HOSEA Aim I – Report (SRS)

Simon Fontaine

October 31, 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**

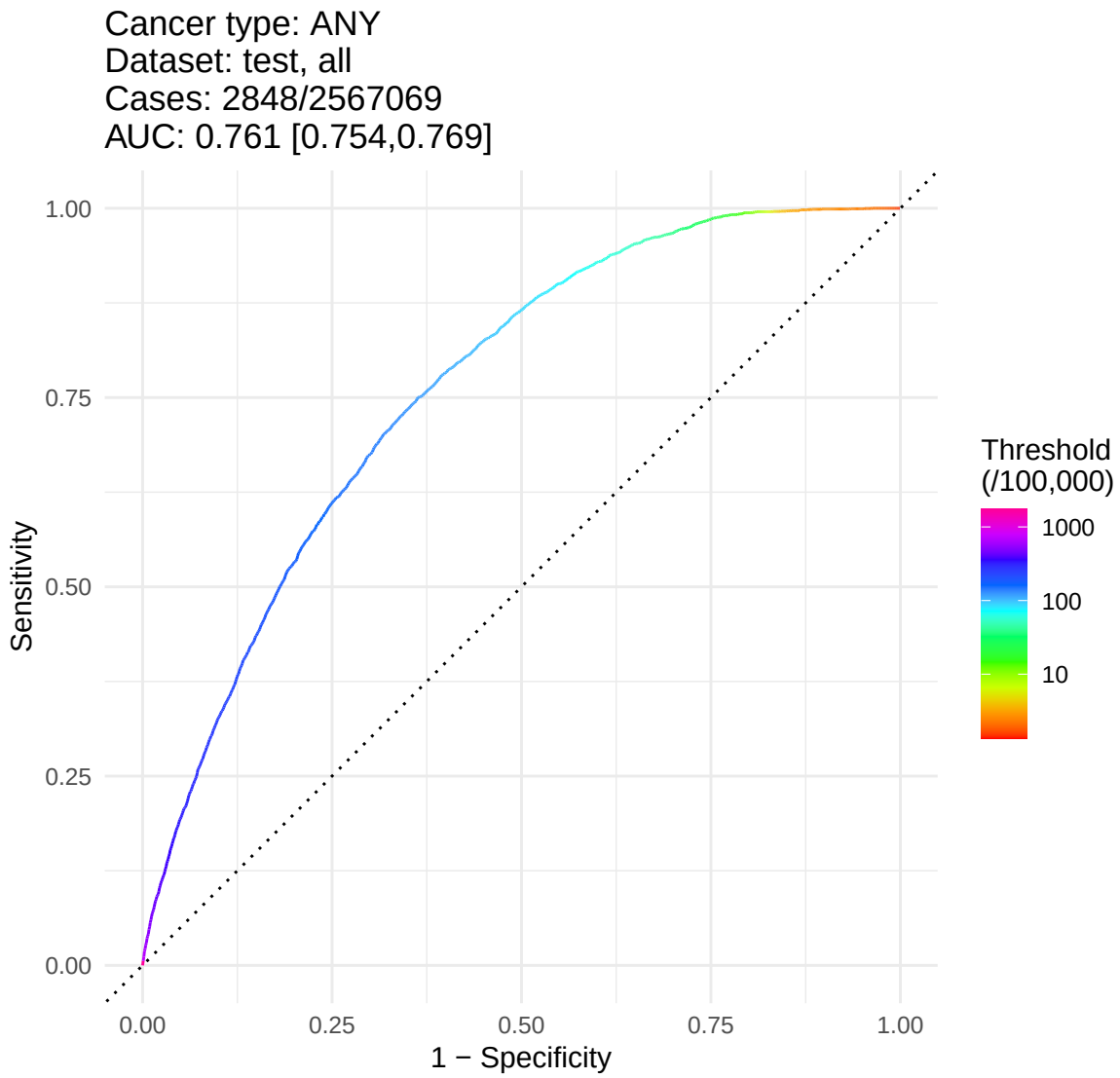
5 Stratification by identity groups **21**
5.1 Age . . . 21
5.2 Sex . . . 24
5.3 Race . . . 27

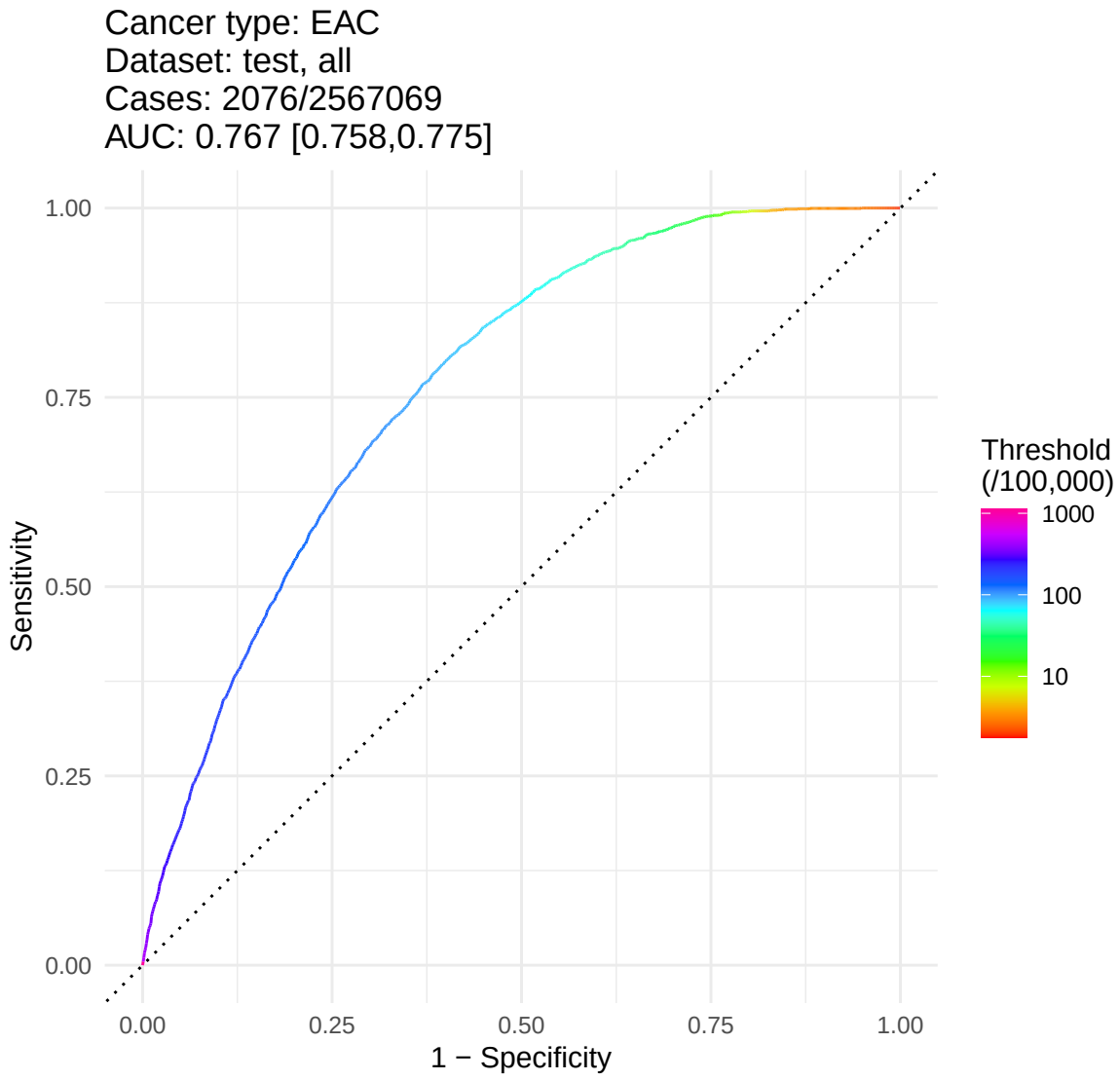
6 Cancer stage **30**

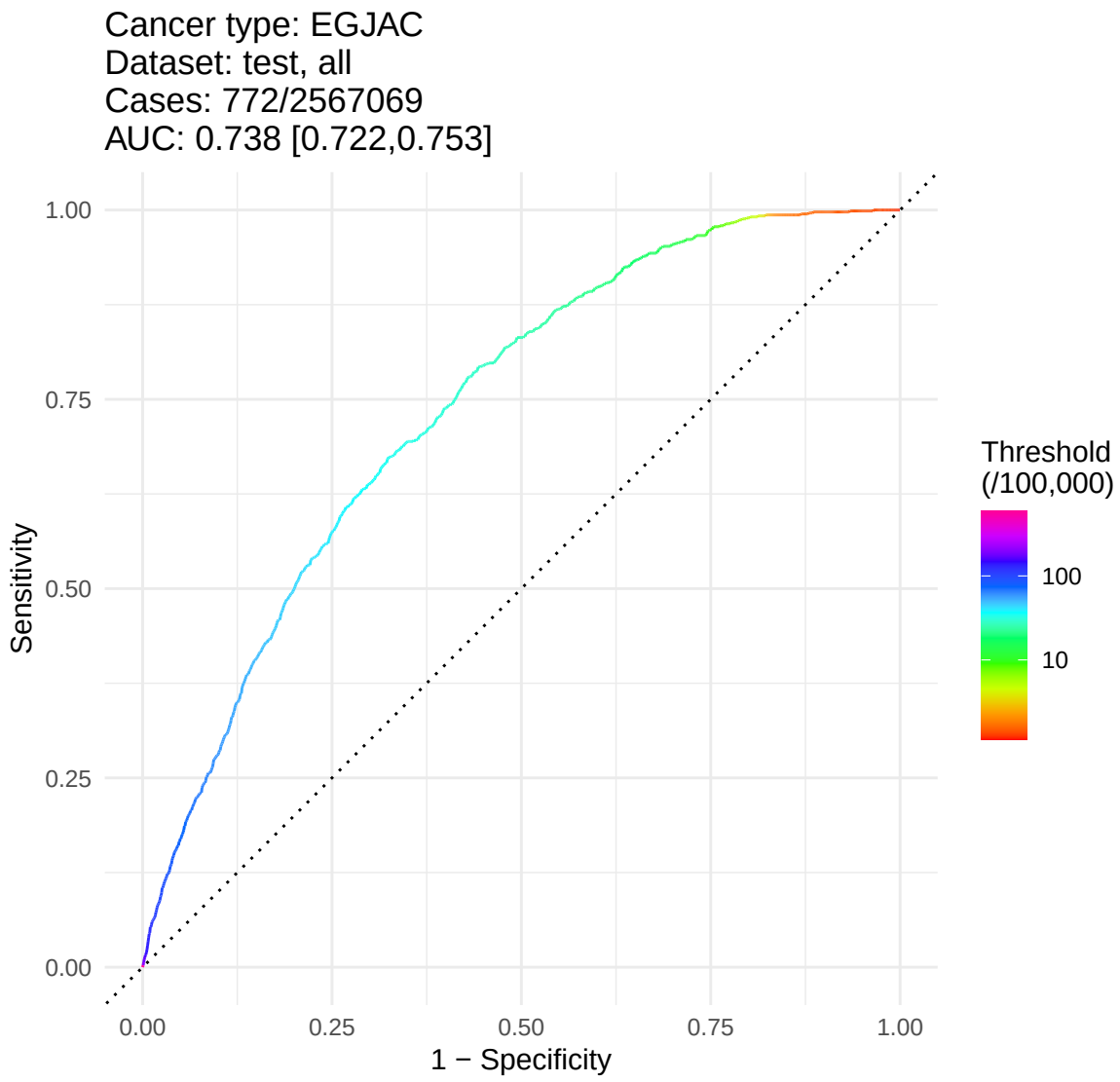
7 Variable importance **31**
7.1 Gain VI . . . 31
7.2 SHAP VI . . . 33

