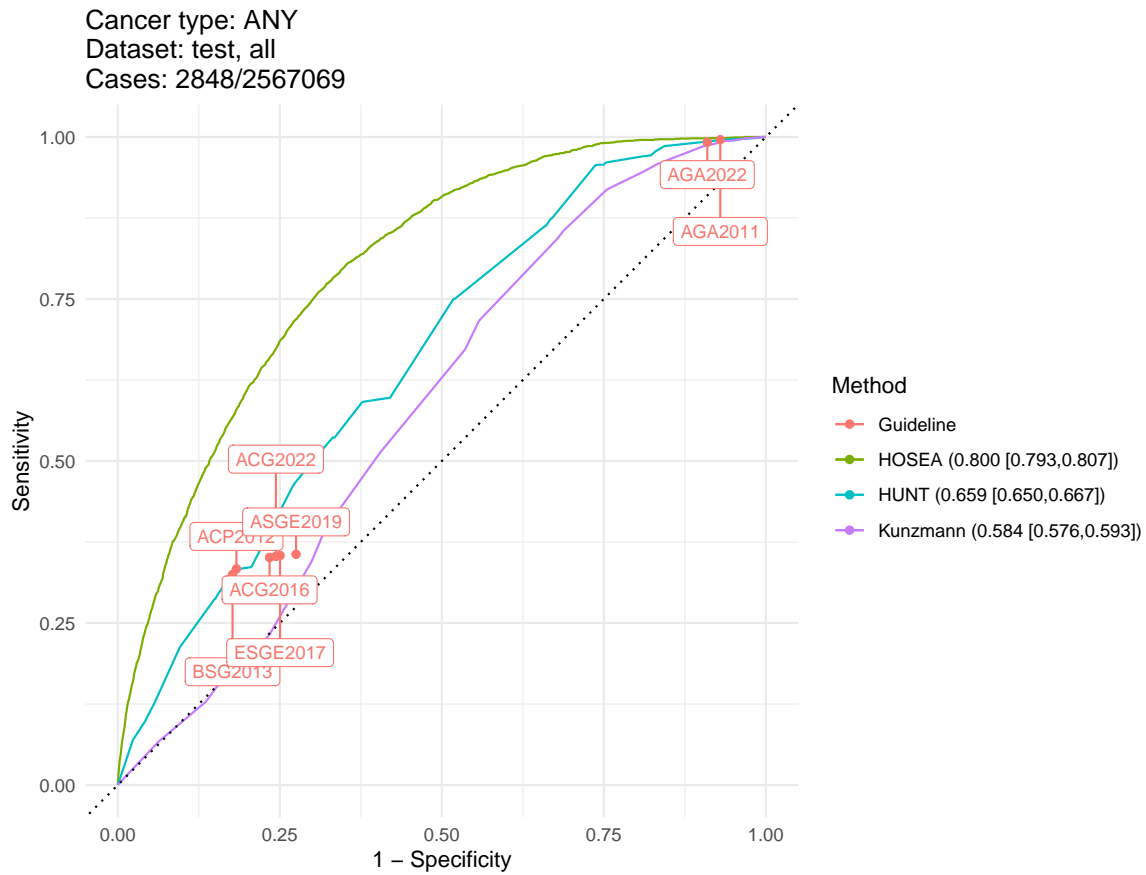


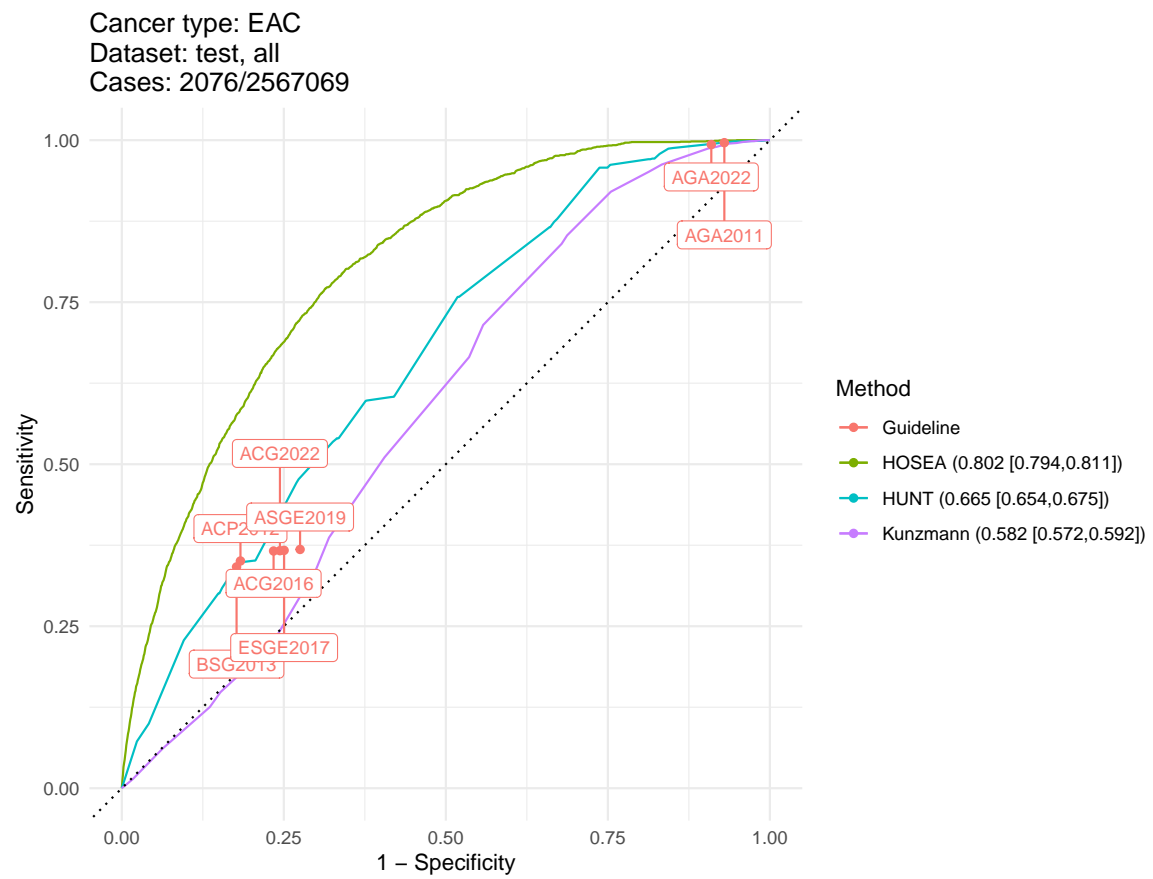
Cancer type: EGJAC
Dataset: test, all
Cases: 772/2567069
AUC: 0.769 [0.755,0.783]



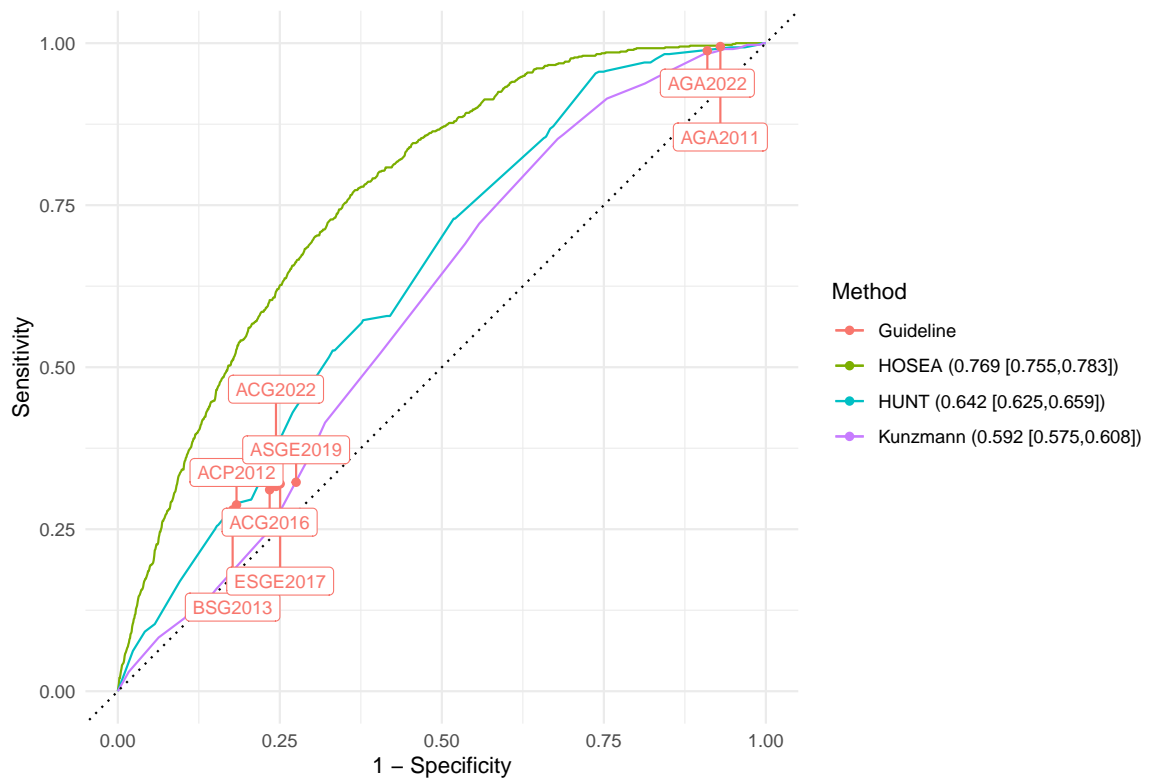
2 Comparison to HUNT, Kunzmann and Guidelines

2.1 Full test data



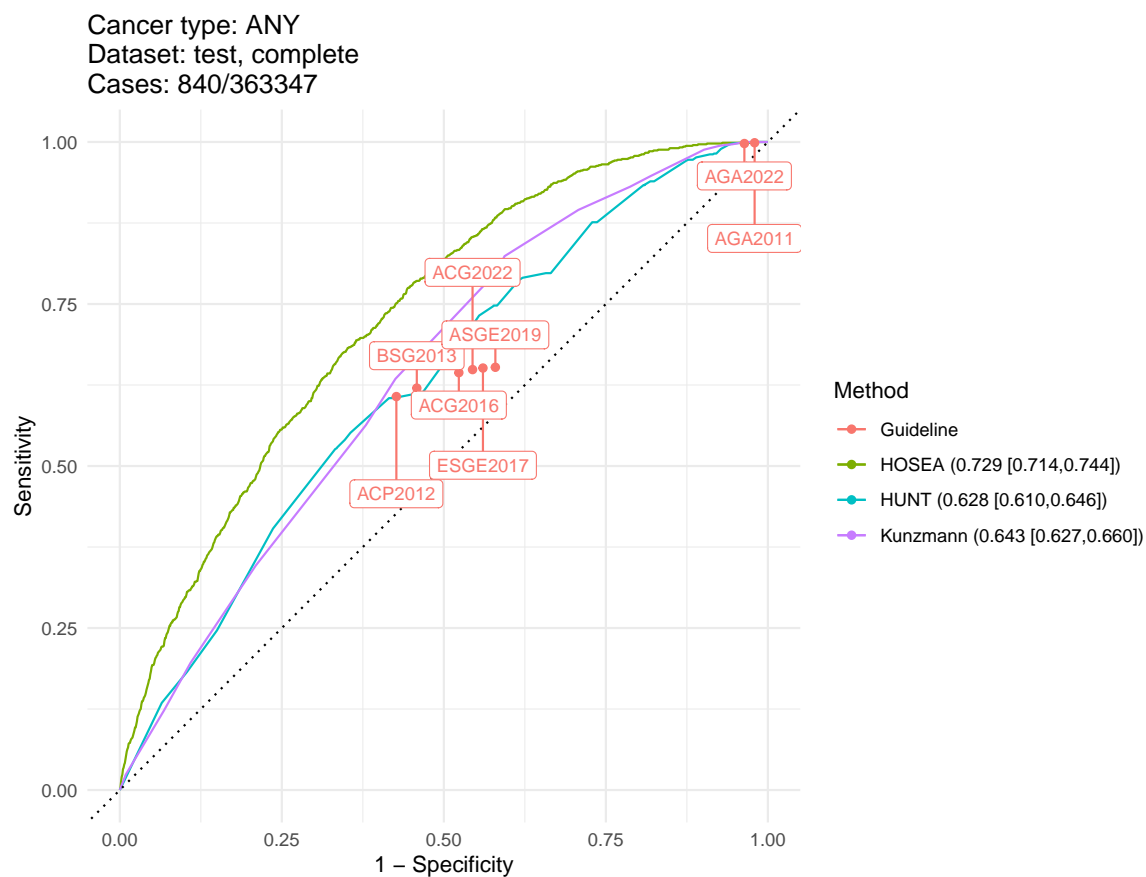


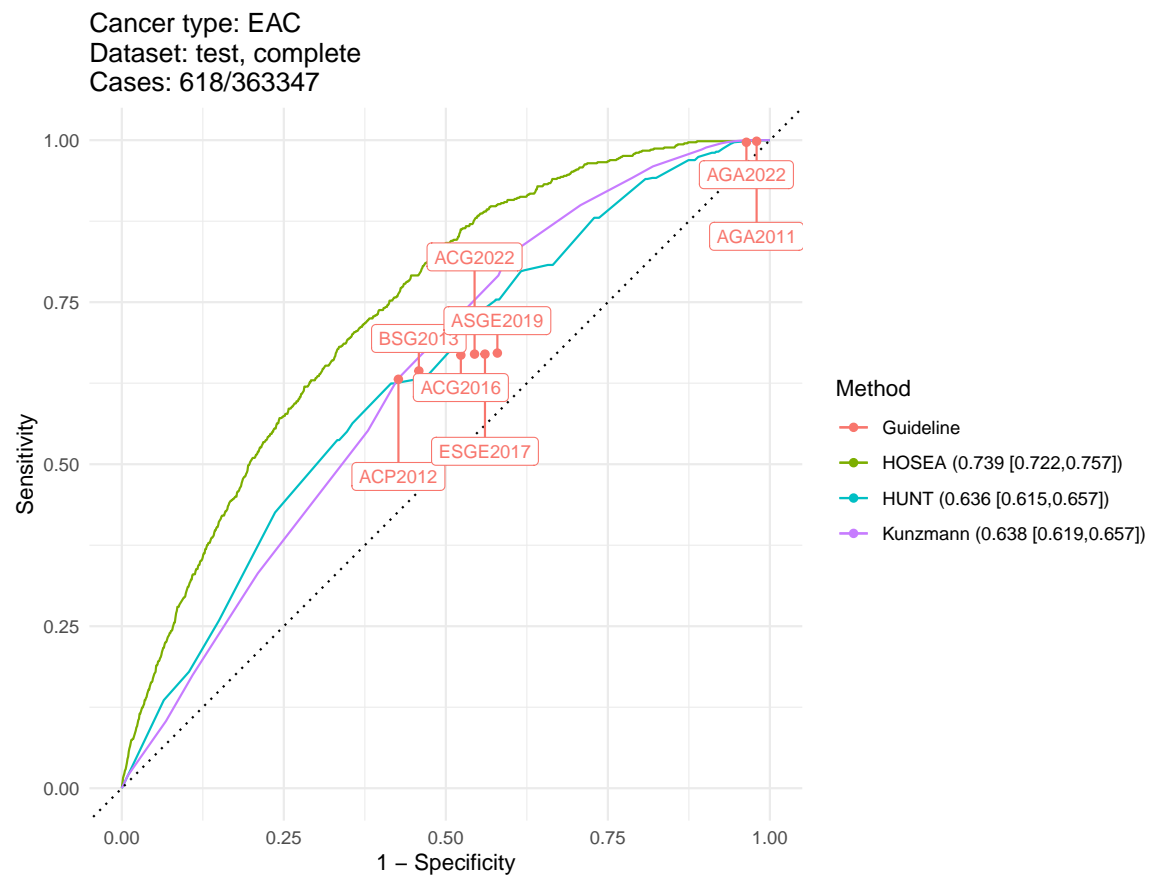
Cancer type: EGJAC
 Dataset: test, all
 Cases: 772/2567069