8 Years prior **35**

1 Overall performance

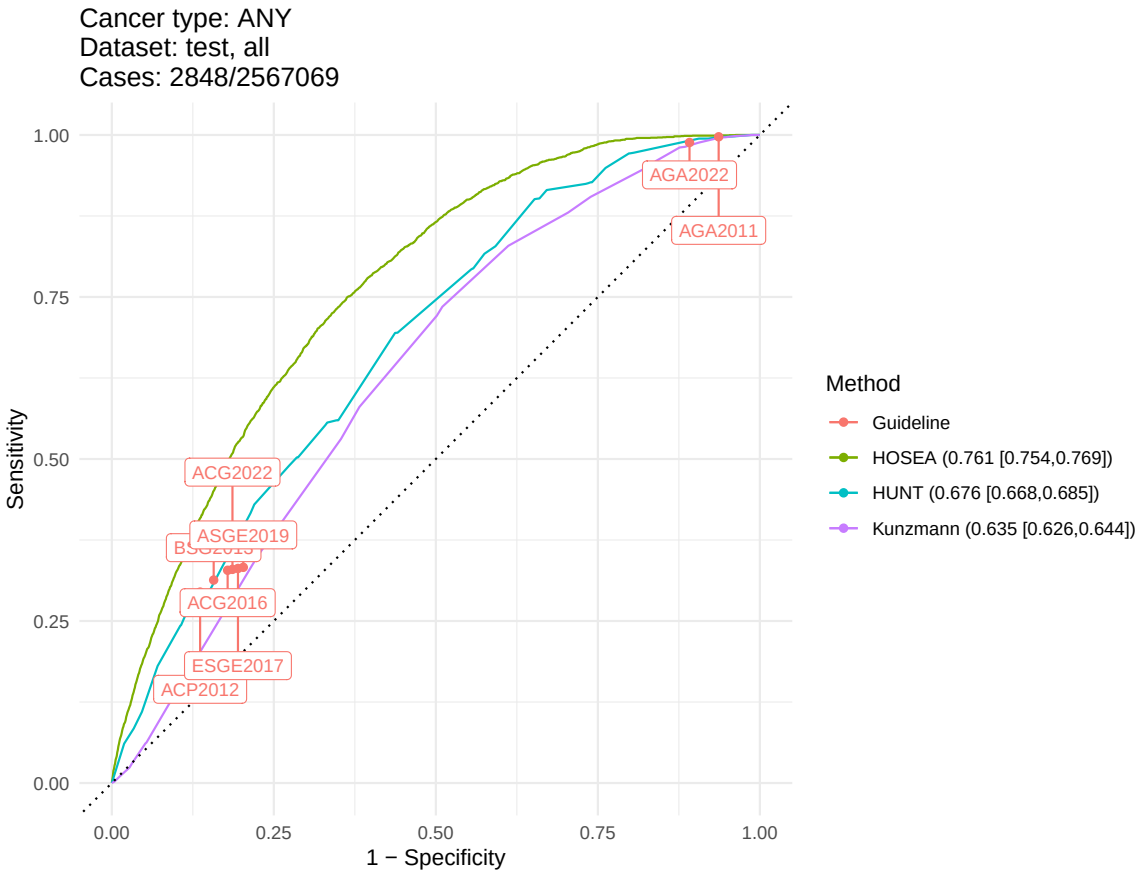


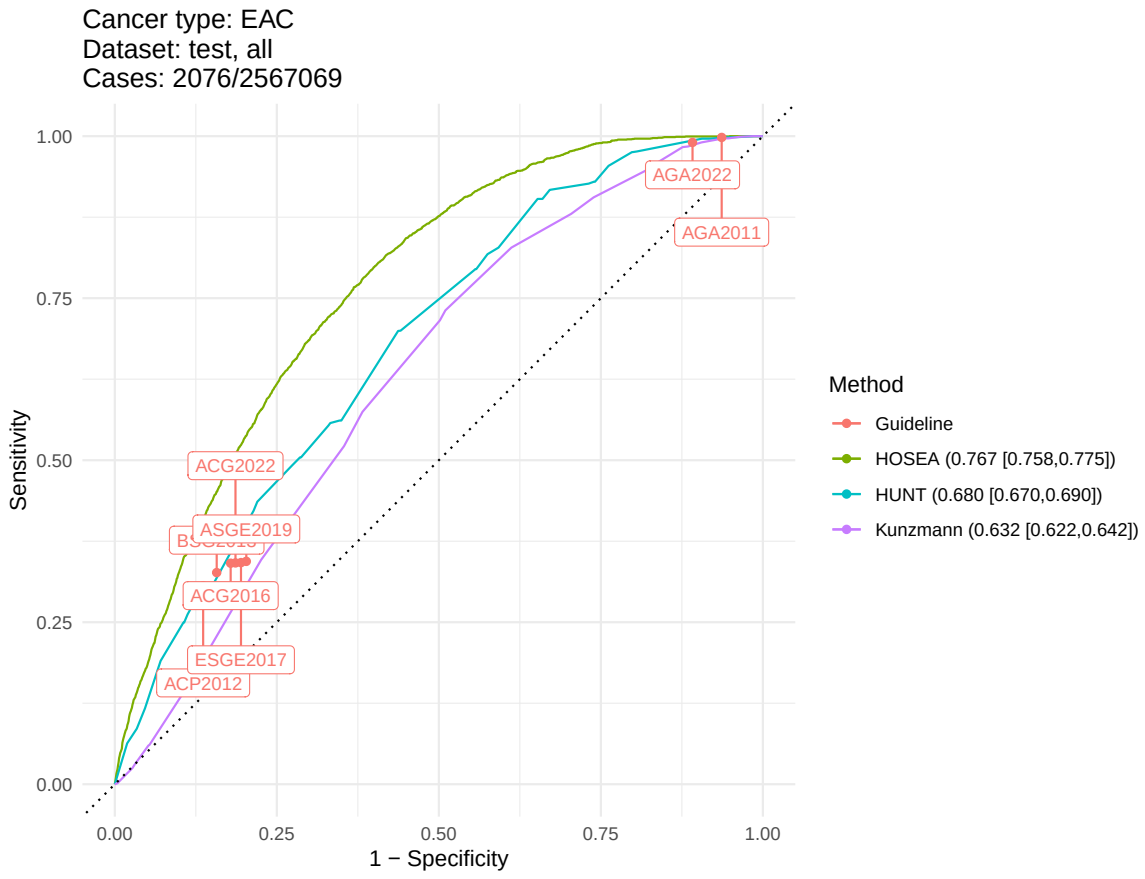

Cancer type: EAC
Dataset: test, all
Cases: 2076/2567069
AUC: 0.767 [0.758,0.775]
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Threshold
(/100,000)
1000
100
10

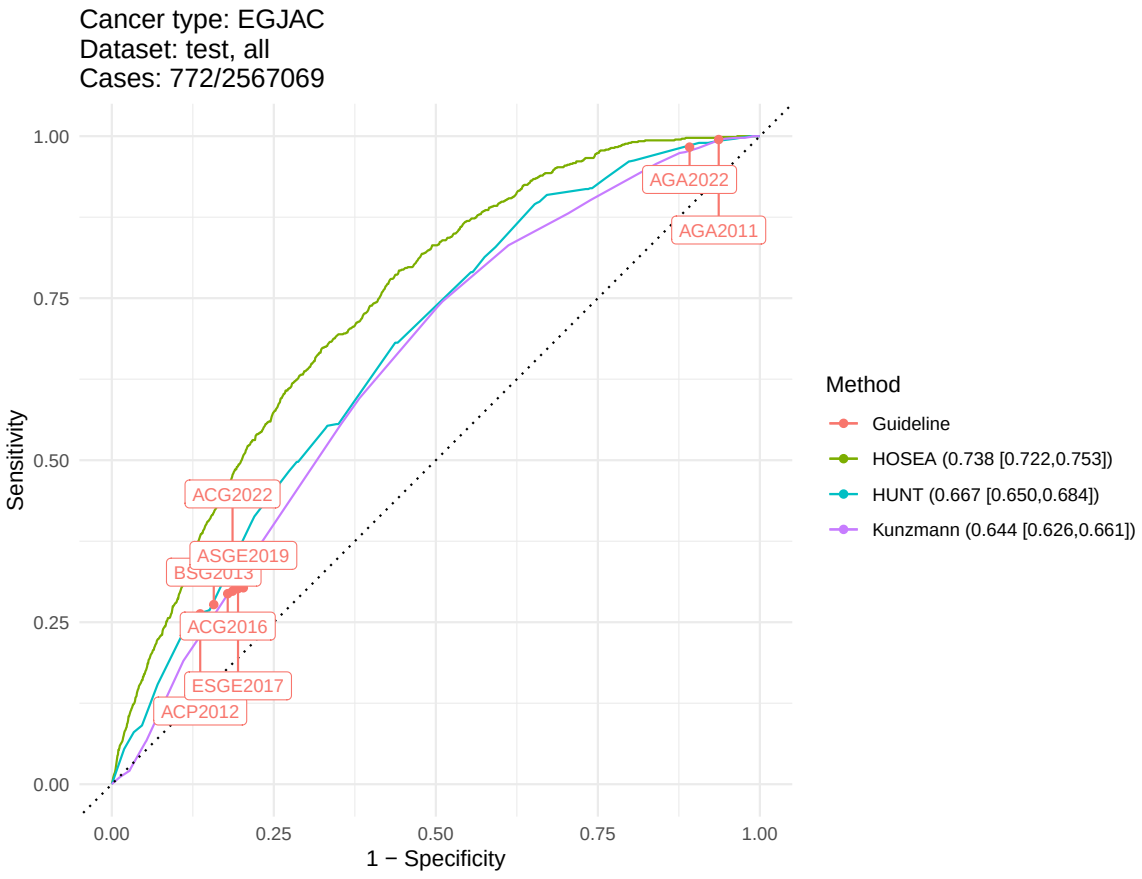

Cancer type: EGJAC
Dataset: test, all
Cases: 772/2567069
AUC: 0.738 [0.722,0.753]
Sensitivity
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
1 − Specificity
Threshold
(/100,000)
100
10

2 Comparison to HUNT, Kunzmann and Guidelines

2.1 Full test data

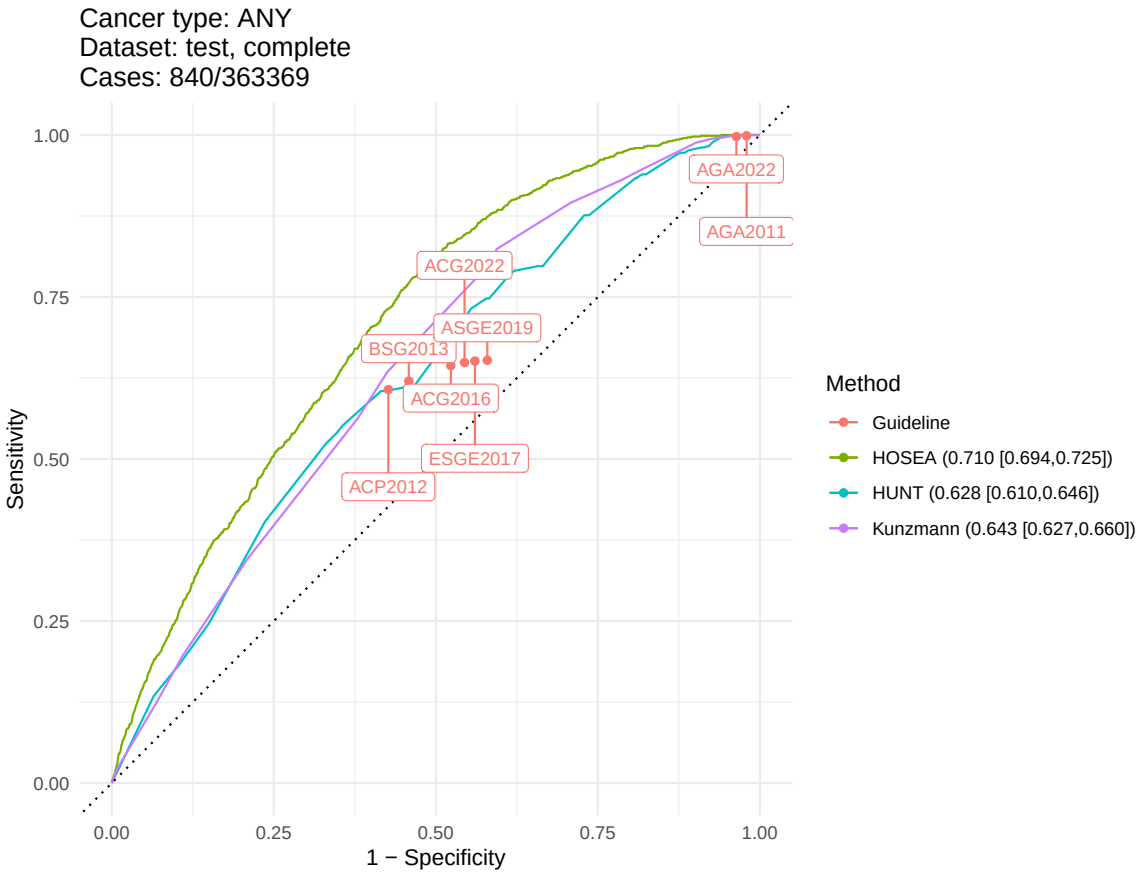


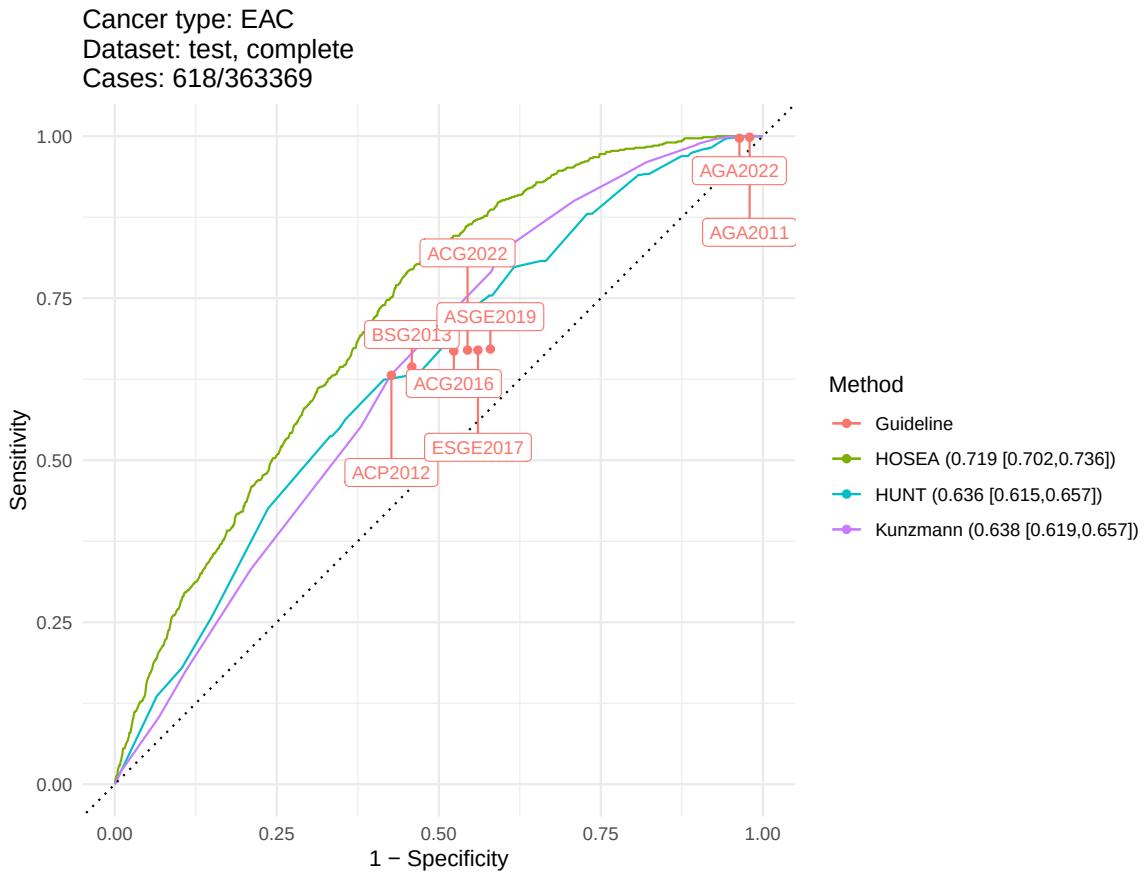

Cancer type: EAC
Dataset: test, all
Cases: 2076/2567069
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
AGA2022
AGA2011
ACG2022
ASGE2019
ACG2016
ESGE2017
ACP2012
Method
Guideline
HOSEA (0.767 [0.758,0.775])
HUNT (0.680 [0.670,0.690])
Kunzmann (0.632 [0.622,0.642])

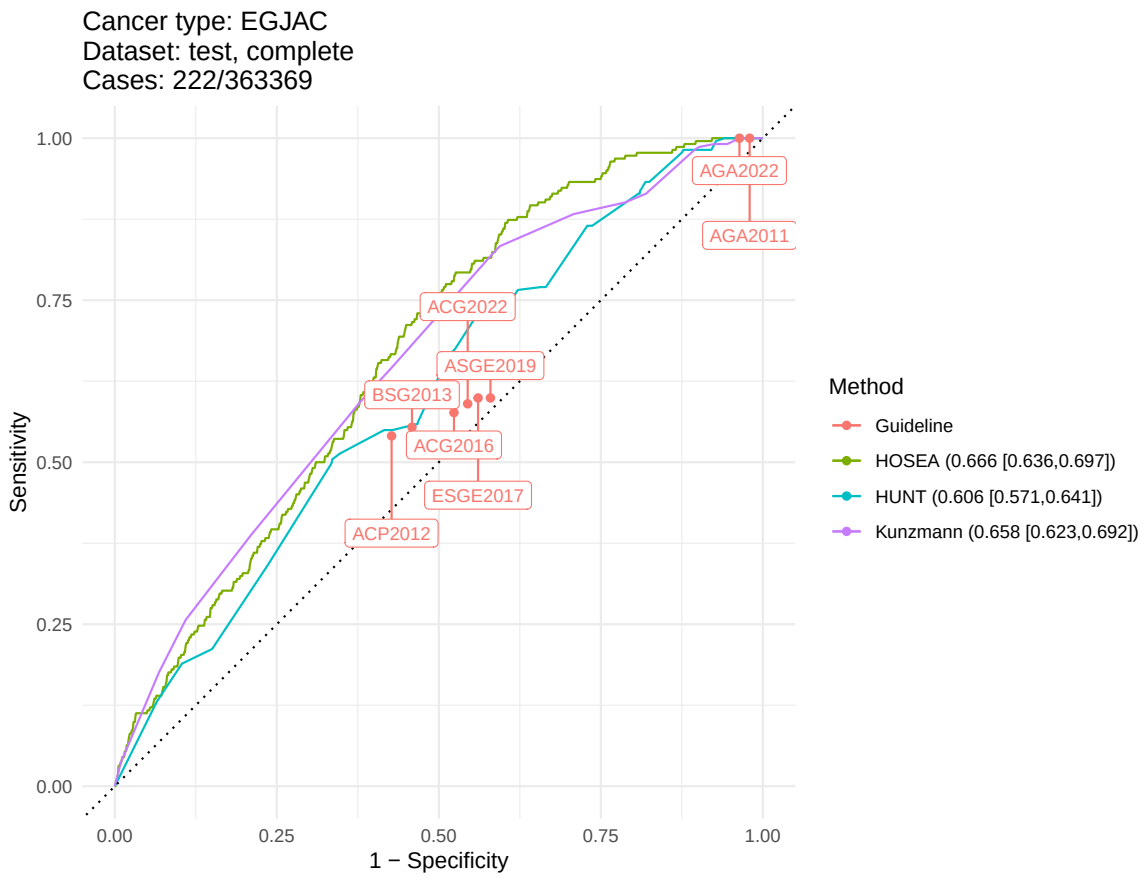

Cancer type: EGJAC
Dataset: test, all
Cases: 772/2567069
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
ACG2022
ASGE2019
BSG2013
ACG2016
ESGE2017
ACP2012
AGA2022
AGA2011
Method
Guideline
HOSEA (0.738 [0.722,0.753])
HUNT (0.667 [0.650,0.684])
Kunzmann (0.644 [0.626,0.661])

2.2 Complete test patients

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

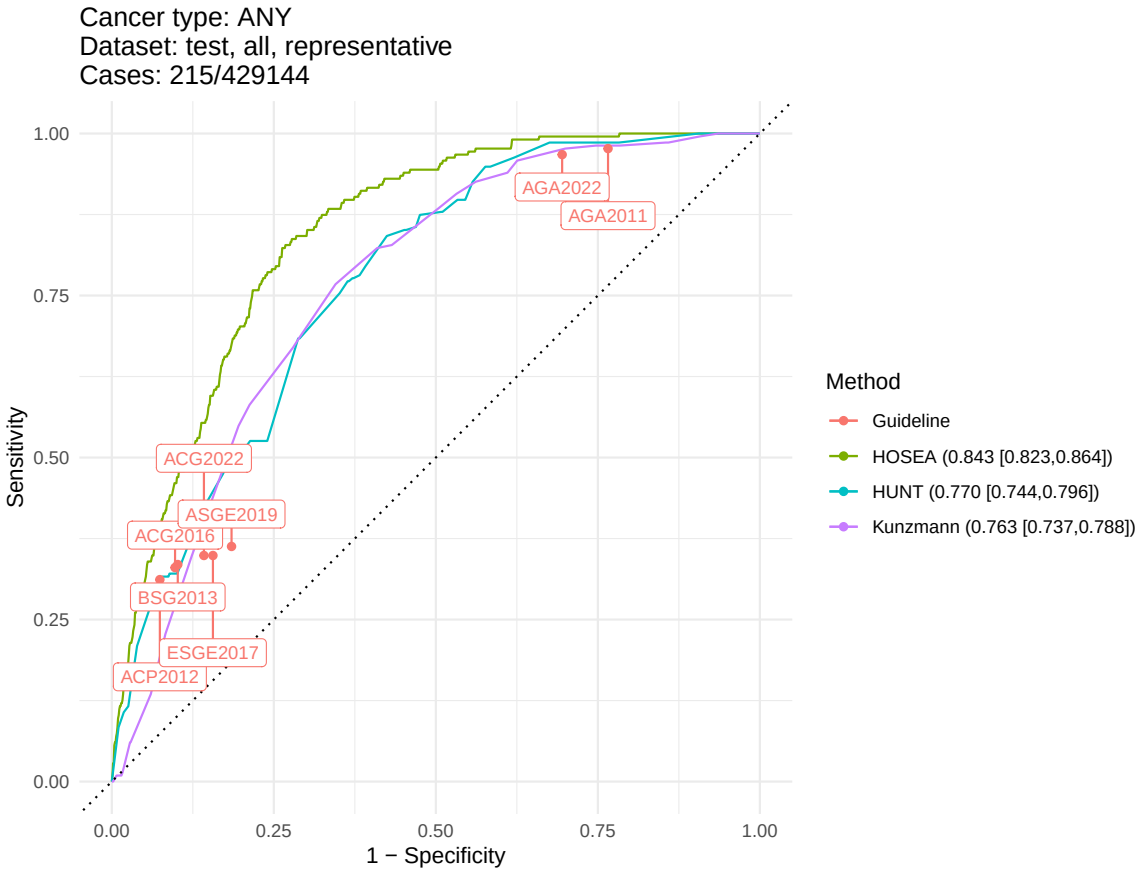


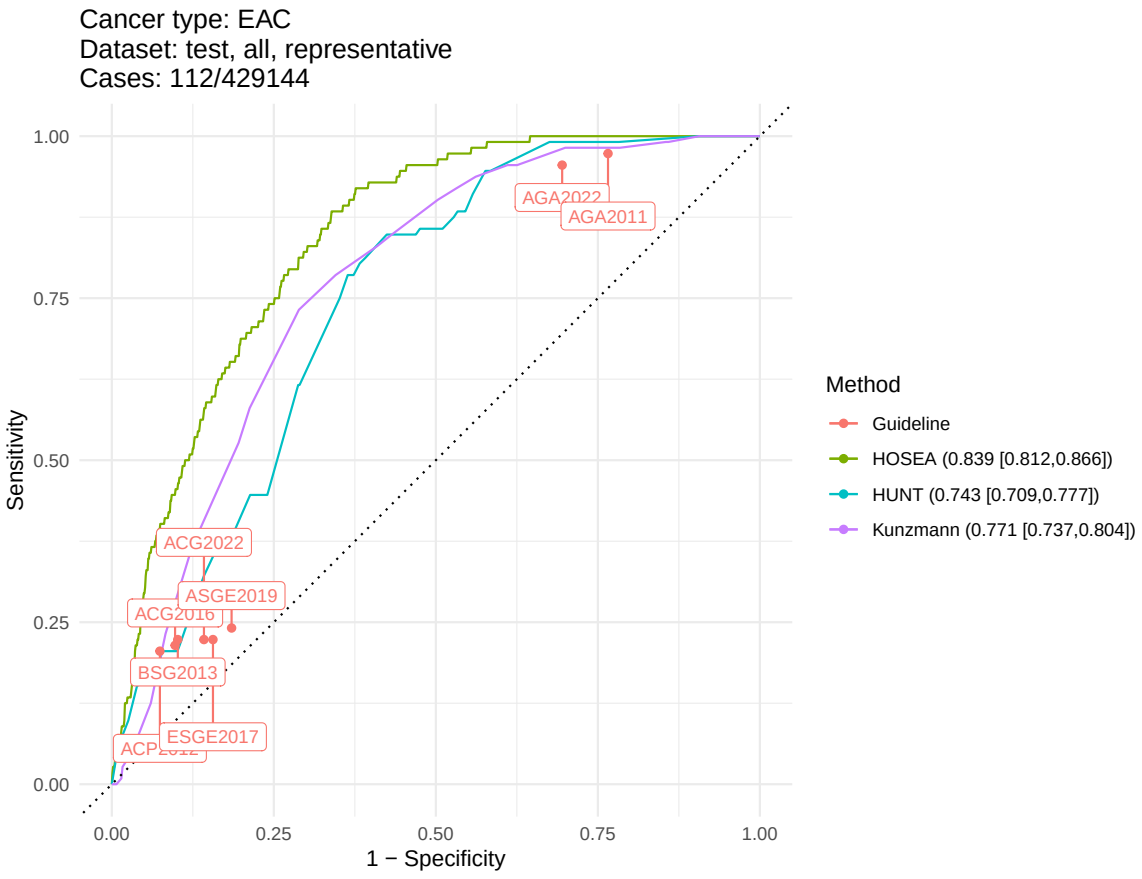

Cancer type: EAC
Dataset: test, complete
Cases: 618/363369
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
AGA2022
AGA2011
ACG2022
ASGE2019
BSG2013
ACG2016
ESGE2017
ACP2012
Method
Guideline
HOSEA (0.719 [0.702,0.736])
HUNT (0.636 [0.615,0.657])
Kunzmann (0.638 [0.619,0.657])

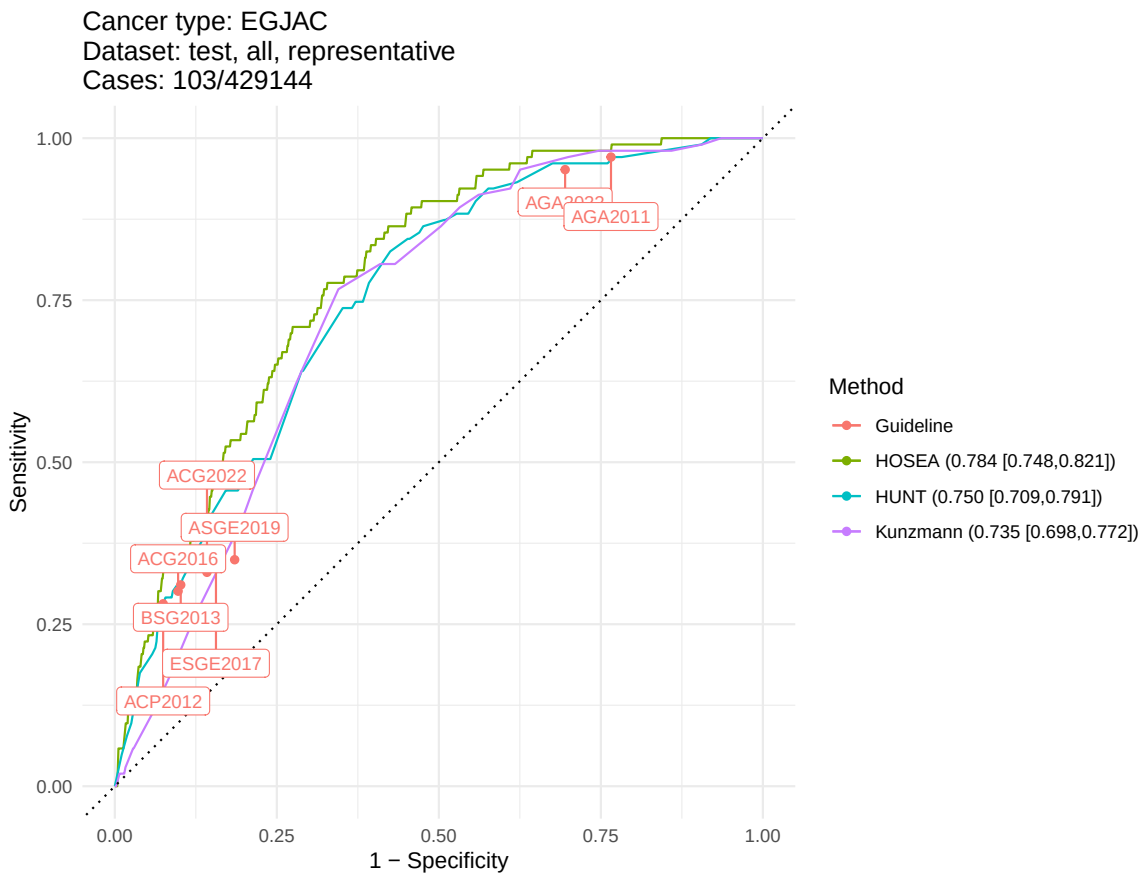

Cancer type: EGJAC
Dataset: test, complete
Cases: 222/363369
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
AGA2022
AGA2011
ACG2022
ASGE2019
BSG2013
ACG2016
ESGE2017
ACP2012
Method
Guideline
HOSEA (0.666 [0.636,0.697])
HUNT (0.606 [0.571,0.641])
Kunzmann (0.658 [0.623,0.692])