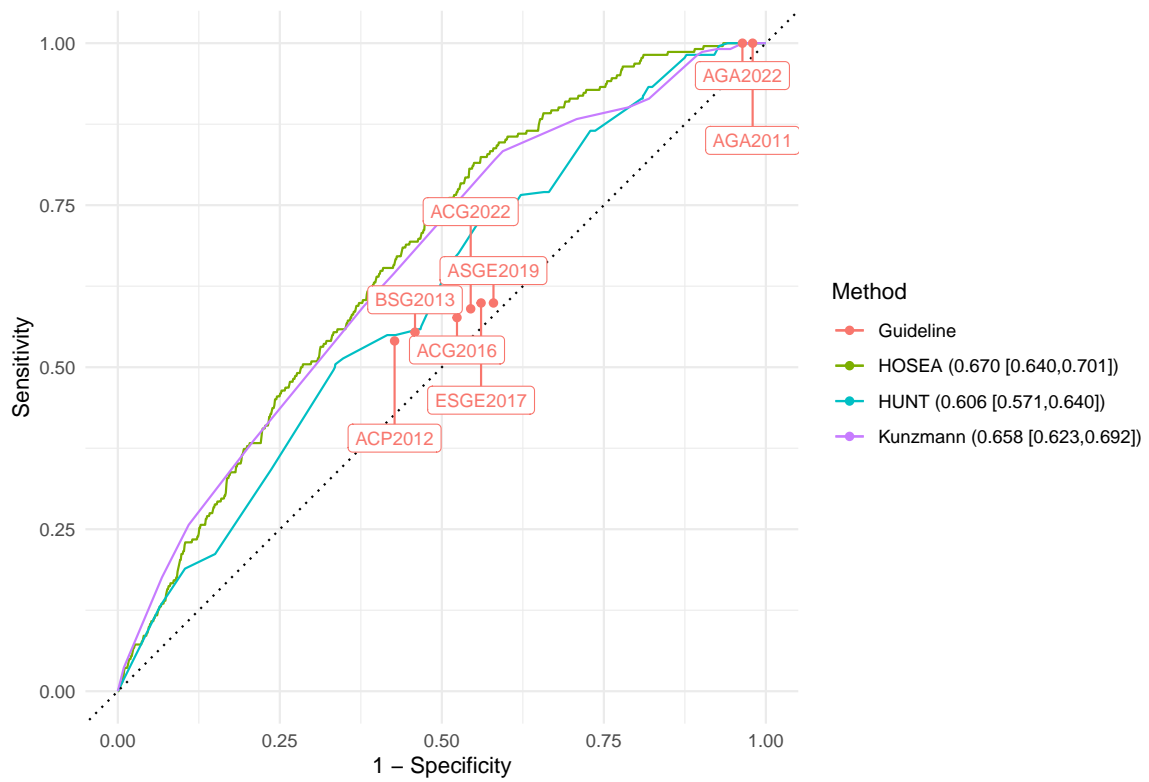
2.2 Complete test patients

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



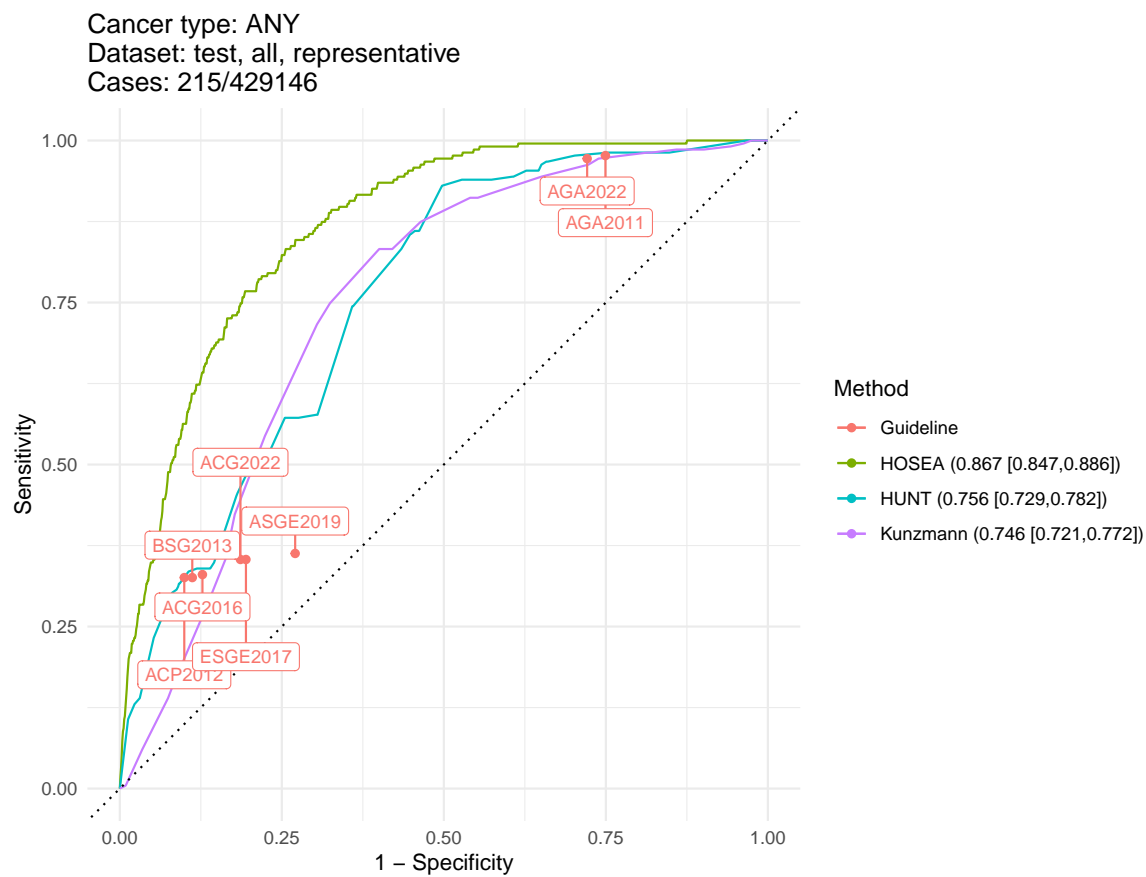


Cancer type: EGJAC
 Dataset: test, complete
 Cases: 222/363347

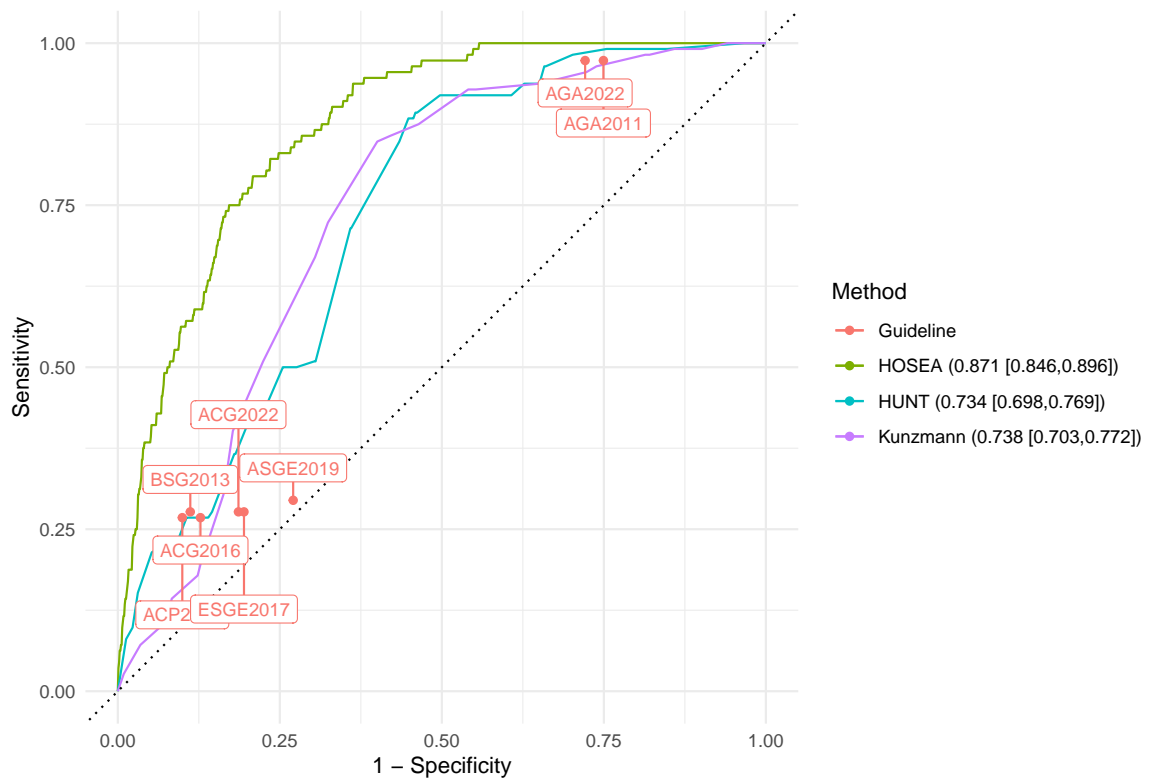


2.3 Representative sample

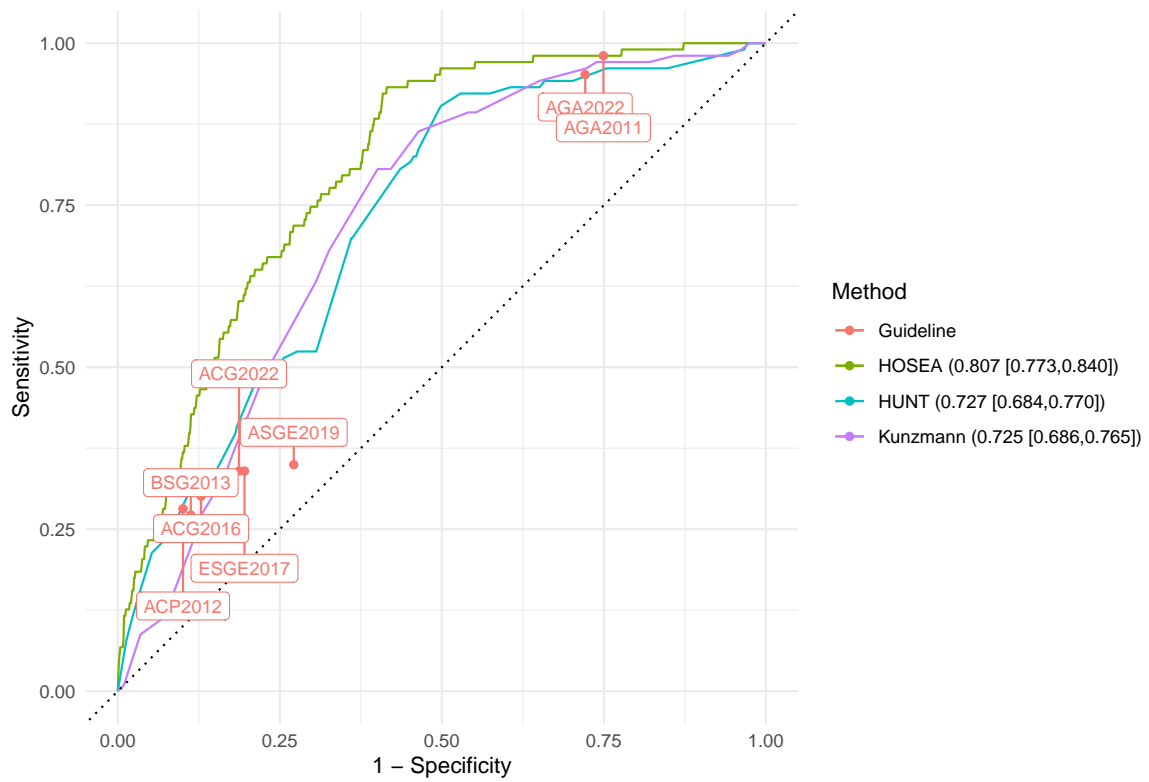
w.r.t. sex and prevalence ratio by sex



Cancer type: EAC
 Dataset: test, all, representative
 Cases: 112/429146

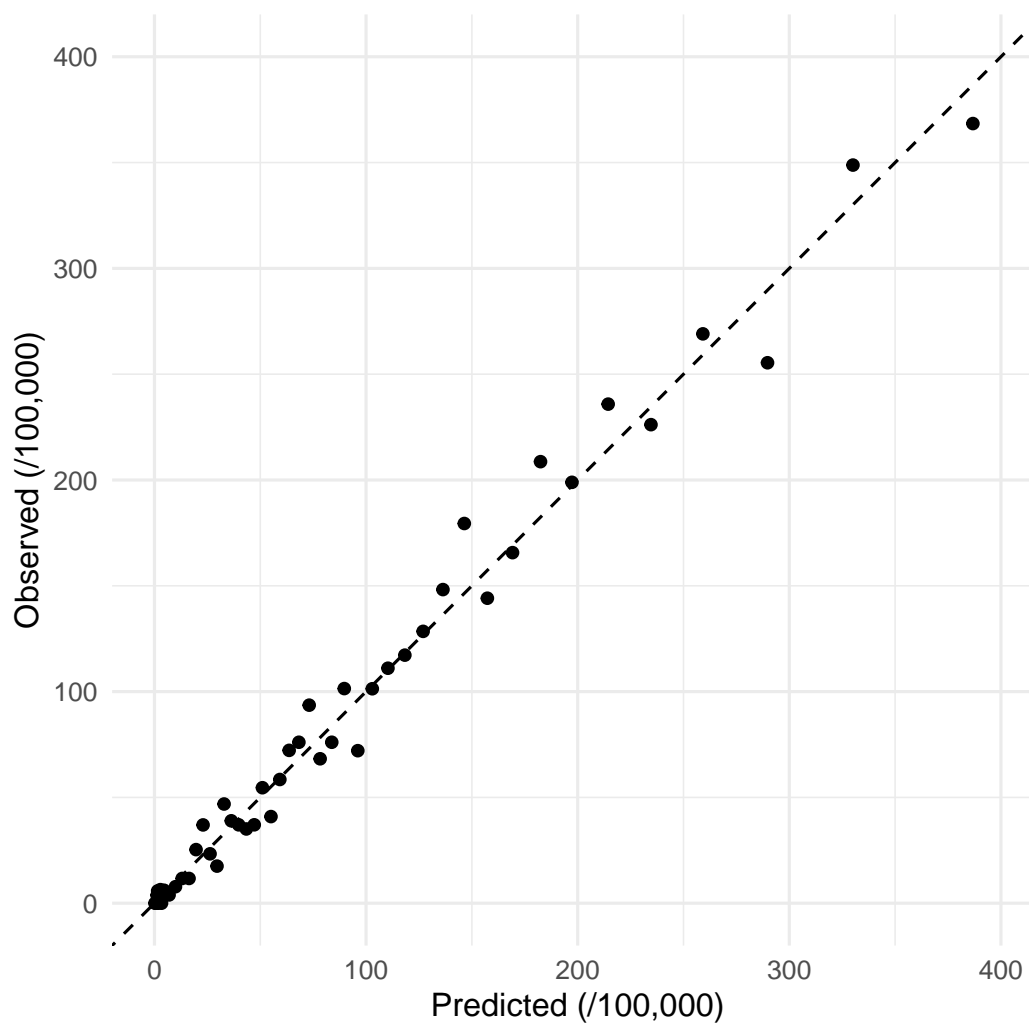


Cancer type: EGJAC
 Dataset: test, all, representative
 Cases: 103/429146

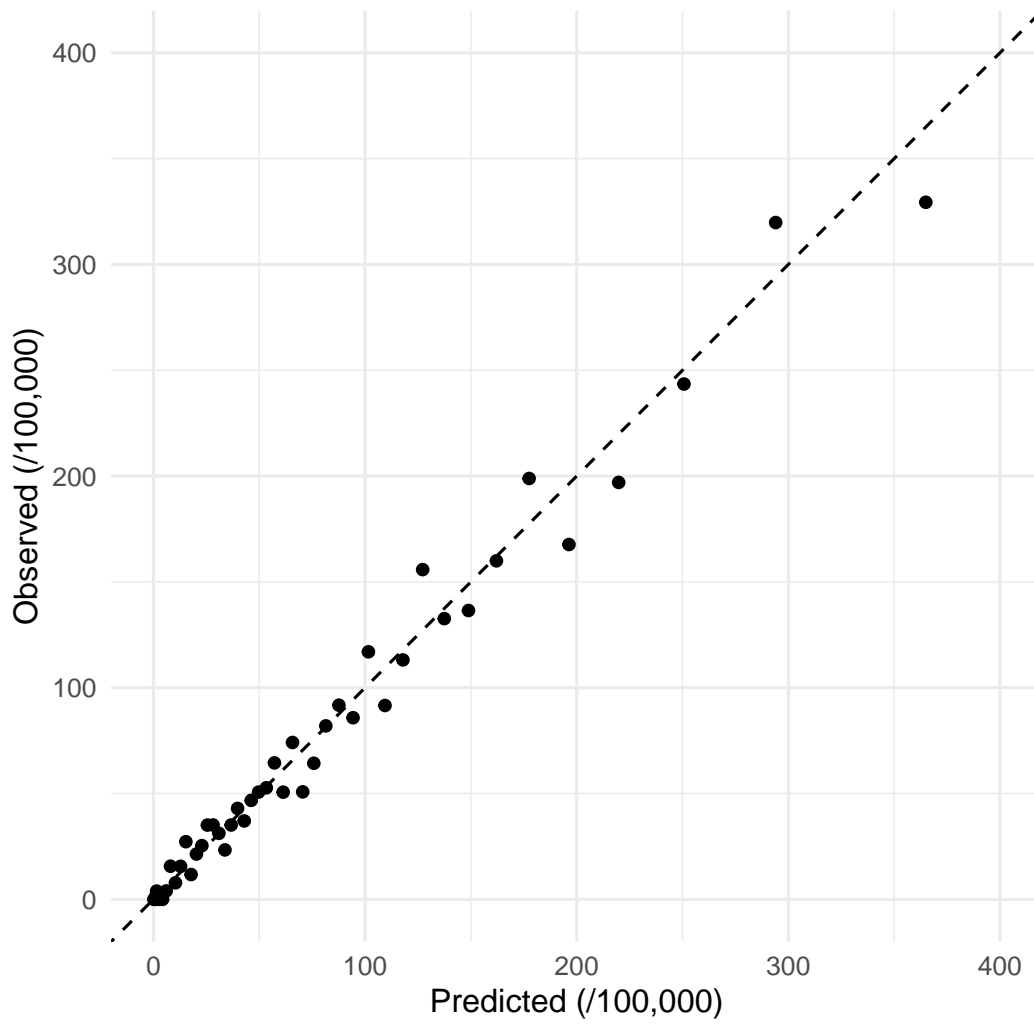


3 Calibration

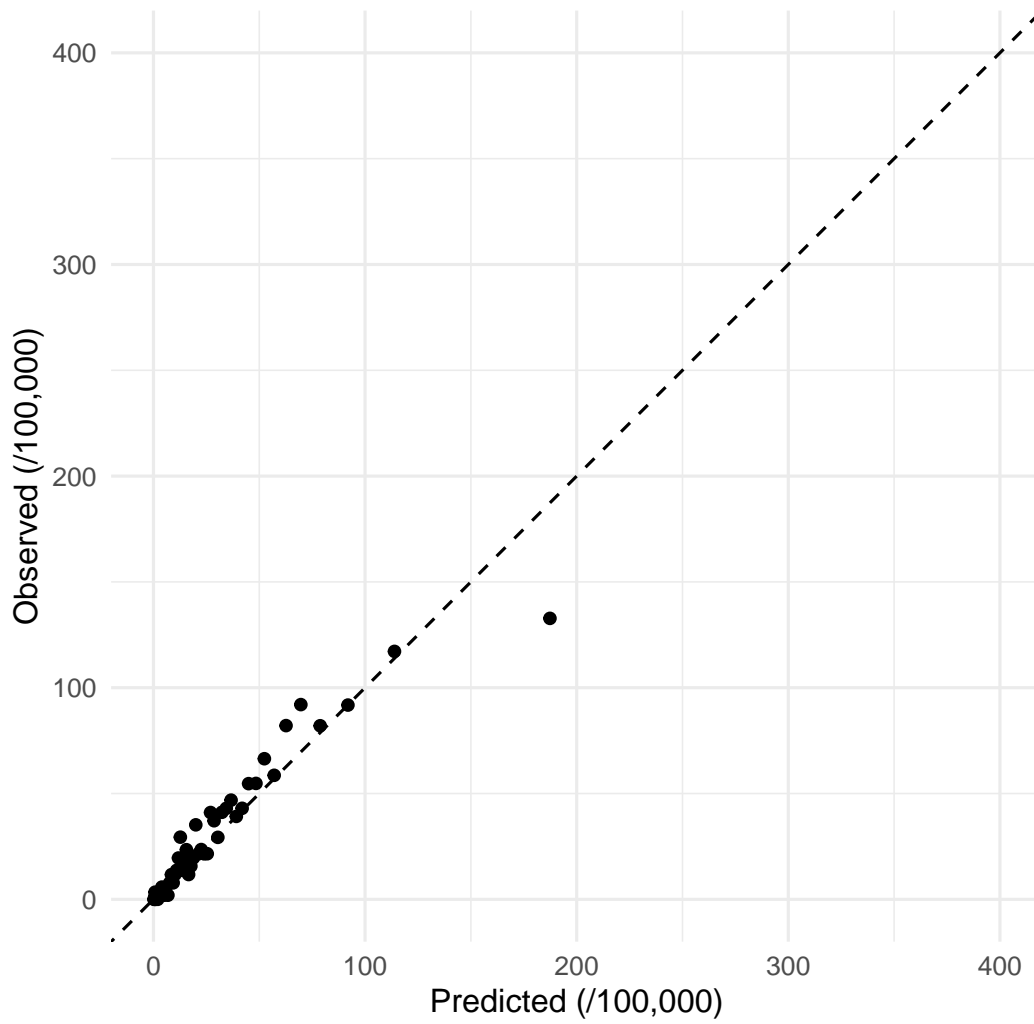
Cancer type: ANY
Dataset: test, all
Cases: 2848/2567069
HL: H=53.04, df=50, p=0.358



Cancer type: EAC
Dataset: test, all
Cases: 2076/2567069
HL: $H=44.41$, $df=50$, $p=0.696$



Cancer type: EGJAC
Dataset: test, all
Cases: 772/2567069
HL: H=64.74, df=50, p=0.079



4 Thresholds

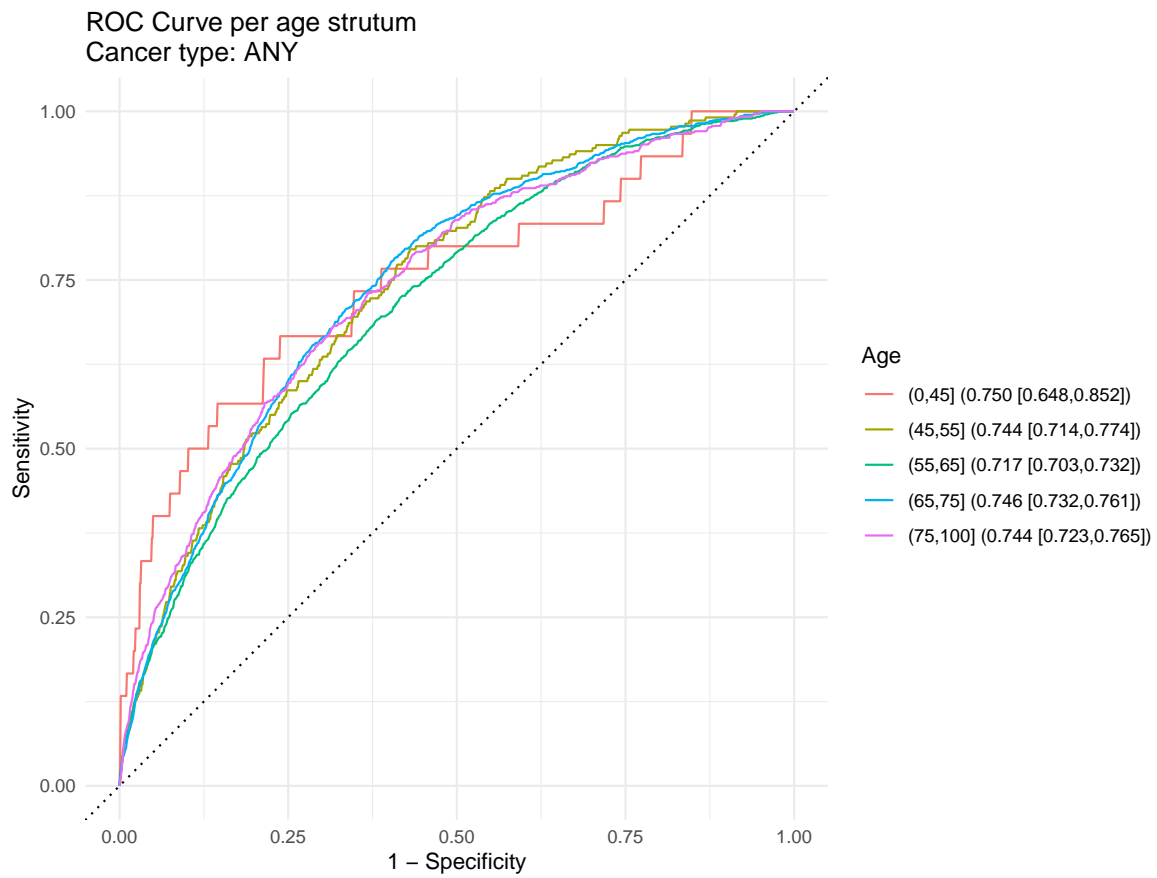
ANY Threshold	TPR	PPV	DetPrevalence
0	100.00	0.11	100.00
5	99.54	0.13	82.38
10	99.40	0.14	78.88
15	99.09	0.15	75.76
20	98.60	0.15	72.73
25	97.68	0.16	69.71
30	97.16	0.16	66.67
35	96.10	0.17	63.70
40	95.19	0.17	60.83
45	94.24	0.18	58.08
50	93.22	0.19	55.47
55	92.17	0.19	52.98
60	91.12	0.20	50.60
65	89.68	0.21	48.36
70	88.24	0.21	46.22
75	86.73	0.22	44.21
80	85.32	0.22	42.32
85	84.23	0.23	40.53
90	82.94	0.24	38.87
95	81.64	0.24	37.30
100	80.65	0.25	35.81
105	79.39	0.26	34.39
110	77.98	0.26	33.06
115	76.65	0.27	31.78
120	75.42	0.27	30.57
125	73.98	0.28	29.41
130	72.75	0.29	28.30
135	71.52	0.29	27.24
140	70.19	0.30	26.23
145	68.71	0.30	25.24
150	67.10	0.31	24.30
155	65.77	0.31	23.39
160	64.64	0.32	22.51
165	63.38	0.32	21.66
170	62.08	0.33	20.85
175	61.17	0.34	20.07
180	59.59	0.34	19.33
185	58.39	0.35	18.61
190	57.09	0.35	17.93
195	56.04	0.36	17.27
200	54.95	0.37	16.64
220	50.49	0.39	14.38
240	46.52	0.41	12.50
260	42.84	0.44	10.92
280	39.54	0.46	9.56
300	37.18	0.49	8.40
320	34.02	0.51	7.41
340	31.07	0.53	6.56
360	29.14	0.56	5.81
380	26.83	0.58	5.16
400	24.72	0.60	4.59
420	23.24	0.63	4.09
440	21.38	0.65	3.65
460	19.73	0.67	3.26
480	18.64	0.71	2.92
500	17.38	0.74	2.62
550	14.54	0.81	2.00
600	12.46	0.90	1.53
650	10.29	0.96	1.19
700	8.25	0.98	0.93
750	7.16	1.08	0.73
800	5.76	1.09	0.58
850	4.88	1.16	0.47
900	4.39	1.30	0.37
950	3.79	1.38	0.31
1000	3.41	1.52	0.25