2.3 Representative sample

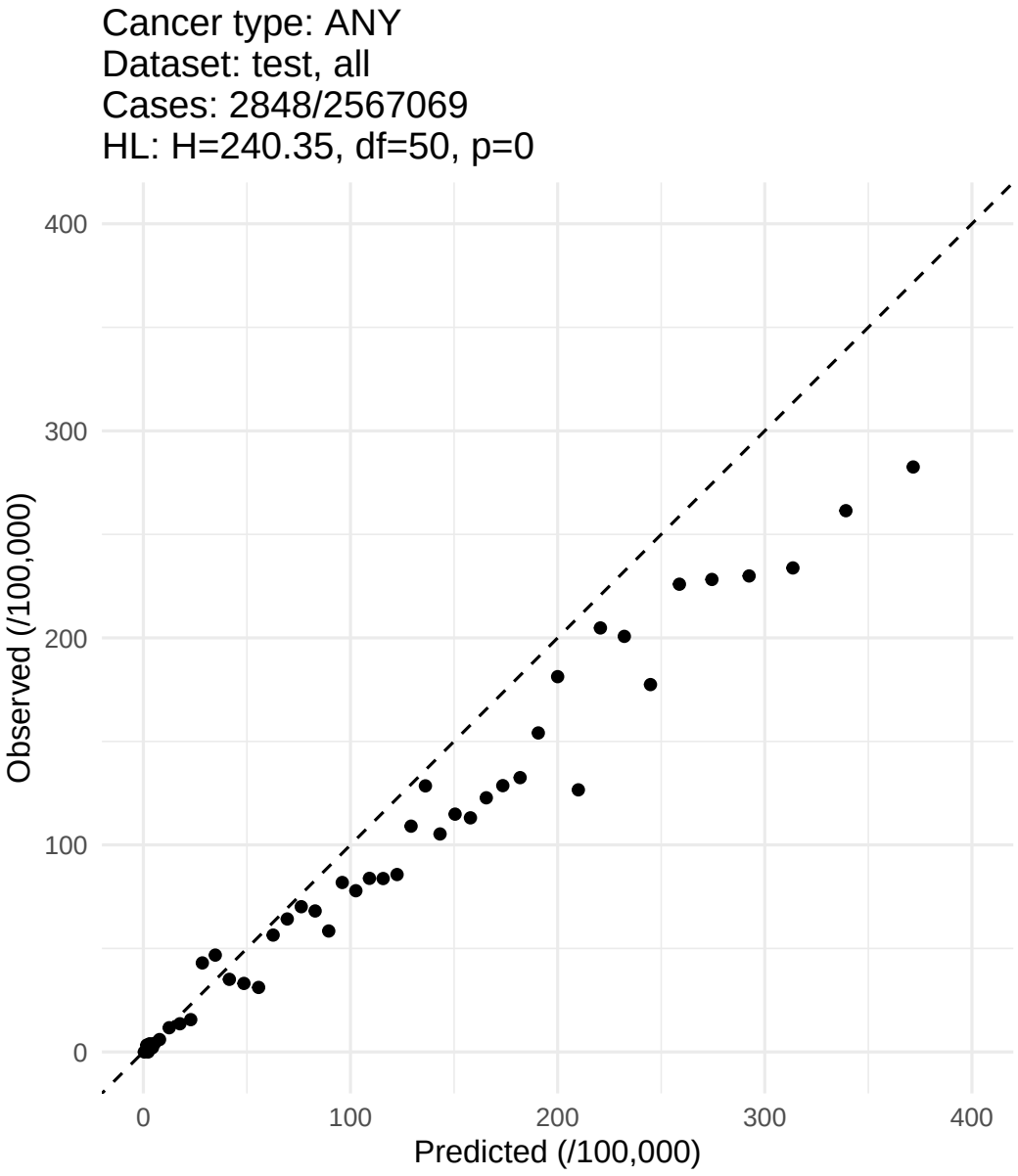
w.r.t. sex and prevalance ratio by sex

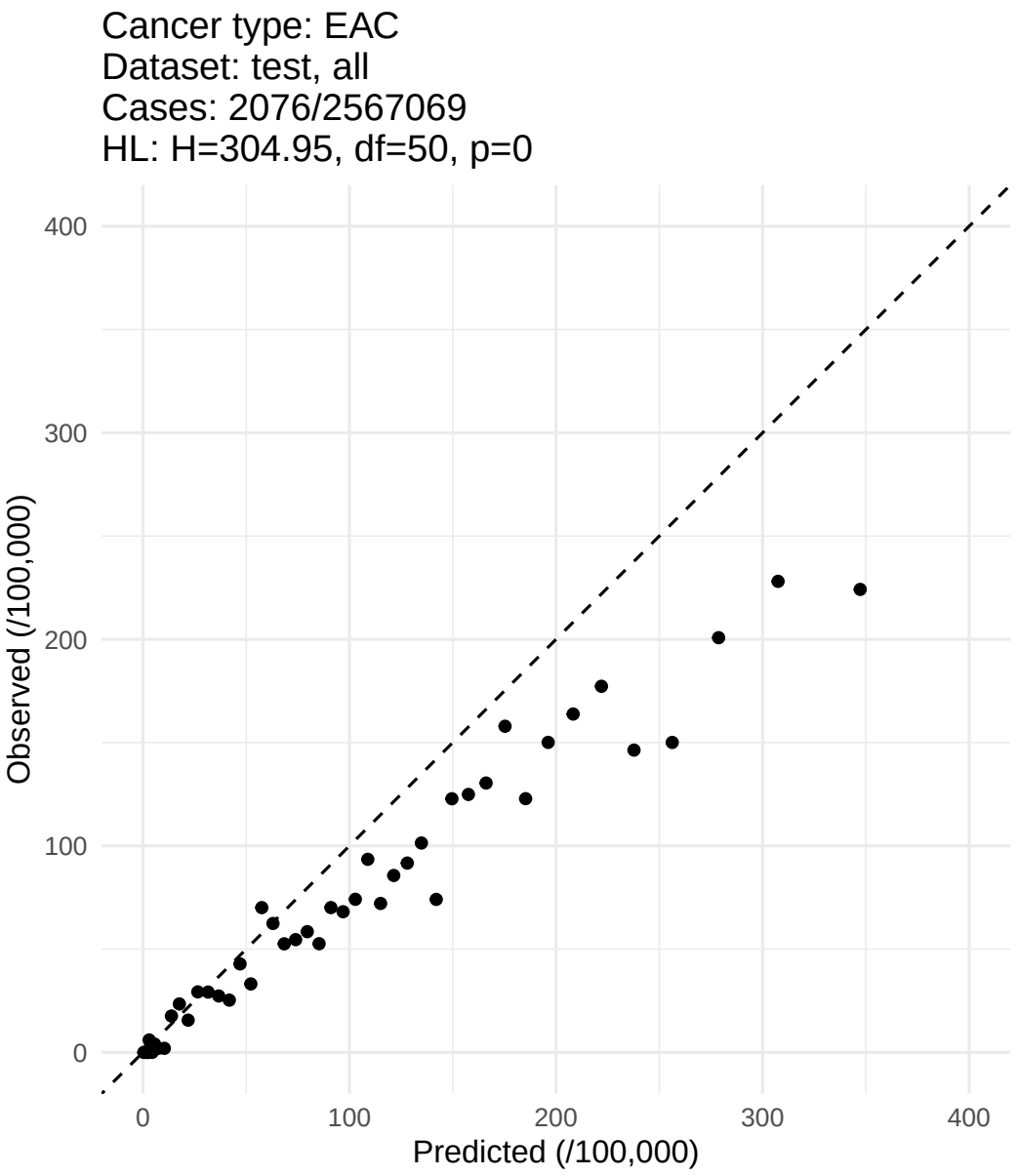


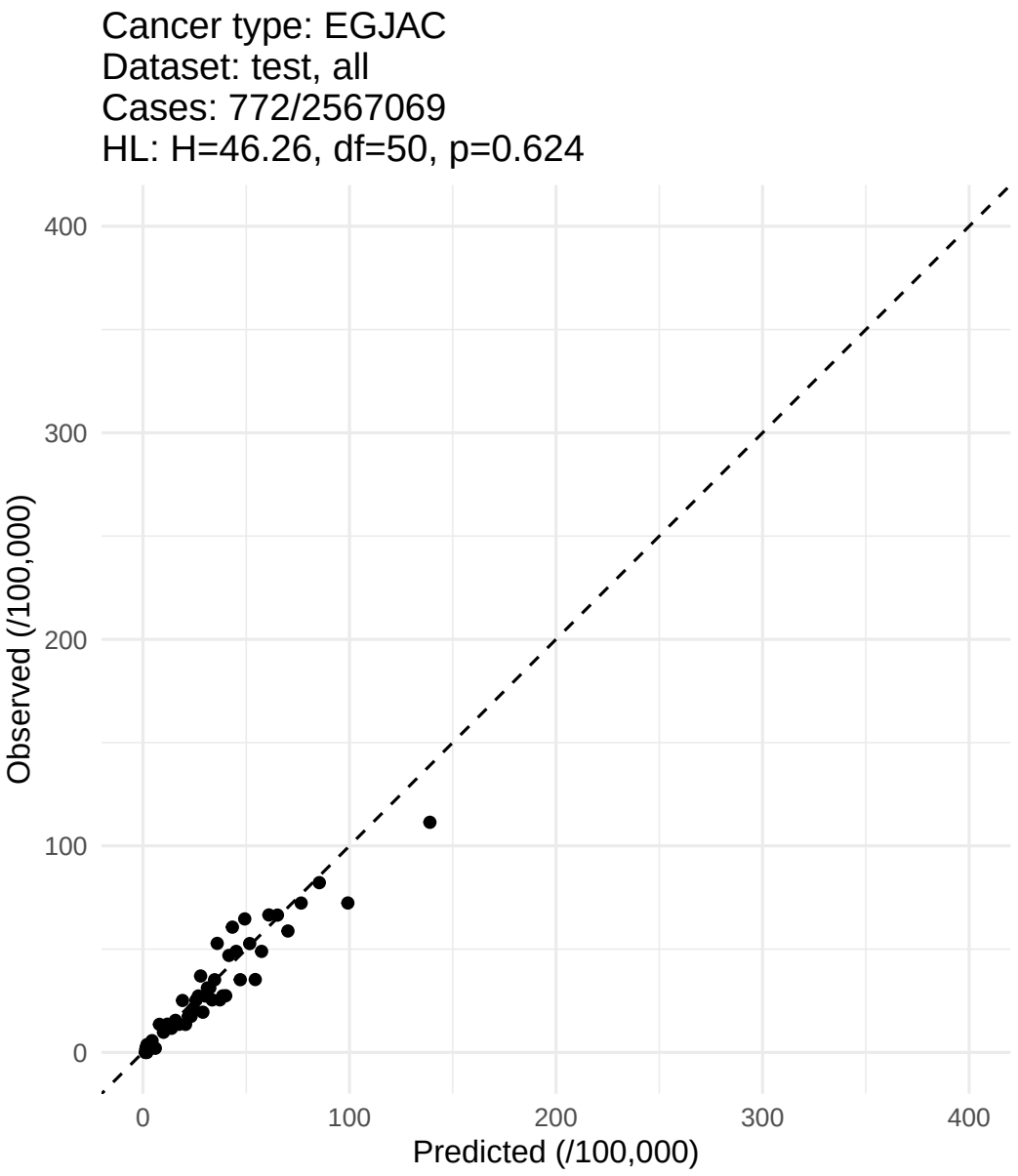

Cancer type: EAC
Dataset: test, all, representative
Cases: 112/429144
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
AGA2022
AGA2011
ACG2022
ASGE2019
ACG2016
BSG2013
ESGE2017
Method
Guideline
HOSEA (0.839 [0.812,0.866])
HUNT (0.743 [0.709,0.777])
Kunzmann (0.771 [0.737,0.804])


Cancer type: EGJAC
Dataset: test, all, representative
Cases: 103/429144
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
ACG2022
ASGE2019
ACG2016
BSG2013
ESGE2017
ACP2012
AGA2011
Method
Guideline
HOSEA (0.784 [0.748,0.821])
HUNT (0.750 [0.709,0.791])
Kunzmann (0.735 [0.698,0.772])

3 Calibration




Cancer type: EAC
Dataset: test, all
Cases: 2076/2567069
HL: H=304.95, df=50, p=0
Observed (/100,000)
Predicted (/100,000)
0
100
200
300
400


Cancer type: EGJAC
Dataset: test, all
Cases: 772/2567069
HL: H=46.26, df=50, p=0.624
Observed (/100,000)
Predicted (/100,000)
0
100
200
300
400