EAC Threshold	TPR	PPV	DetPrevalence
0	100.00	0.08	100.00
5	99.71	0.10	79.96
10	99.18	0.11	75.25
15	98.55	0.11	71.17
20	97.59	0.12	67.18
25	96.15	0.12	63.30
30	94.75	0.13	59.59
35	93.55	0.13	56.09
40	92.05	0.14	52.76
45	90.37	0.15	49.65
50	88.87	0.15	46.76
55	86.95	0.16	44.05
60	84.97	0.17	41.56
65	83.33	0.17	39.26
70	81.79	0.18	37.16
75	80.39	0.18	35.24
80	78.85	0.19	33.45
85	77.31	0.20	31.80
90	75.53	0.20	30.23
95	73.70	0.21	28.77
100	72.11	0.21	27.39
105	70.42	0.22	26.08
110	68.64	0.22	24.82
115	67.29	0.23	23.63
120	65.70	0.24	22.50
125	64.07	0.24	21.43
130	62.19	0.25	20.41
135	60.16	0.25	19.43
140	58.67	0.26	18.50
145	57.42	0.26	17.62
150	56.02	0.27	16.78
155	54.62	0.28	16.00
160	53.42	0.28	15.27
165	51.73	0.29	14.57
170	50.43	0.29	13.90
175	49.37	0.30	13.27
180	47.54	0.30	12.69
185	45.86	0.31	12.13
190	44.89	0.31	11.61
195	43.93	0.32	11.10
200	42.82	0.33	10.62
220	38.97	0.35	8.96
240	35.50	0.38	7.59
260	32.23	0.40	6.47
280	28.37	0.42	5.52
300	25.72	0.44	4.74
320	23.03	0.46	4.08
340	20.86	0.48	3.52
360	19.08	0.51	3.04
380	17.15	0.53	2.64
400	15.90	0.56	2.29
420	14.60	0.59	1.99
440	12.96	0.60	1.73
460	11.85	0.63	1.52
480	10.69	0.65	1.33
500	9.78	0.68	1.16
550	7.66	0.73	0.85
600	5.92	0.77	0.62
650	4.53	0.79	0.47
700	3.85	0.90	0.35
750	3.32	1.01	0.27
800	2.46	0.98	0.20
850	1.83	0.95	0.16
900	1.49	1.01	0.12
950	1.25	1.08	0.09
1000	1.16	1.27	0.07

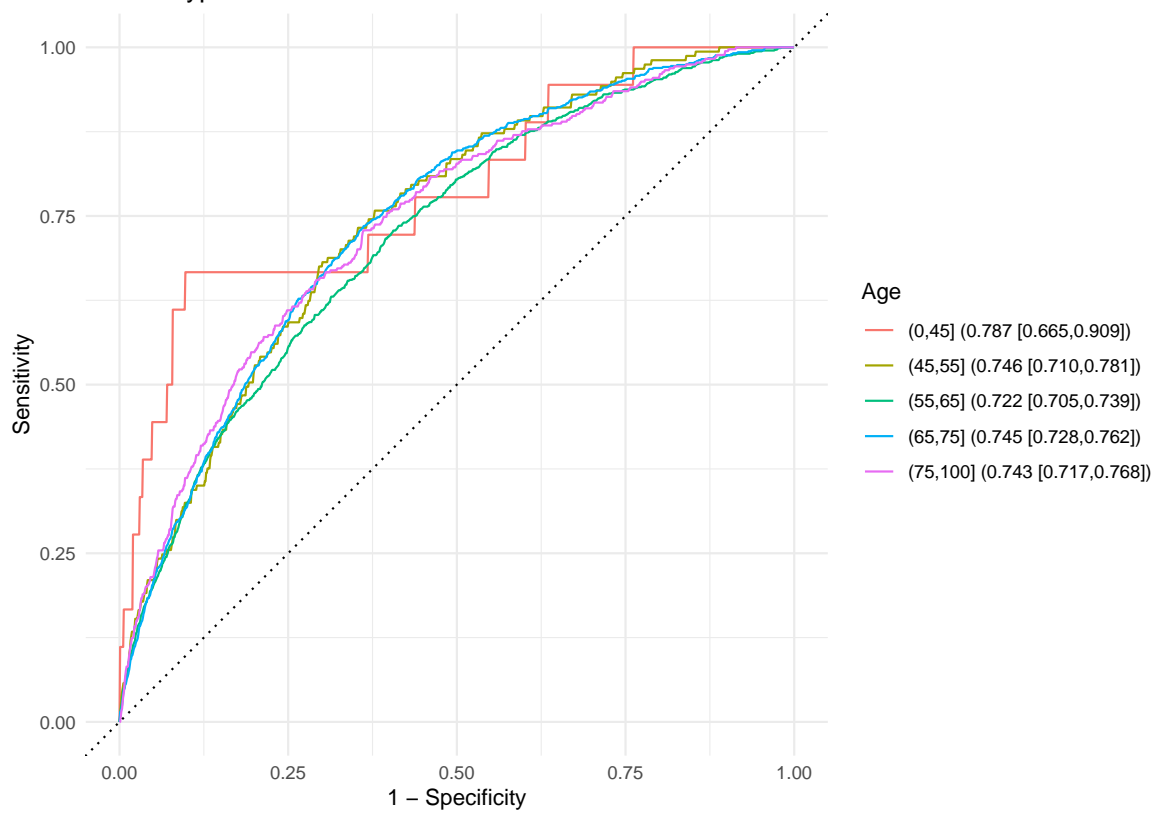
EGJAC Threshold	TPR	PPV	DetPrevalence
0	100.00	0.03	100.00
5	98.58	0.04	76.82
10	96.11	0.04	65.06
15	89.25	0.05	53.93
20	83.68	0.06	44.85
25	77.72	0.06	37.42
30	70.85	0.07	31.35
35	64.64	0.07	26.37
40	58.55	0.08	22.29
45	53.89	0.09	18.91
50	48.32	0.09	16.12
55	43.26	0.09	13.78
60	38.99	0.10	11.84
65	34.97	0.10	10.23
70	30.83	0.10	8.87
75	27.85	0.11	7.71
80	25.26	0.11	6.73
85	22.54	0.11	5.90
90	19.56	0.11	5.19
95	18.01	0.12	4.58
100	16.84	0.12	4.06
105	15.41	0.13	3.60
110	14.51	0.14	3.20
115	12.82	0.14	2.86
120	11.40	0.13	2.55
125	10.23	0.13	2.28
130	8.94	0.13	2.05
135	8.16	0.13	1.84
140	7.38	0.13	1.66
145	6.87	0.14	1.50
150	6.35	0.14	1.36
155	5.96	0.15	1.23
160	5.83	0.16	1.12
165	5.05	0.15	1.02
170	4.79	0.16	0.92
175	4.27	0.15	0.84
180	4.27	0.17	0.77
185	4.15	0.18	0.70
190	4.02	0.19	0.64
195	3.63	0.19	0.59
200	3.37	0.19	0.54
220	2.46	0.19	0.39
240	2.07	0.22	0.28
260	1.55	0.22	0.21
280	1.17	0.22	0.16
300	0.78	0.20	0.12
320	0.78	0.26	0.09
340	0.39	0.17	0.07
360	0.39	0.21	0.05
380	0.39	0.27	0.04
400	0.39	0.34	0.03
420	0.13	0.14	0.03
440	0.13	0.18	0.02
460	0.13	0.22	0.02
480	0.13	0.28	0.01
500	0.13	0.34	0.01
550	0.13	0.53	0.01
600	0.00	0.00	0.00
650	0.00	0.00	0.00
700	0.00	0.00	0.00
750	0.00	0.00	0.00
800	0.00	0.00	0.00
850	0.00	0.00	0.00
900	0.00	0.00	0.00
950	0.00	0.00	0.00
1000	0.00	0.00	0.00

5 Stratification by identity groups

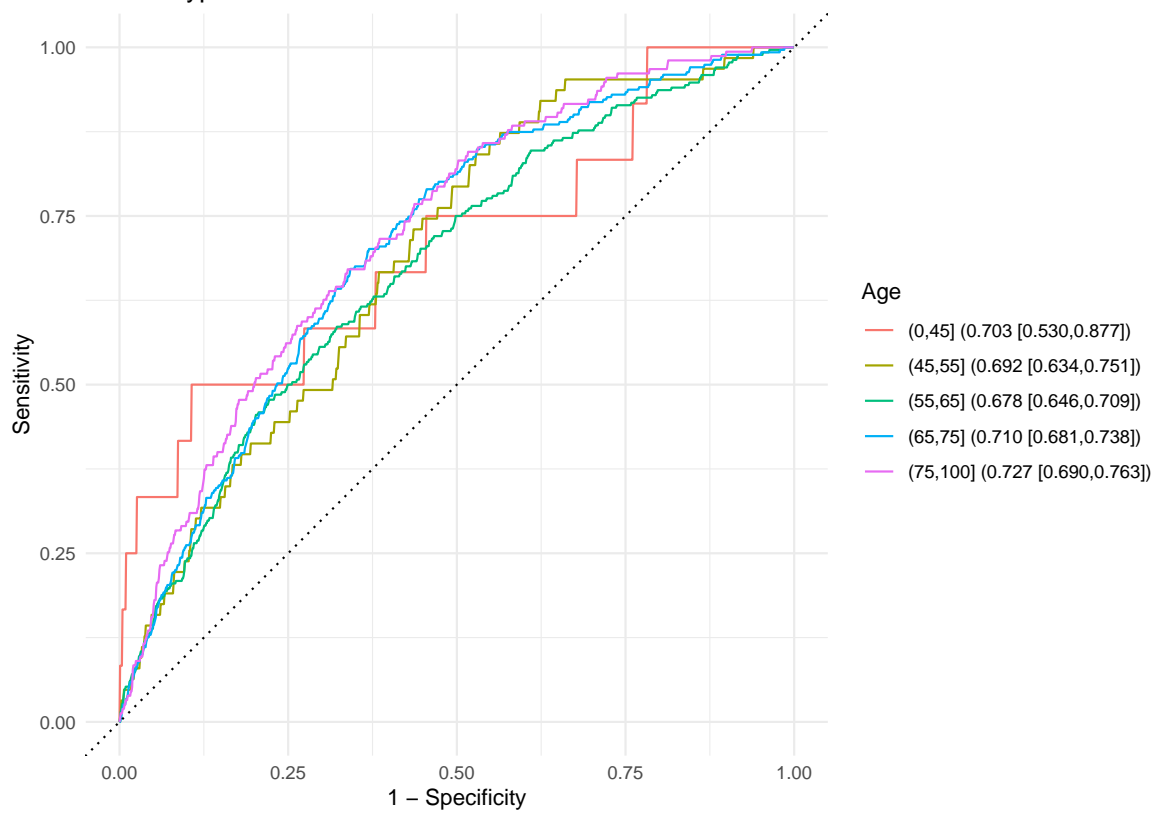
5.1 Age



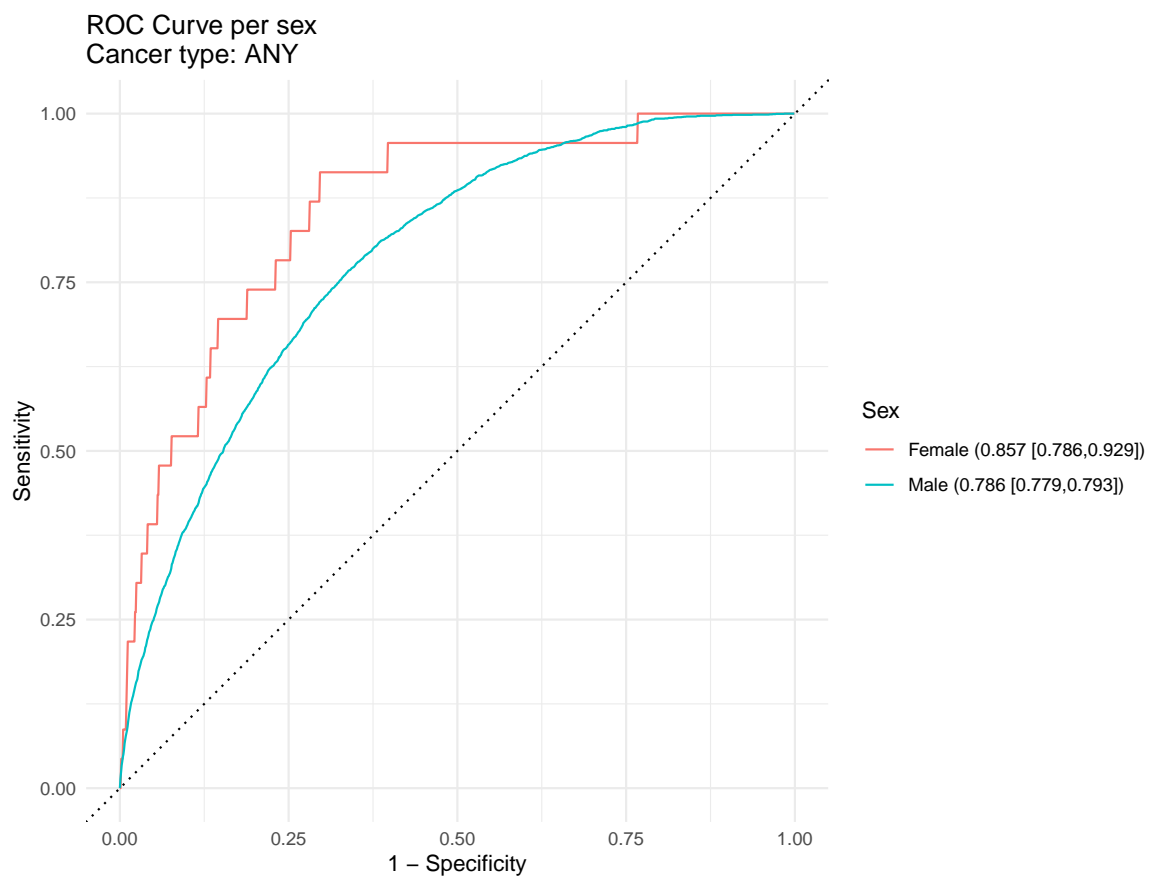
ROC Curve per age stratum
Cancer type: EAC