4 Thresholds

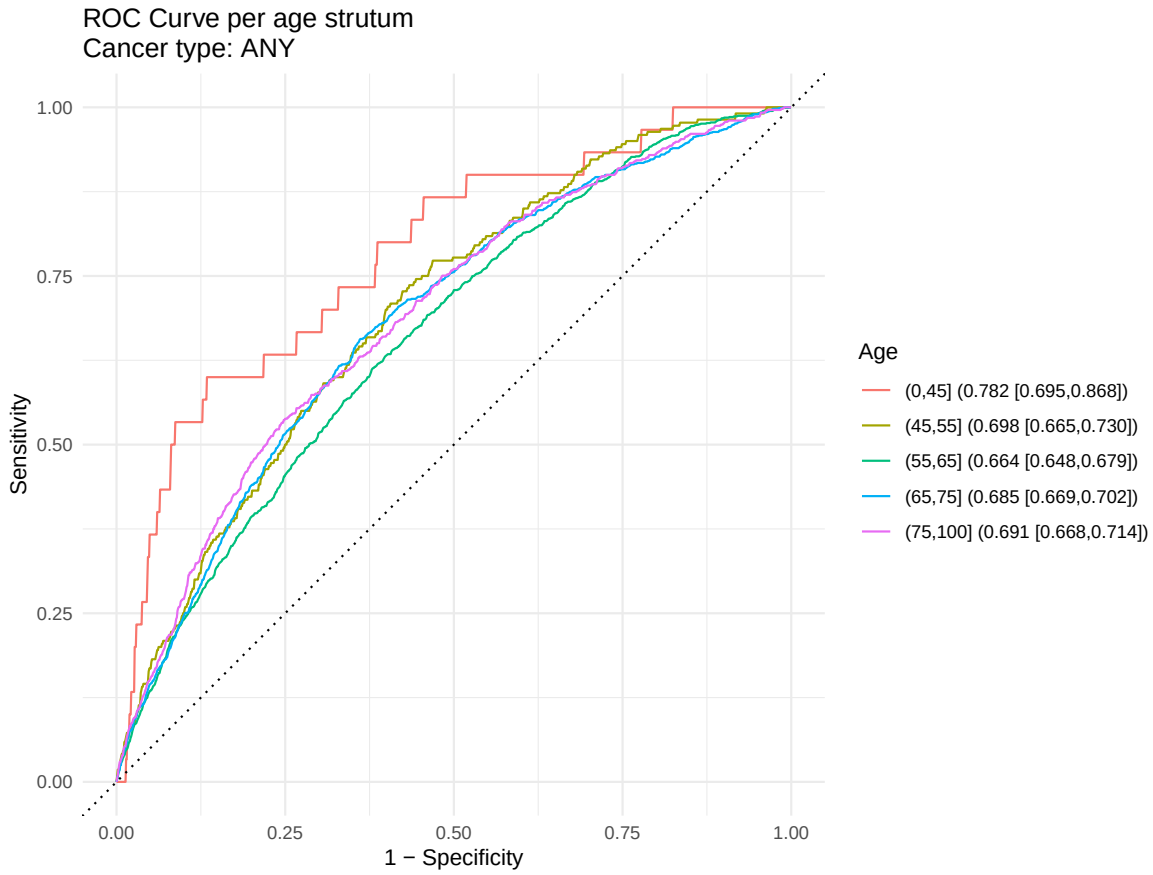
Threshold	TPR	PPV	DetPrevalence
0	100.00	0.11	100.00
5	99.68	0.13	85.38
10	99.54	0.13	81.91
15	99.33	0.14	79.99
20	99.09	0.14	78.05
25	98.81	0.14	76.17
30	98.31	0.15	74.45
35	97.65	0.15	72.88
40	96.91	0.15	71.40
45	96.56	0.15	69.98
50	96.10	0.16	68.56
55	95.68	0.16	67.16
60	95.33	0.16	65.74
65	94.56	0.16	64.29
70	93.71	0.17	62.81
75	92.87	0.17	61.34
80	91.92	0.17	59.85
85	91.01	0.17	58.35
90	90.13	0.18	56.81
95	89.22	0.18	55.28
100	87.96	0.18	53.76
105	86.90	0.18	52.24
110	85.96	0.19	50.71
115	84.69	0.19	49.20
120	83.60	0.19	47.71
125	82.51	0.20	46.22
130	81.04	0.20	44.75
135	79.49	0.20	43.30
140	77.91	0.21	41.87
145	76.33	0.21	40.47
150	75.35	0.21	39.11
155	73.81	0.22	37.75
160	72.54	0.22	36.41
165	70.79	0.22	35.11
170	69.56	0.23	33.85
175	67.94	0.23	32.61
180	66.92	0.24	31.41
185	65.41	0.24	30.25
190	63.87	0.24	29.12
195	62.36	0.25	28.02
200	60.60	0.25	26.98
220	55.13	0.26	23.09
240	49.05	0.28	19.72
260	44.03	0.29	16.81
280	39.04	0.30	14.33
300	34.30	0.31	12.22
320	30.58	0.33	10.43
340	26.86	0.33	8.92
360	23.95	0.35	7.63
380	20.93	0.36	6.53
400	18.68	0.37	5.60
420	16.64	0.38	4.81
440	15.03	0.40	4.15

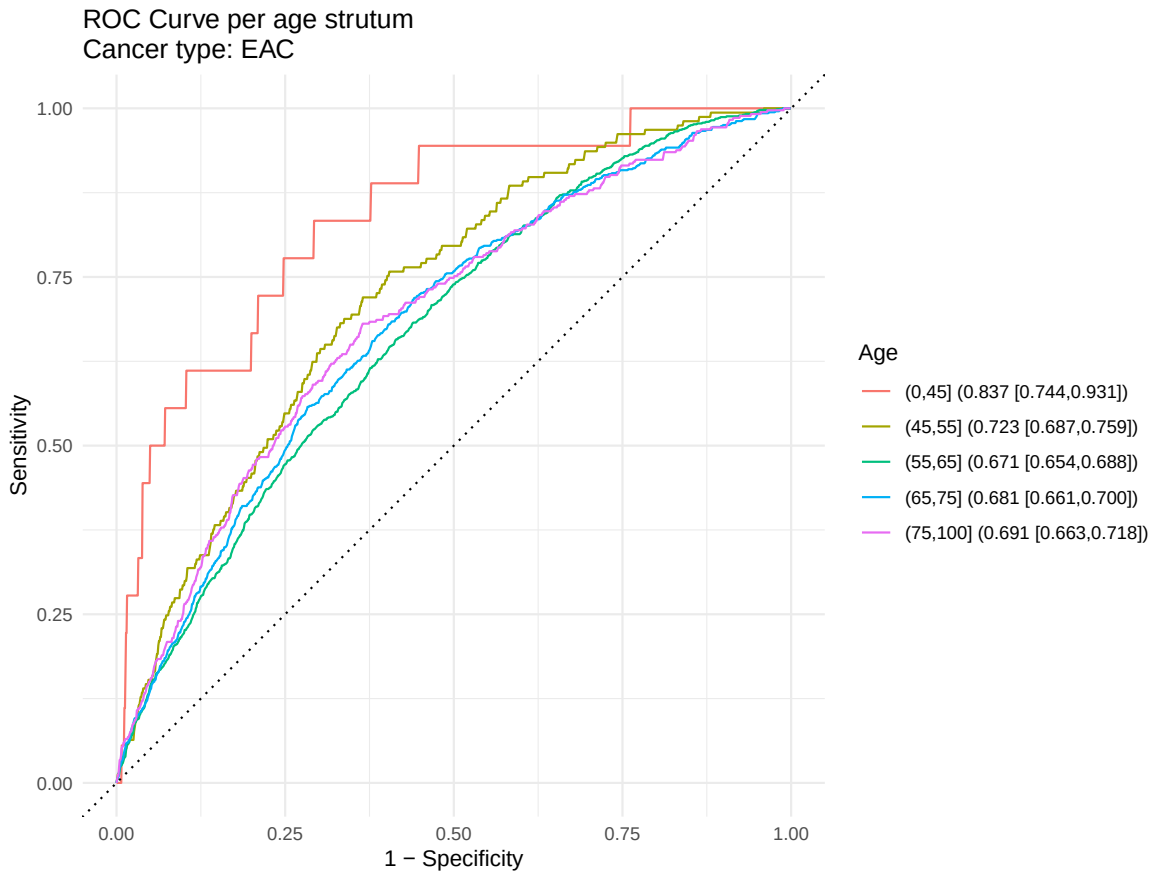
Threshold	TPR	PPV	DetPrevalence
0	100.00	0.08	100.00
5	99.71	0.10	83.69
10	99.57	0.10	79.13
15	99.13	0.11	76.28
20	98.51	0.11	73.81
25	97.93	0.11	71.60
30	97.40	0.11	69.57
35	96.53	0.12	67.62
40	95.91	0.12	65.67
45	95.28	0.12	63.73
50	94.27	0.12	61.81
55	93.45	0.13	59.91
60	91.96	0.13	58.05
65	90.51	0.13	56.21
70	89.16	0.13	54.38
75	88.01	0.14	52.58
80	86.75	0.14	50.81
85	85.40	0.14	49.07
90	84.15	0.14	47.31
95	82.76	0.15	45.60
100	81.36	0.15	43.92
105	79.91	0.15	42.24
110	78.03	0.16	40.59
115	76.16	0.16	38.99
120	74.81	0.16	37.41
125	73.27	0.17	35.87
130	71.48	0.17	34.37
135	70.23	0.17	32.92
140	68.21	0.18	31.51
145	67.00	0.18	30.17
150	65.22	0.18	28.86
155	63.10	0.18	27.60
160	61.27	0.19	26.39
165	59.68	0.19	25.23
170	57.66	0.19	24.11
175	55.35	0.19	23.04
180	53.61	0.20	22.01
185	51.88	0.20	21.03
190	50.58	0.20	20.09
195	48.99	0.21	19.19
200	47.25	0.21	18.31
220	41.18	0.22	15.25
240	36.27	0.23	12.71
260	31.89	0.24	10.61
280	28.23	0.26	8.87
300	25.14	0.27	7.44
320	21.15	0.27	6.25
340	18.74	0.29	5.27
360	16.43	0.30	4.46
380	14.21	0.30	3.79
400	12.24	0.31	3.22
420	11.22	0.33	2.75
440	9.97	0.34	2.35

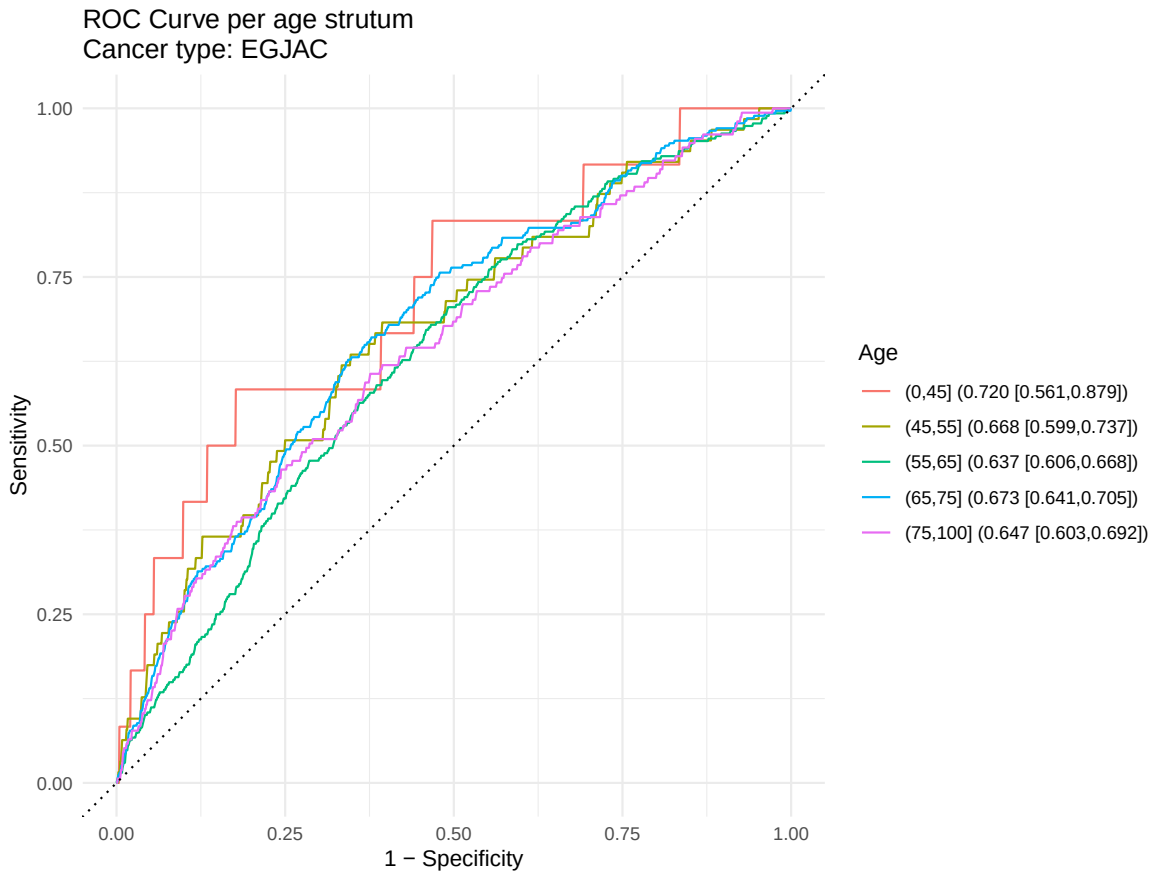
Threshold	TPR	PPV	DetPrevalence
0	100.00	0.03	100.00
5	98.83	0.04	80.09
10	97.54	0.04	74.84
15	95.47	0.04	69.68
20	91.58	0.04	63.75
25	87.31	0.05	56.02
30	79.15	0.05	47.04
35	70.85	0.06	38.34
40	62.69	0.06	30.95
45	52.59	0.06	25.03
50	44.43	0.07	20.28
55	38.73	0.07	16.50
60	33.81	0.08	13.45
65	28.89	0.08	10.99
70	24.87	0.08	9.02
75	20.47	0.08	7.40
80	17.88	0.09	6.08
85	15.54	0.09	5.02
90	13.47	0.10	4.15
95	11.27	0.10	3.44
100	10.10	0.11	2.84
105	8.94	0.11	2.36
110	7.38	0.11	1.96
115	6.74	0.12	1.64
120	6.35	0.14	1.36
125	5.57	0.15	1.14
130	4.79	0.15	0.96
135	3.89	0.15	0.80
140	3.37	0.15	0.67
145	2.85	0.15	0.57
150	2.07	0.13	0.48
155	1.68	0.12	0.41
160	1.42	0.12	0.35
165	1.30	0.13	0.29
170	1.04	0.12	0.25
175	1.04	0.14	0.22
180	0.65	0.11	0.19
185	0.39	0.07	0.16
190	0.39	0.09	0.14
195	0.39	0.10	0.12
200	0.39	0.11	0.10
220	0.26	0.13	0.06
240	0.00	0.00	0.03
260	0.00	0.00	0.02
280	0.00	0.00	0.01
300	0.00	0.00	0.01
320	0.00	0.00	0.00
340	0.00	0.00	0.00
360	0.00	0.00	0.00
380	0.00	0.00	0.00
400	0.00	0.00	0.00
420	0.00	0.00	0.00
440	0.00	0.00	0.00