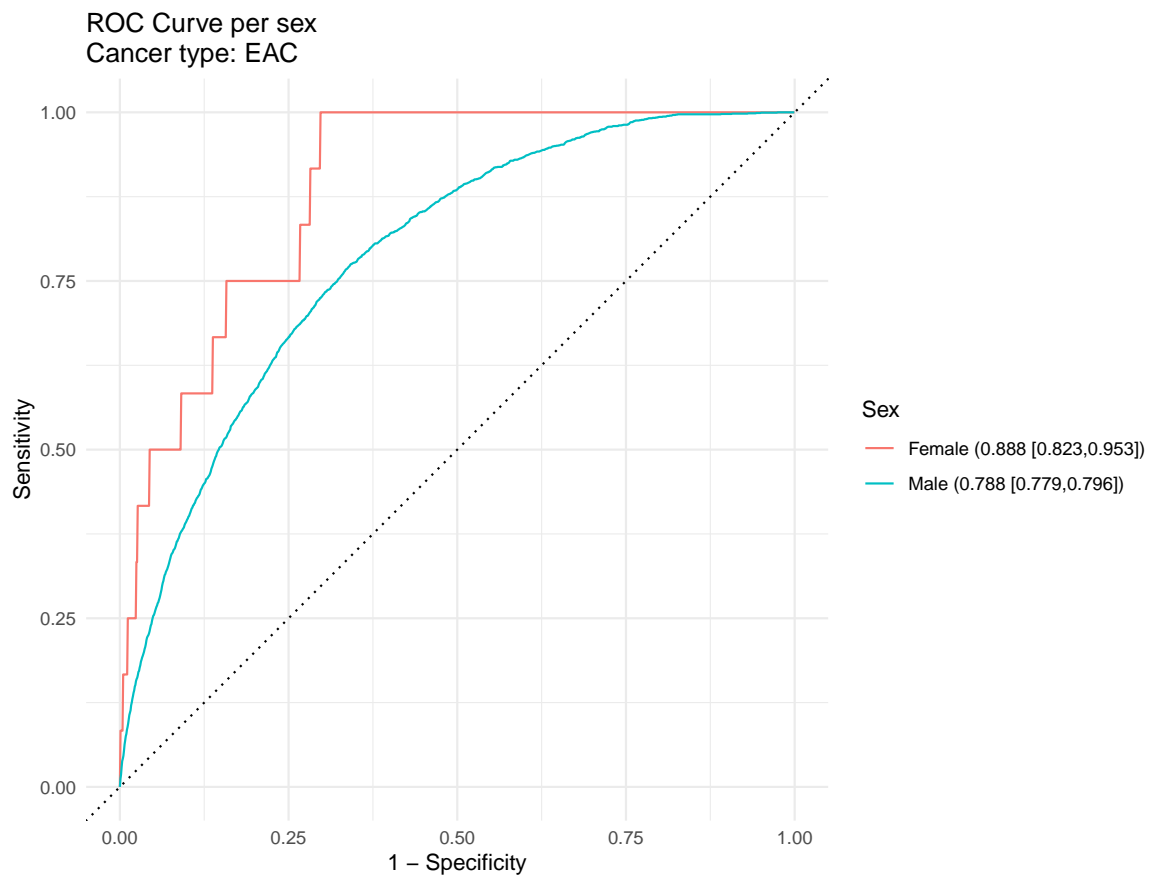


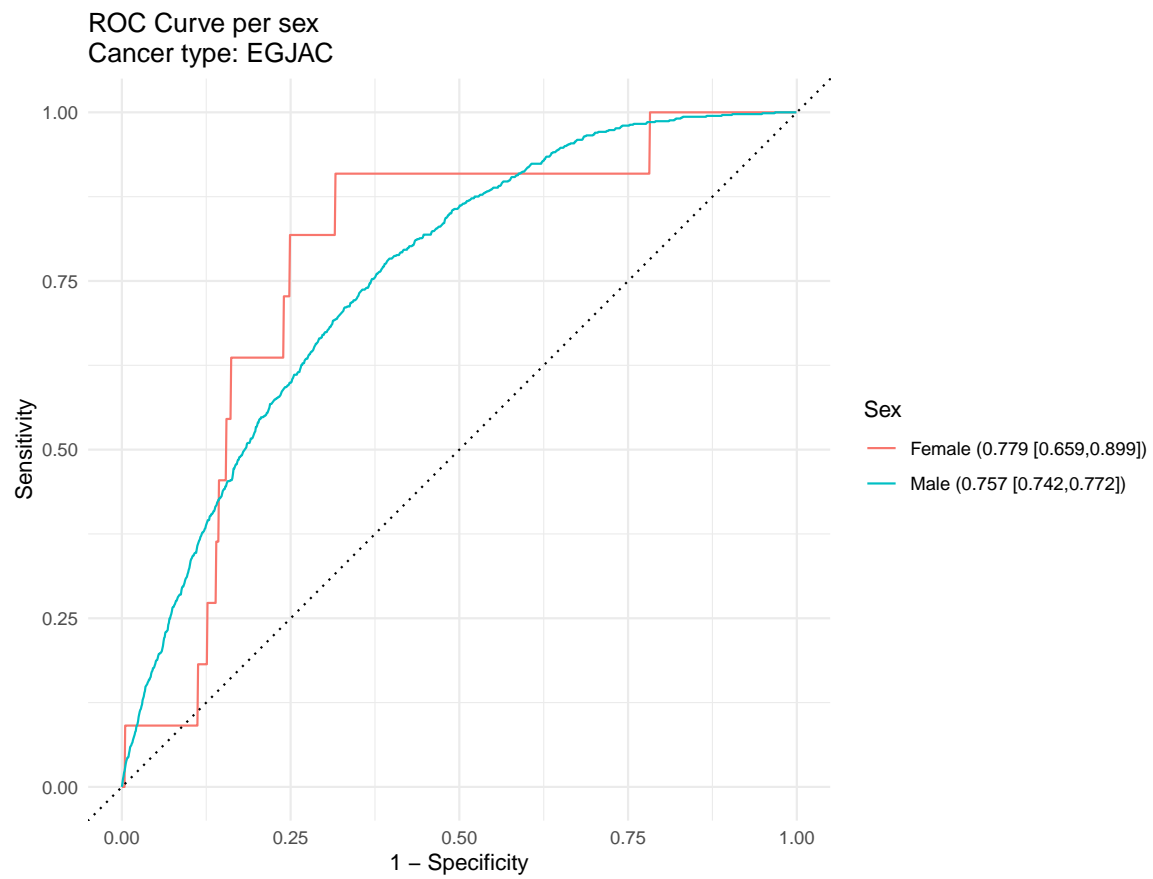
ROC Curve per age stratum
Cancer type: EGJAC