5 Stratification by identity groups

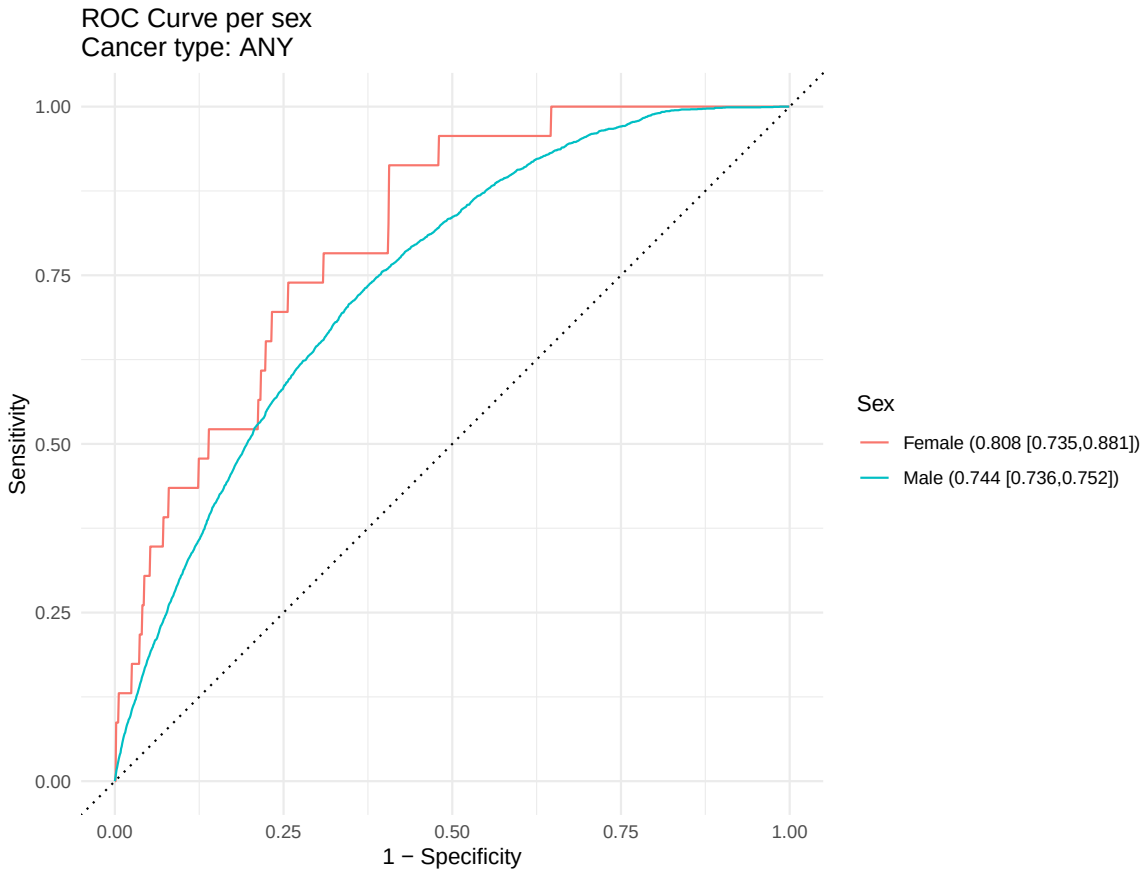
5.1 Age

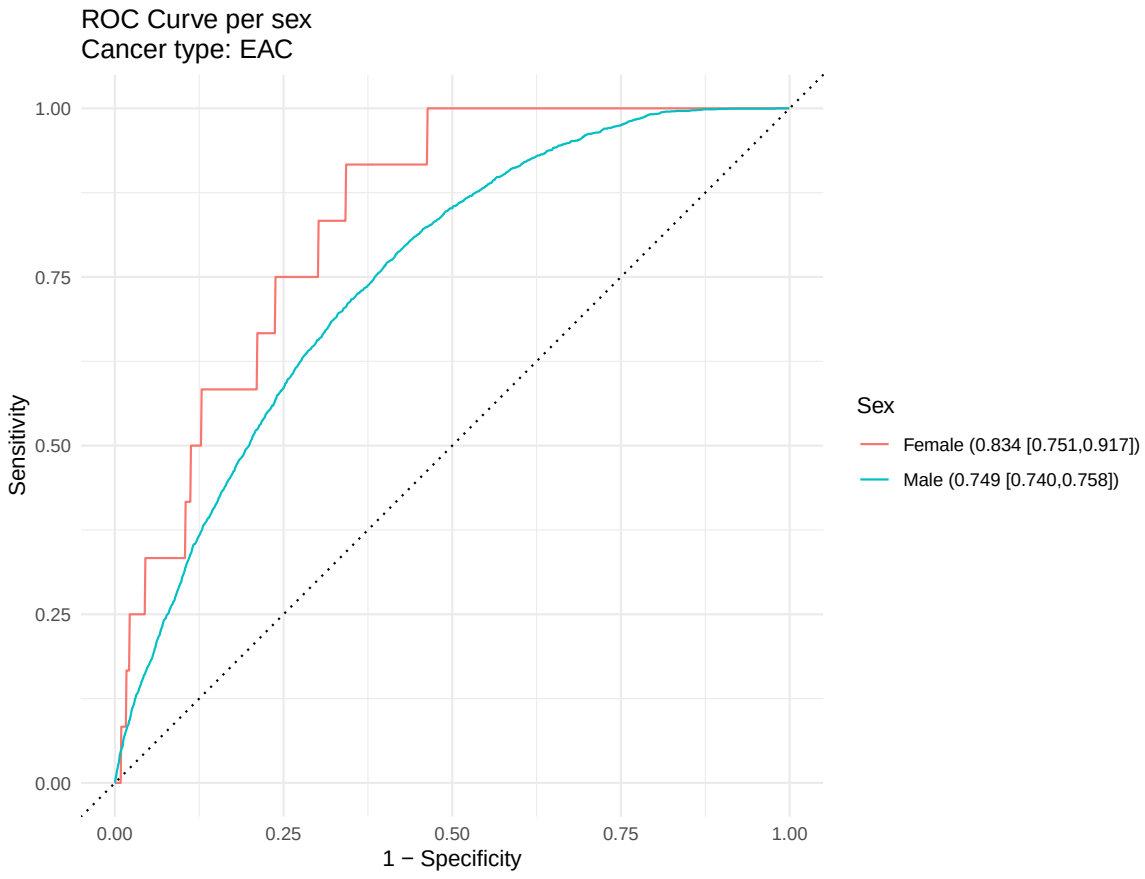


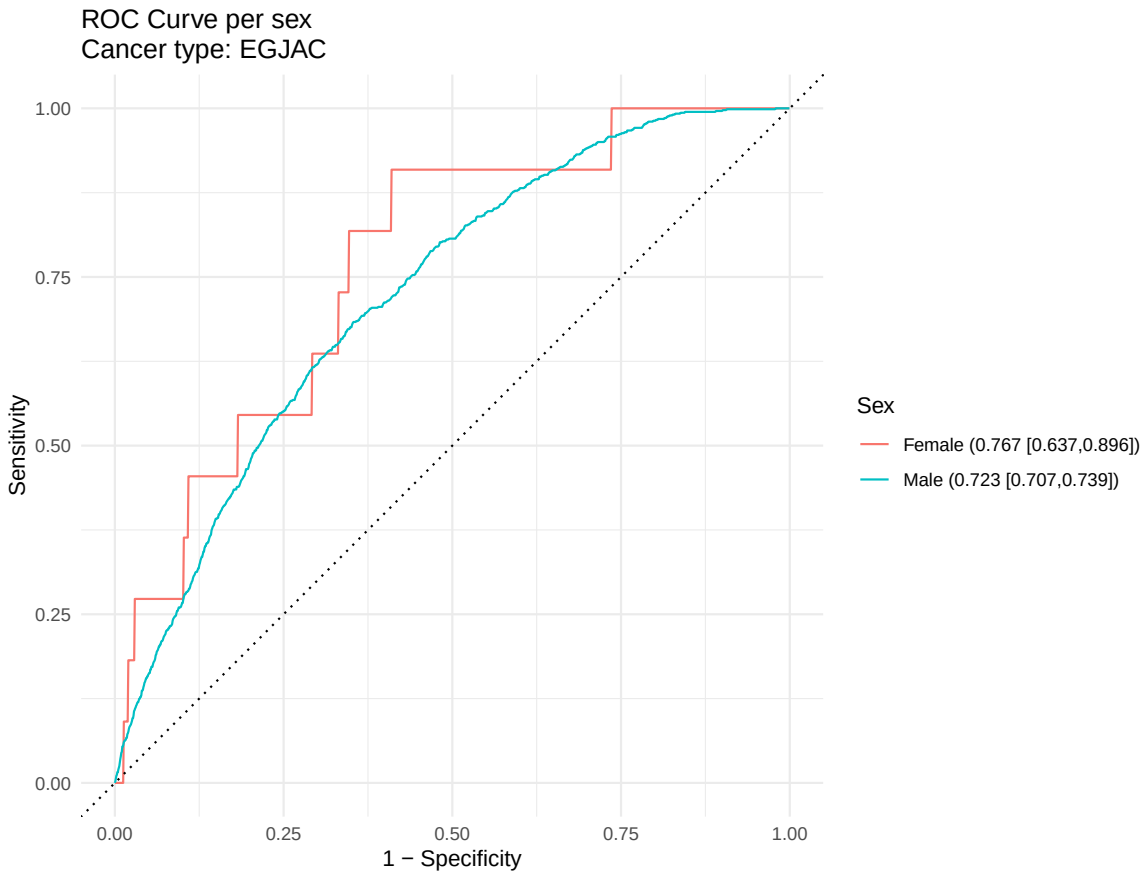

ROC Curve per age strutum
Cancer type: EAC
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Age
(0,45] (0.837 [0.744,0.931])
(45,55] (0.723 [0.687,0.759])
(55,65] (0.671 [0.654,0.688])
(65,75] (0.681 [0.661,0.700])
(75,100] (0.691 [0.663,0.718])


ROC Curve per age strutum
Cancer type: EGJAC
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Age
(0,45] (0.720 [0.561,0.879])
(45,55] (0.668 [0.599,0.737])
(55,65] (0.637 [0.606,0.668])
(65,75] (0.673 [0.641,0.705])
(75,100] (0.647 [0.603,0.692])

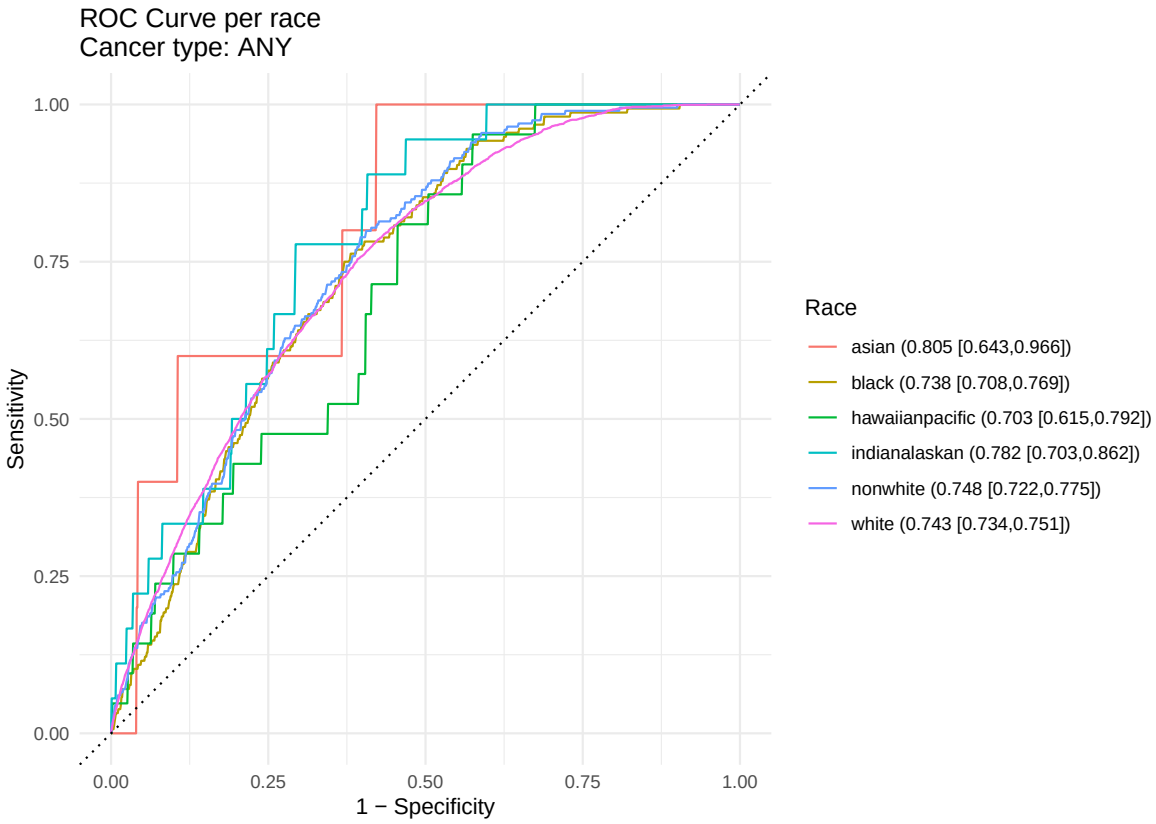
5.2 Sex

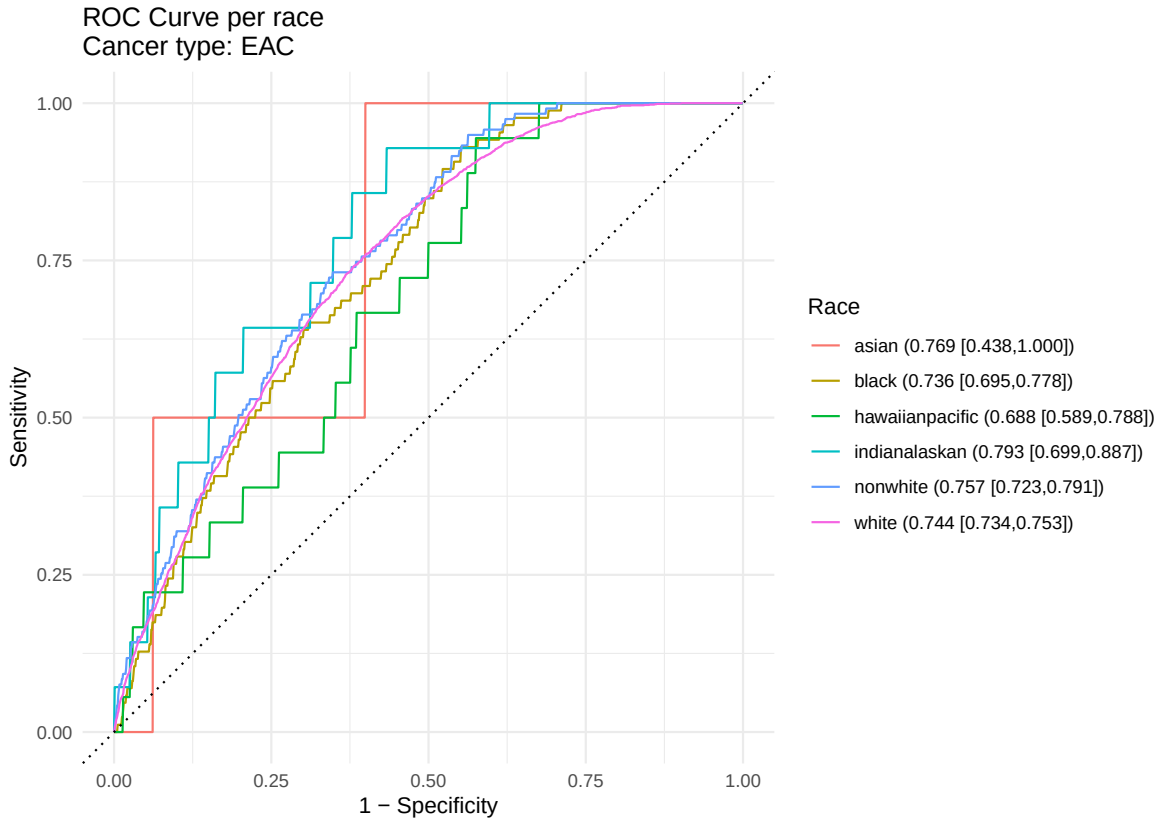

ROC Curve per sex
Cancer type: ANY
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Sex
Female (0.808 [0.735,0.881])
Male (0.744 [0.736,0.752])