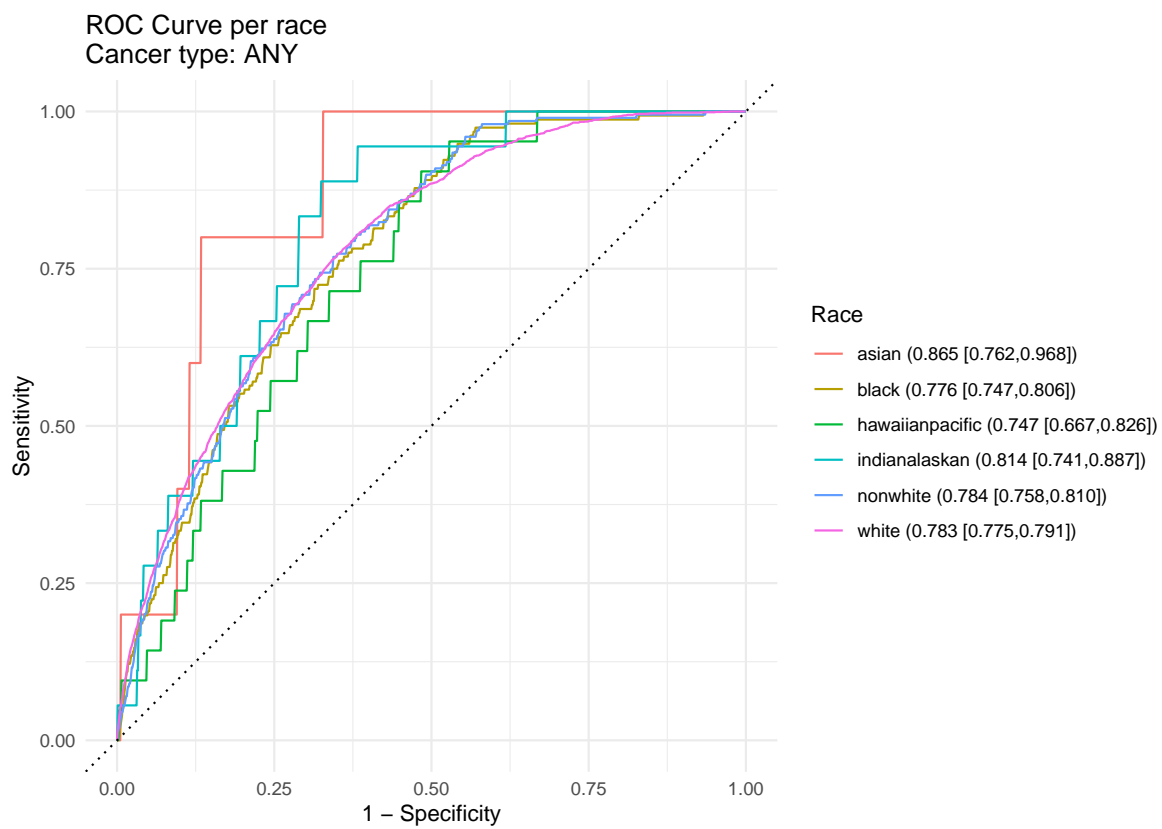
5.2 Sex

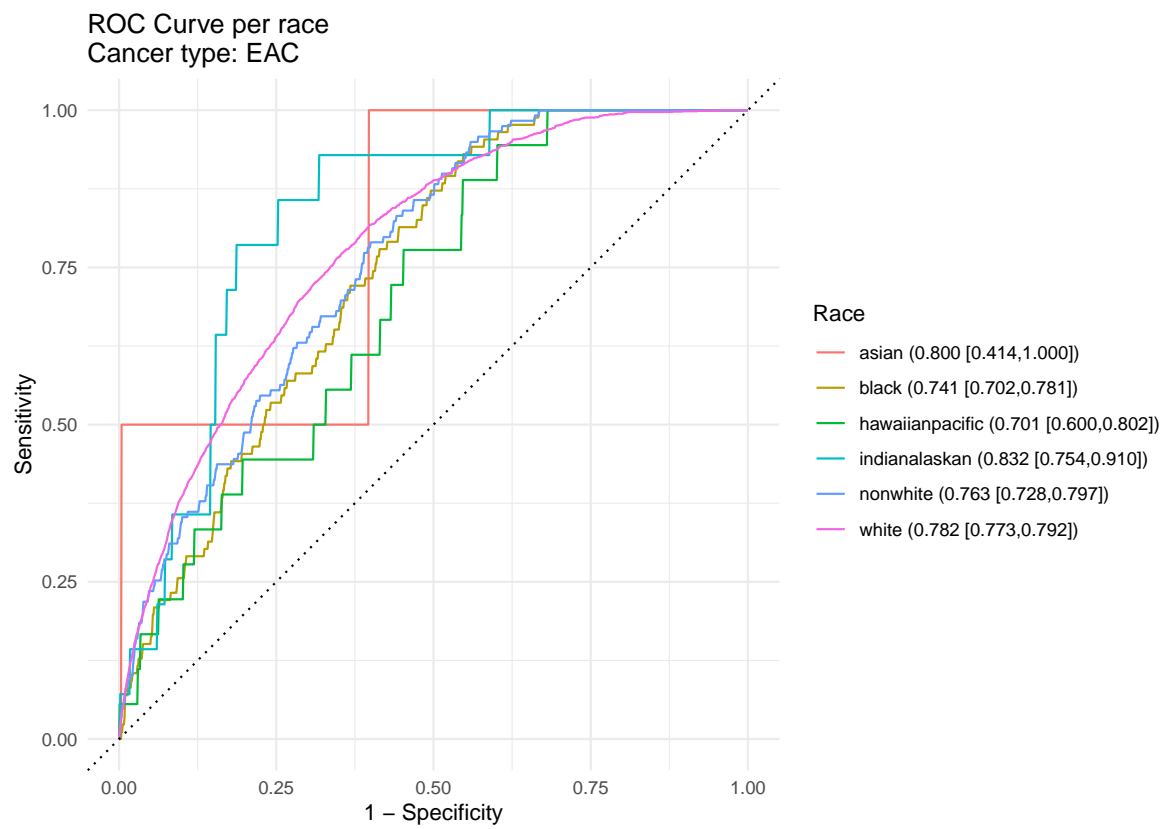


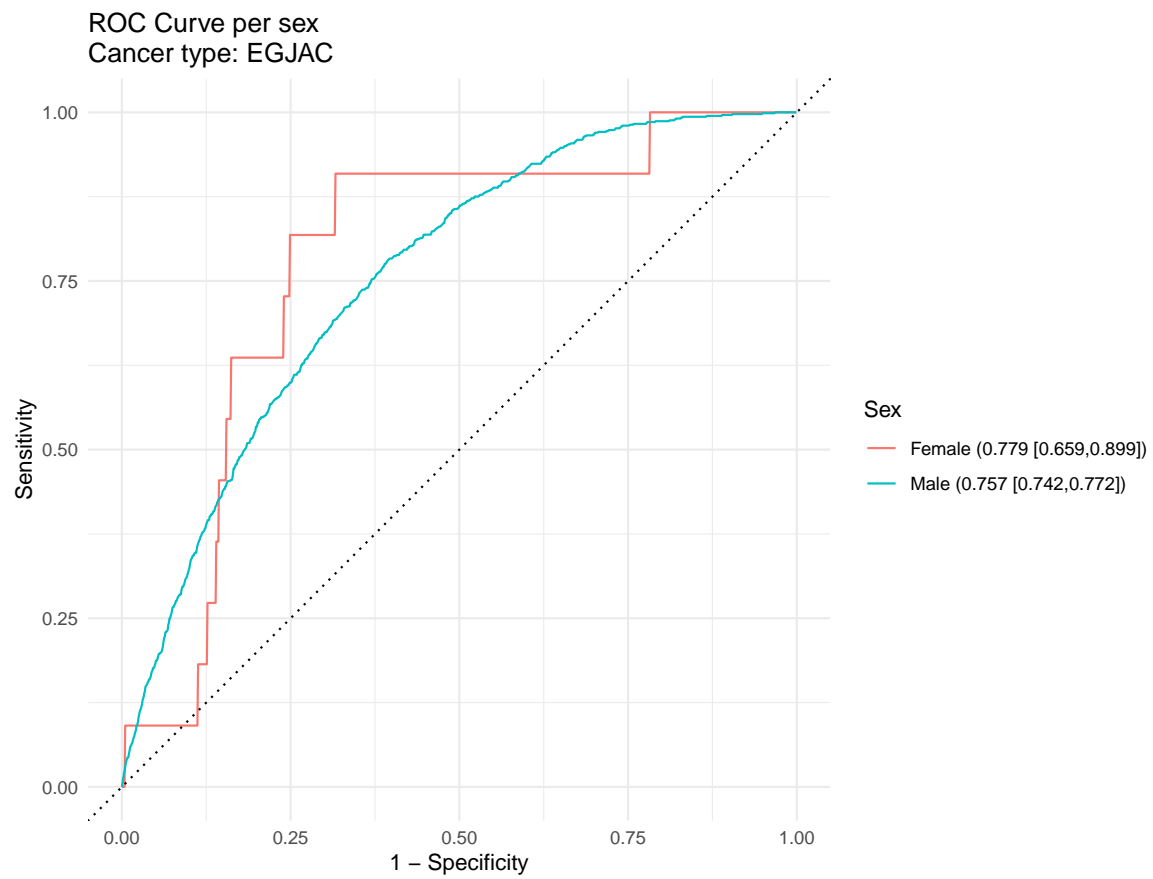




5.3 Race







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.

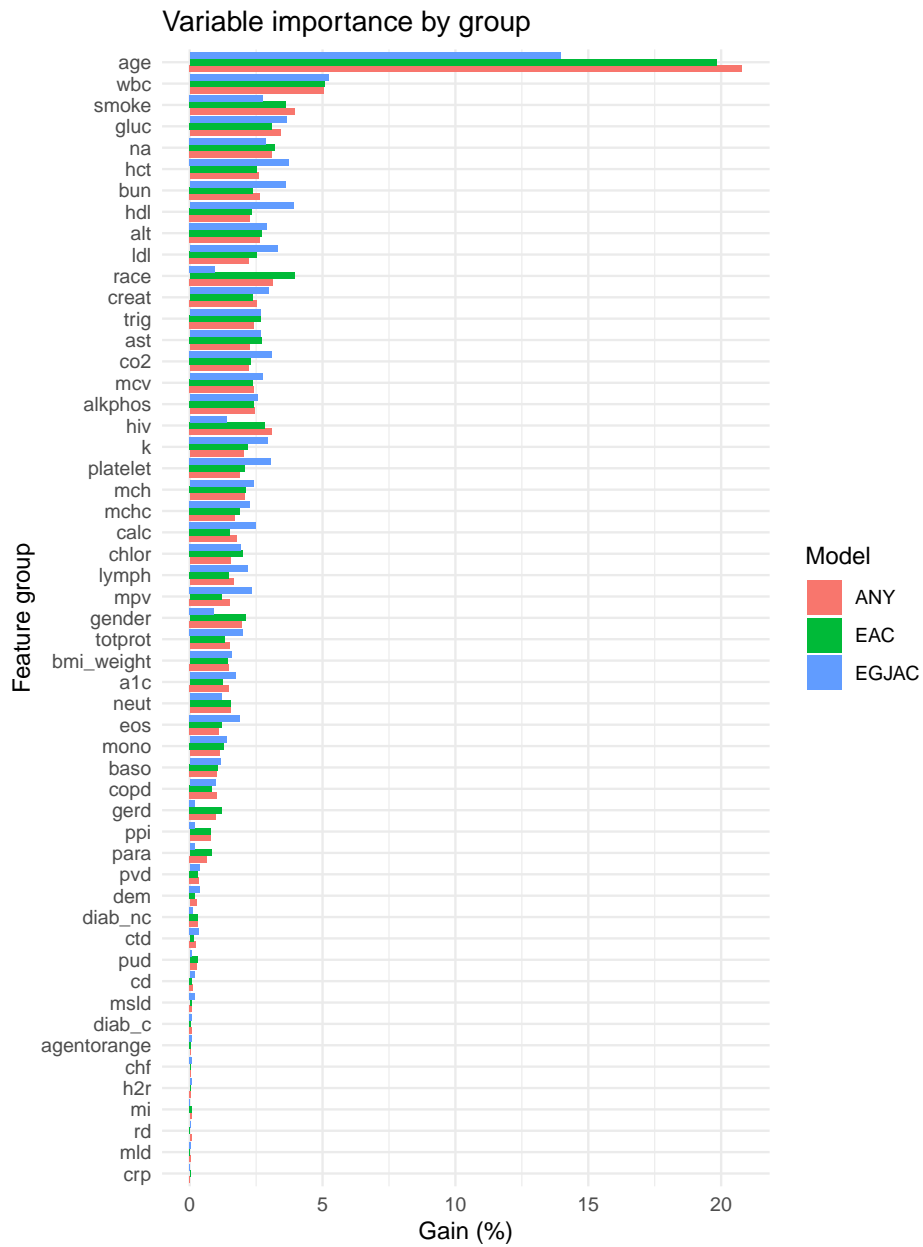
ANY		
Stage	Test. AUC	Nb. cases
Any	0.799 [0.792,0.807]	2818
I	0.831 [0.815,0.847]	495
II	0.813 [0.796,0.831]	435
III	0.803 [0.791,0.816]	878
IV	0.793 [0.782,0.804]	1244
I+	0.801 [0.793,0.808]	2448
II+	0.796 [0.788,0.804]	2169
III+	0.794 [0.785,0.803]	1950
IV+	0.793 [0.782,0.804]	1244

EAC		
Stage	Test. AUC	Nb. cases
Any	0.803 [0.794,0.811]	2054
I	0.838 [0.820,0.856]	350
II	0.816 [0.795,0.837]	302
III	0.804 [0.790,0.819]	650
IV	0.797 [0.785,0.810]	913
I+	0.803 [0.794,0.811]	1813
II+	0.797 [0.788,0.807]	1608
III+	0.796 [0.786,0.806]	1451
IV+	0.797 [0.785,0.810]	913

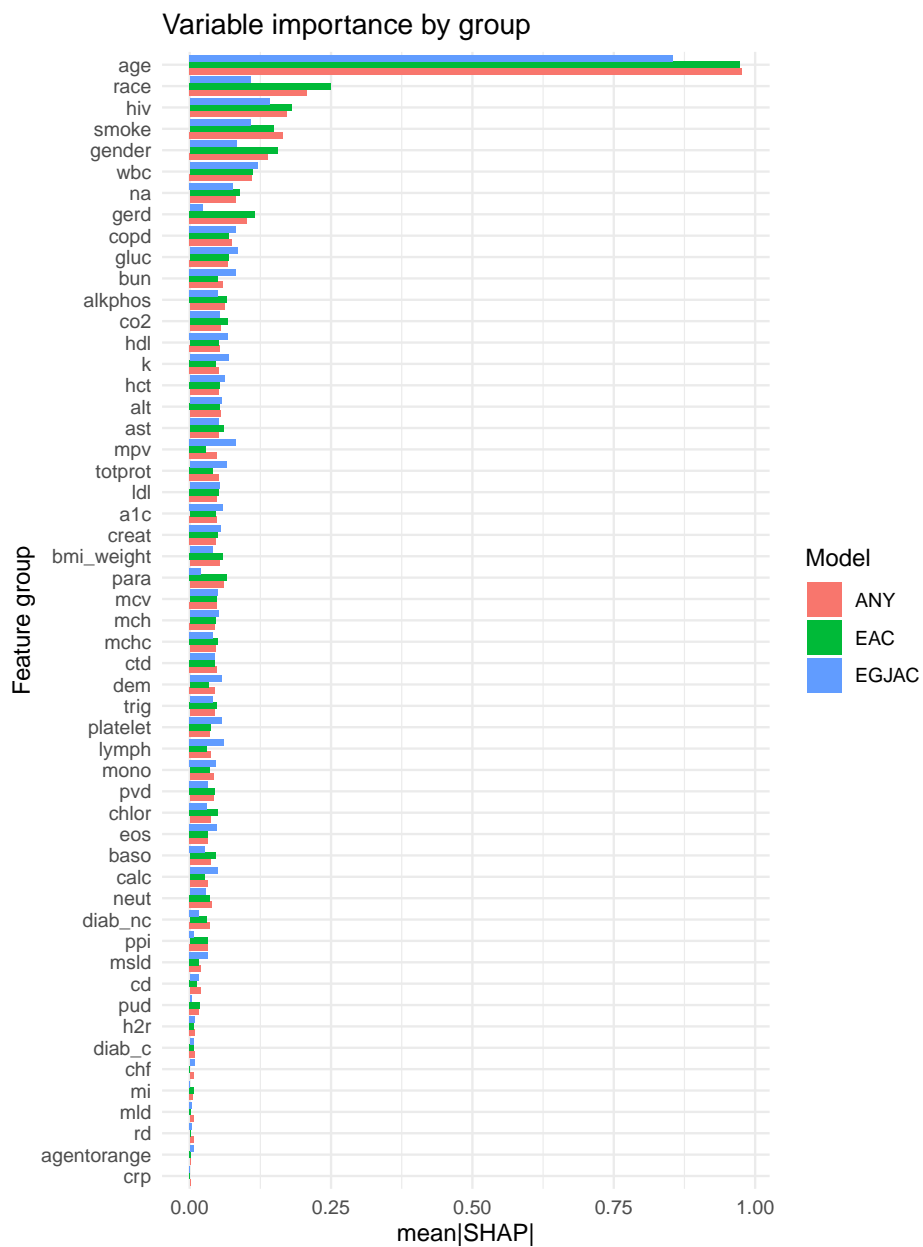
EGJAC		
Stage	Test. AUC	Nb. cases
Any	0.768 [0.754,0.782]	764
I	0.789 [0.756,0.821]	145
II	0.788 [0.752,0.824]	133
III	0.770 [0.744,0.796]	228
IV	0.762 [0.740,0.785]	331
I+	0.770 [0.754,0.786]	635
II+	0.767 [0.750,0.784]	561
III+	0.763 [0.745,0.781]	499
IV+	0.762 [0.740,0.785]	331

7 Variable importance

7.1 Gain VI



7.2 SHAP VI



8 Years prior