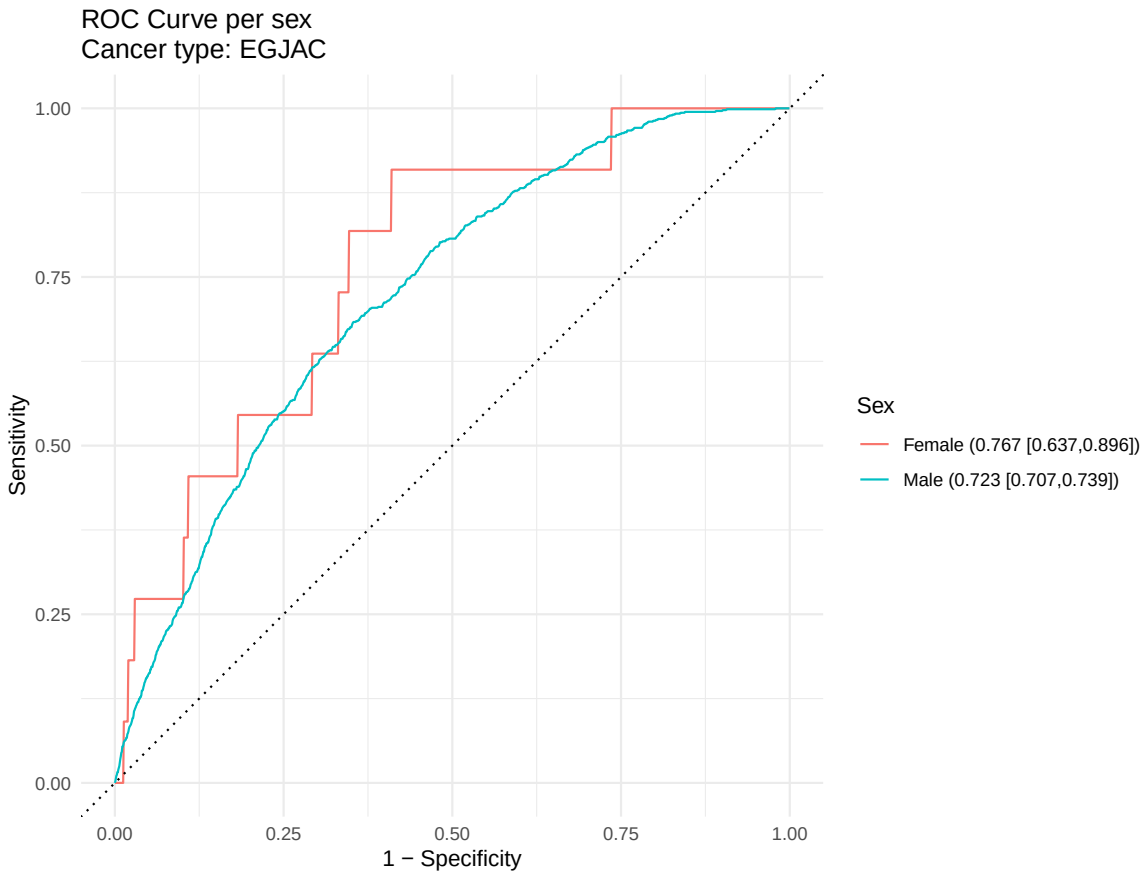

ROC Curve per sex
Cancer type: EAC
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Sex
Female (0.834 [0.751,0.917])
Male (0.749 [0.740,0.758])


ROC Curve per sex
Cancer type: EGJAC
Sensitivity
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
1 − Specificity
Sex
Female (0.767 [0.637,0.896])
Male (0.723 [0.707,0.739])

5.3 Race


ROC Curve per race
Cancer type: ANY
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Race
asian (0.805 [0.643,0.966])
black (0.738 [0.708,0.769])
hawaiianpacific (0.703 [0.615,0.792])
indianalaskan (0.782 [0.703,0.862])
nonwhite (0.748 [0.722,0.775])
white (0.743 [0.734,0.751])


ROC Curve per race
Cancer type: EAC
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Race
asian (0.769 [0.438,1.000])
black (0.736 [0.695,0.778])
hawaiianpacific (0.688 [0.589,0.788])
indianalaskan (0.793 [0.699,0.887])
nonwhite (0.757 [0.723,0.791])
white (0.744 [0.734,0.753])


ROC Curve per sex
Cancer type: EGJAC
Sensitivity
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
1 − Specificity
Sex
Female (0.767 [0.637,0.896])
Male (0.723 [0.707,0.739])

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.

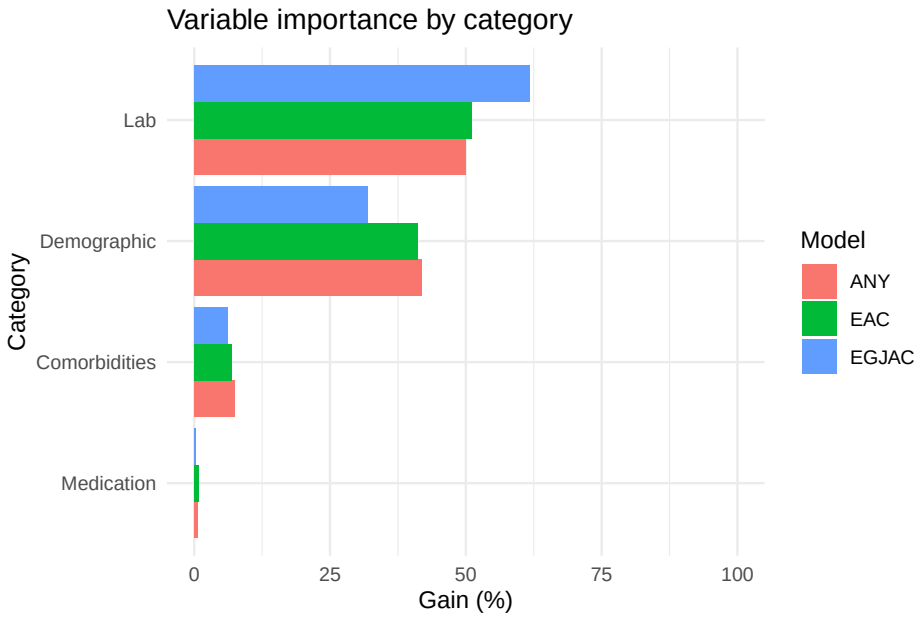
Stage	Test. AUC	Nb. cases
Any	0.761 [0.753,0.768]	2818
I	0.799 [0.781,0.816]	495
II	0.778 [0.759,0.797]	435
III	0.767 [0.754,0.780]	878
IV	0.752 [0.740,0.763]	1244
I+	0.764 [0.756,0.772]	2448
II+	0.757 [0.749,0.766]	2169
III+	0.755 [0.746,0.764]	1950
IV+	0.752 [0.740,0.763]	1244

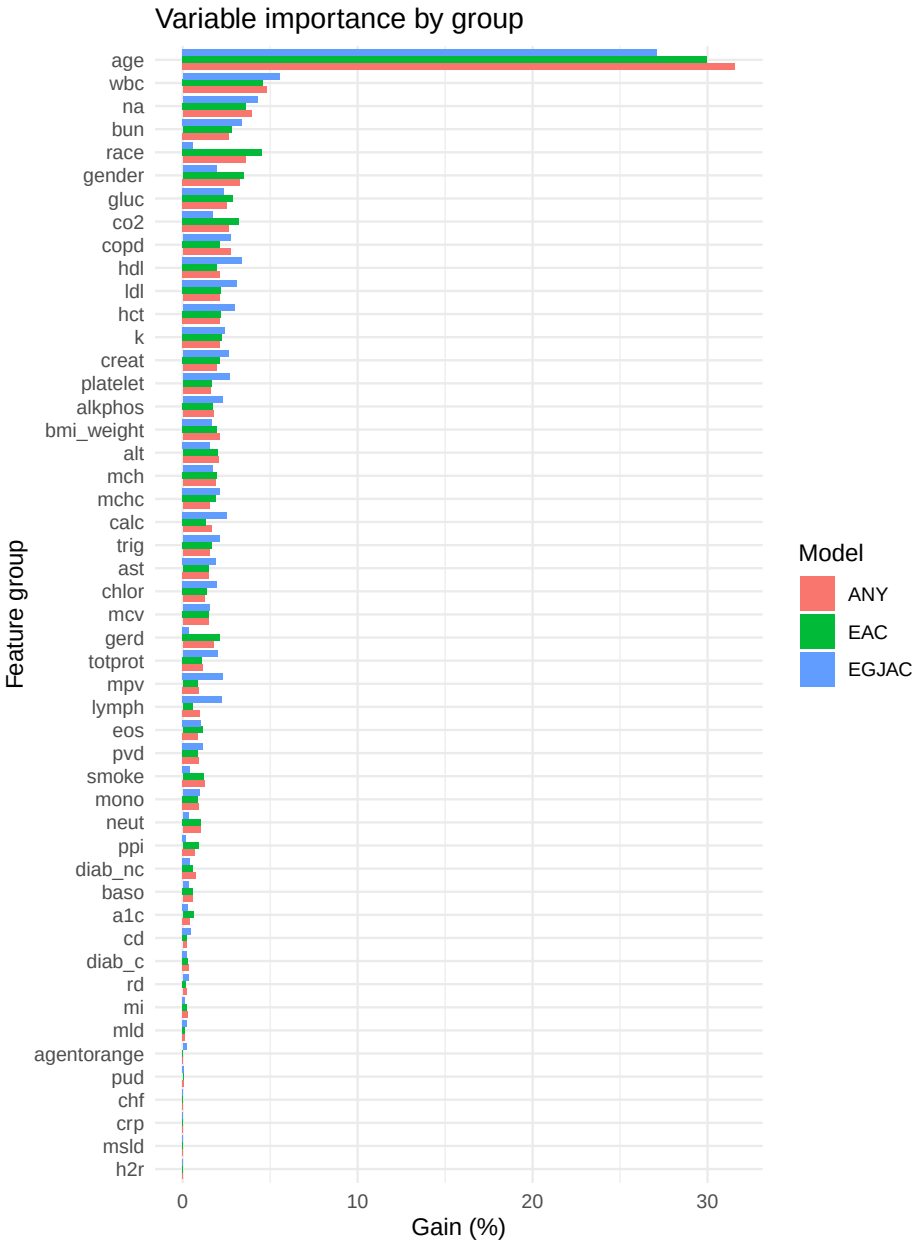
Stage	Test. AUC	Nb. cases
Any	0.766 [0.758,0.775]	2054
I	0.809 [0.789,0.828]	350
II	0.785 [0.763,0.807]	302
III	0.767 [0.753,0.782]	650
IV	0.758 [0.745,0.771]	913
I+	0.767 [0.758,0.776]	1813
II+	0.760 [0.751,0.770]	1608
III+	0.758 [0.748,0.769]	1451
IV+	0.758 [0.745,0.771]	913

Stage	Test. AUC	Nb. cases
Any	0.737 [0.721,0.752]	764
I	0.759 [0.723,0.796]	145
II	0.752 [0.715,0.789]	133
III	0.750 [0.723,0.778]	228
IV	0.723 [0.699,0.747]	331
I+	0.739 [0.722,0.756]	635
II+	0.735 [0.716,0.753]	561
III+	0.731 [0.712,0.750]	499
IV+	0.723 [0.699,0.747]	331

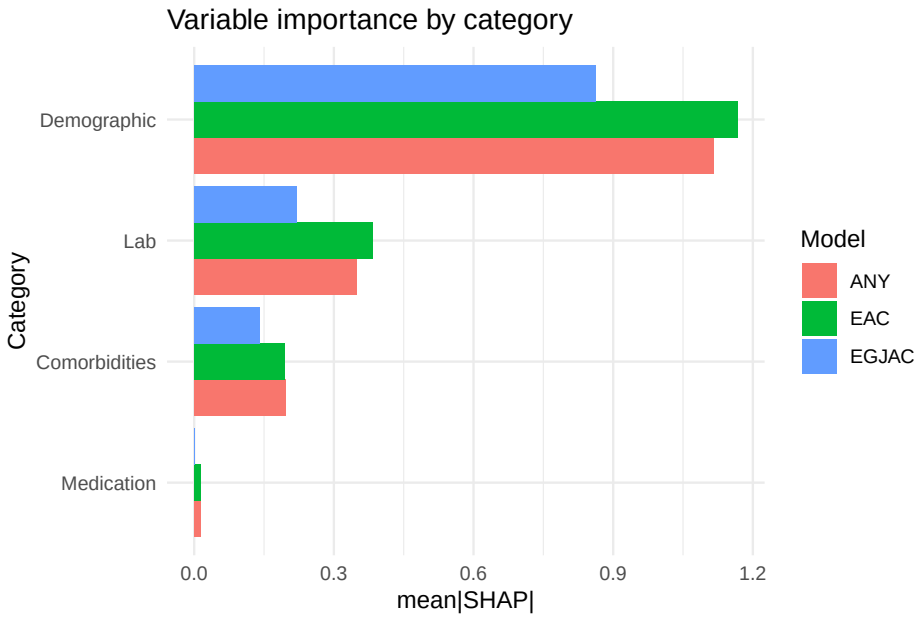
7 Variable importance

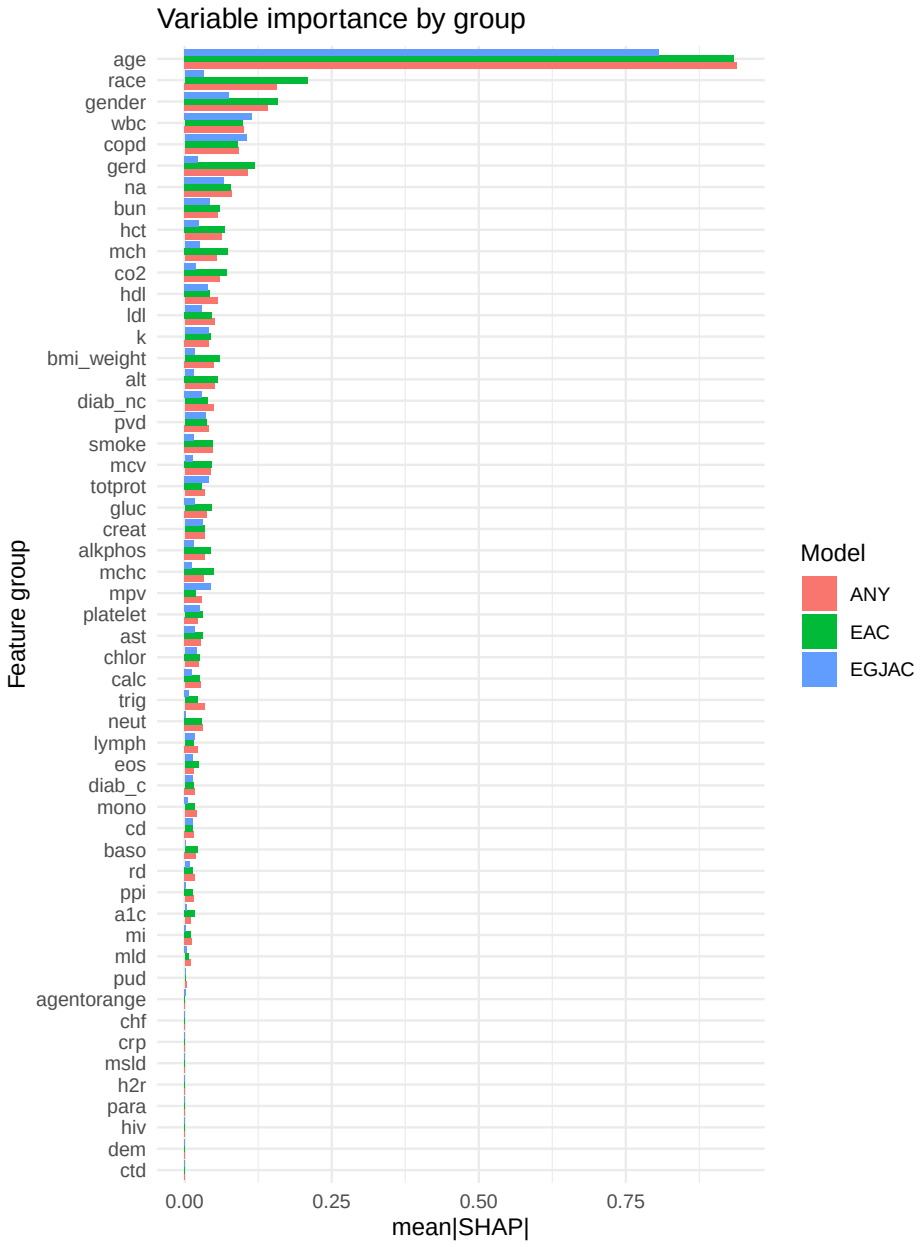
7.1 Gain VI



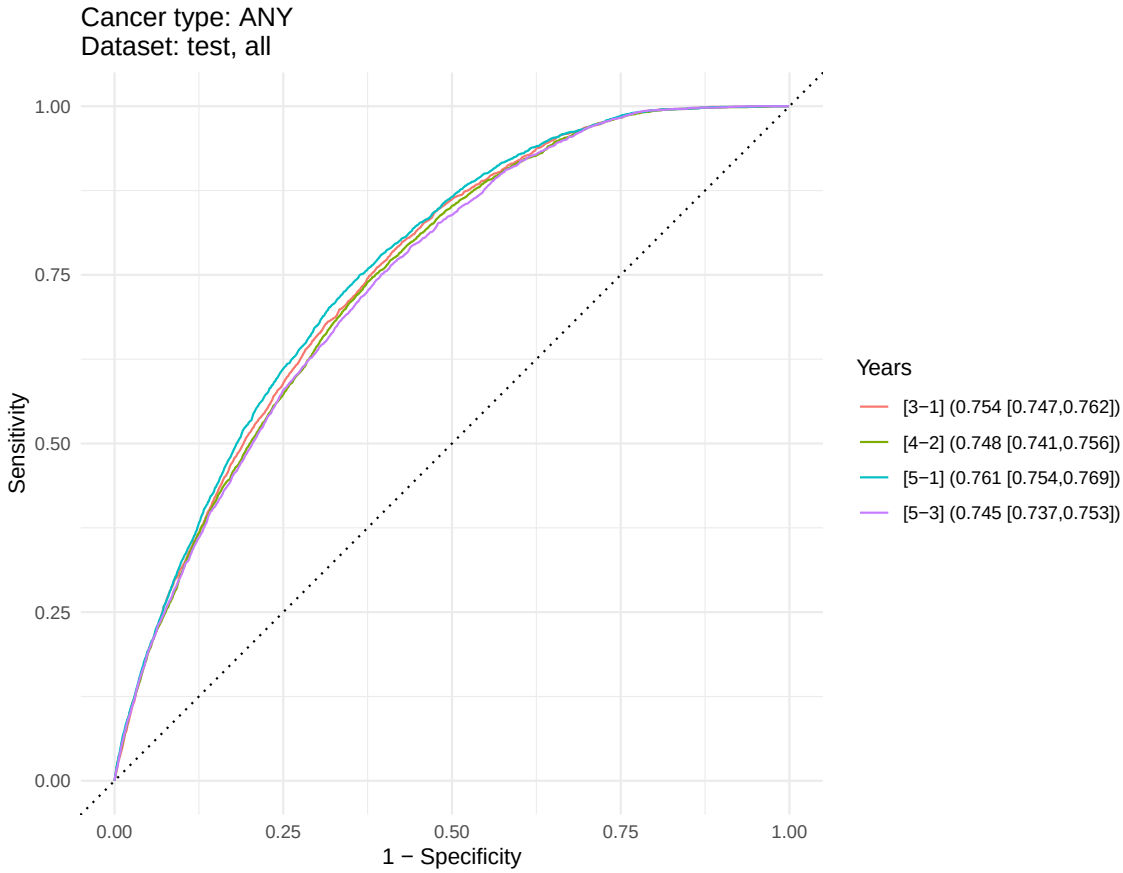

Variable importance by group
age
wbc
na
bun
race
gender
gluc
co2
copd
hdl
ldl
hct
k
creat
platelet
alkphos
bmi_weight
alt
mch
mchc
calc
trig
ast
chlor
mcv
gerd
totprot
mpv
lymph
eos
pvd
smoke
mono
neut
ppi
diab_nc
baso
a1c
cd
diab_c
rd
mi
mld
agentorange
pud
chf
crp
msld
h2r
Feature group
0
10
20
30
Gain (%)
Model
ANY
EAC
EGJAC

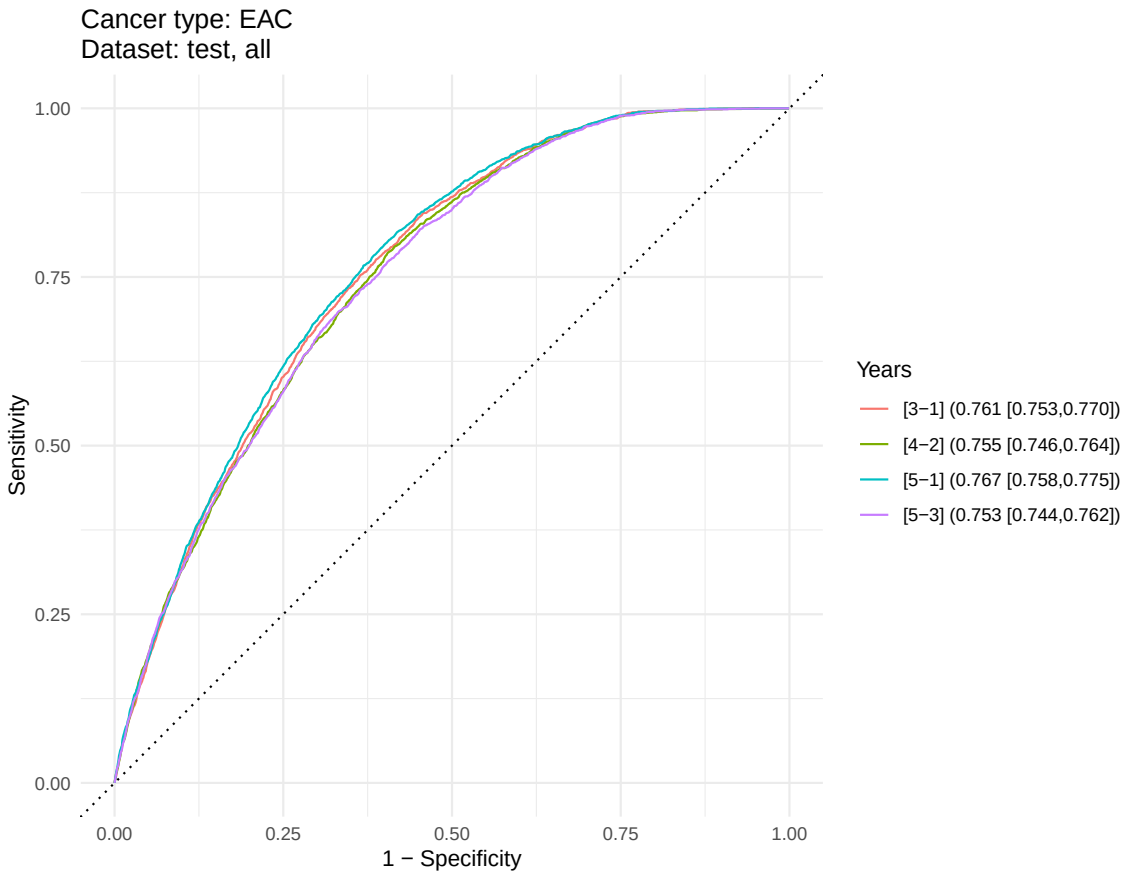
7.2 SHAP VI

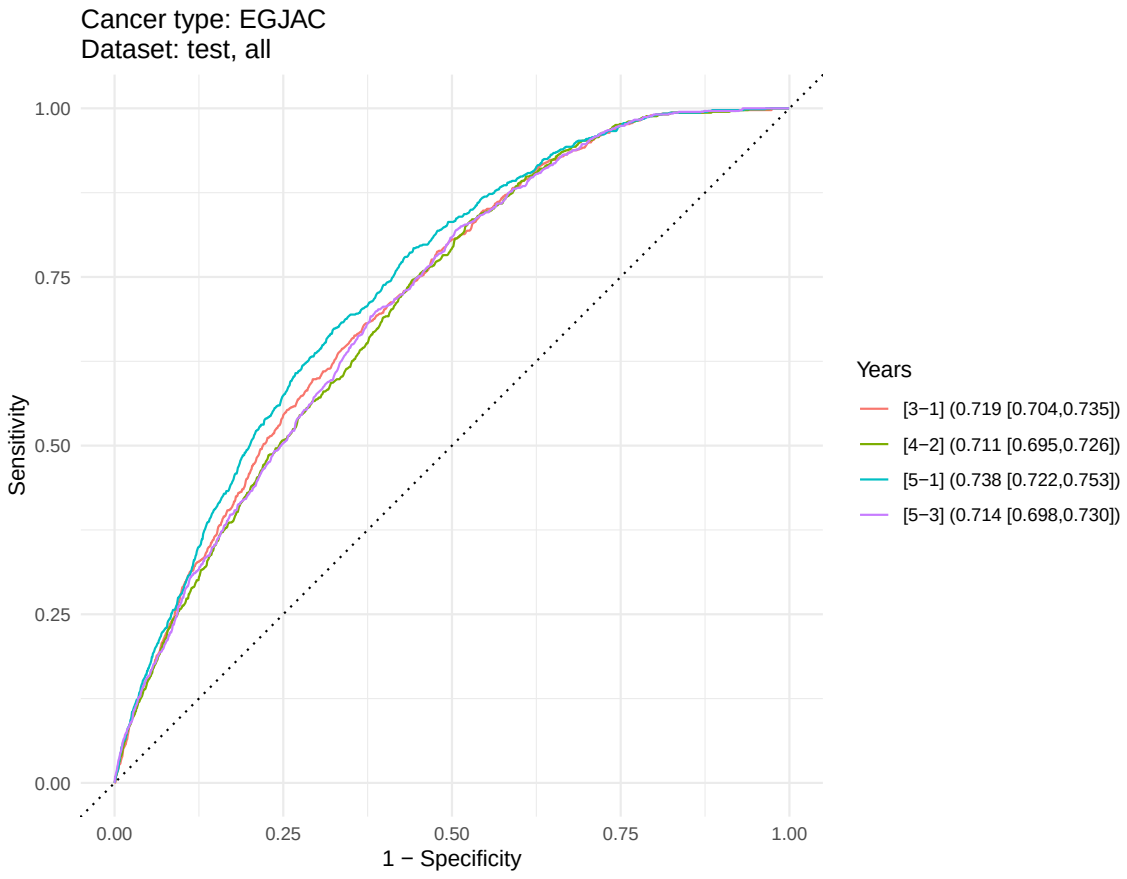

Variable importance by category
Demographic
Lab
Comorbidities
Medication
Category
0.0
0.3
0.6
0.9
1.2
mean|SHAP|
Model
ANY
EAC
EGJAC


Variable importance by group
age
race
gender
wbc
copd
gerd
na
bun
hct
mch
co2
hdl
ldl
k
bmi_weight
alt
diab_nc
pvd
smoke
mcv
totprot
gluc
creat
alkphos
mchc
mpv
platelet
ast
chlor
calc
trig
neut
lymph
eos
diab_c
mono
cd
baso
rd
ppi
a1c
mi
mld
pud
agentorange
chf
crp
msld
h2r
para
hiv
dem
ctd
Feature group
0.00
0.25
0.50
0.75
mean|SHAP|
Model
ANY
EAC
EGJAC

8 Years prior




Cancer type: EAC
Dataset: test, all
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Years
[3−1] (0.761 [0.753,0.770])
[4−2] (0.755 [0.746,0.764])
[5−1] (0.767 [0.758,0.775])
[5−3] (0.753 [0.744,0.762])


Cancer type: EGJAC
Dataset: test, all
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Years
[3−1] (0.719 [0.704,0.735])
[4−2] (0.711 [0.695,0.726])
[5−1] (0.738 [0.722,0.753])
[5−3] (0.714 [0.698,0.730])