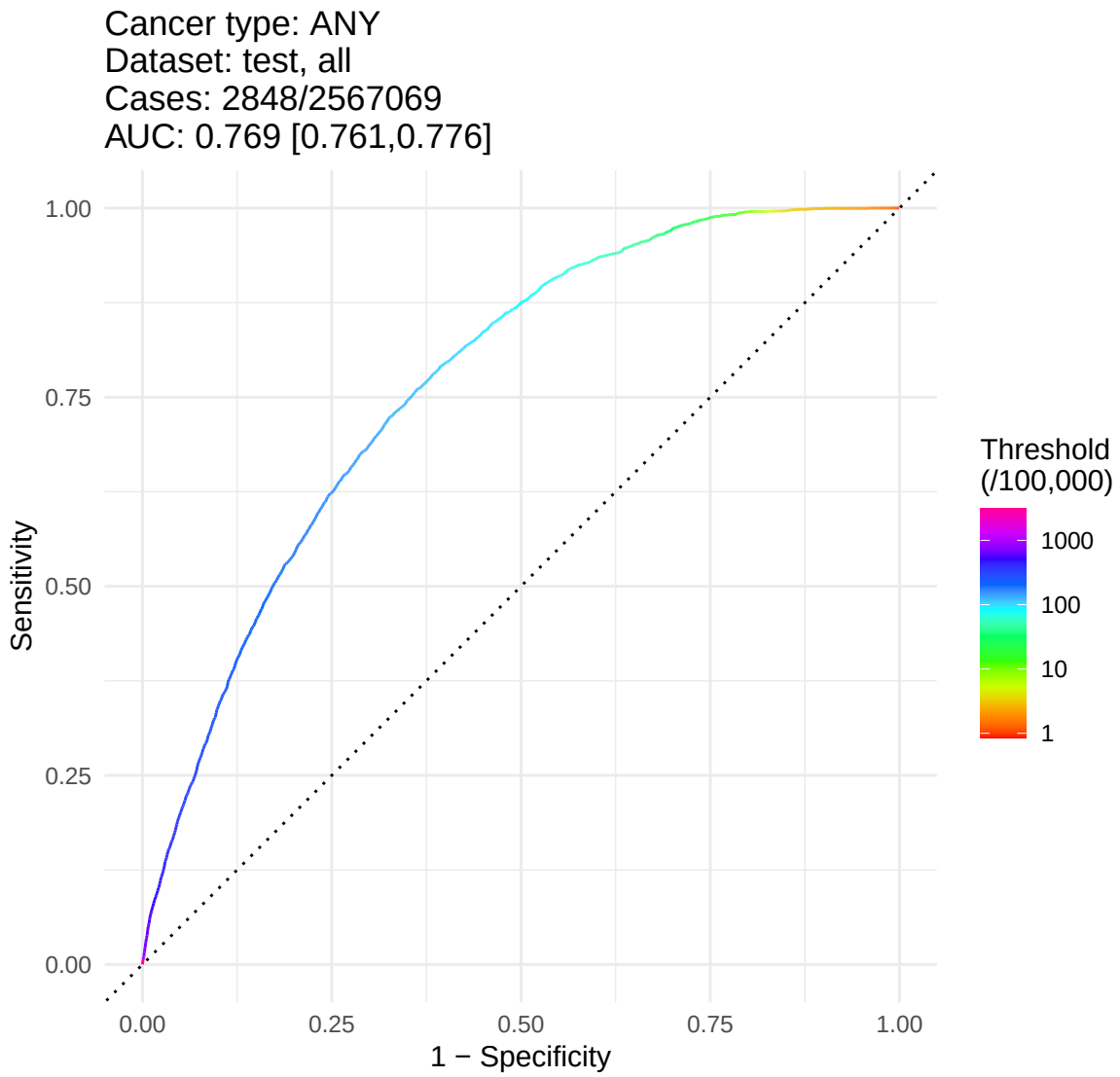
HOSEA Aim I – Results (SRS)

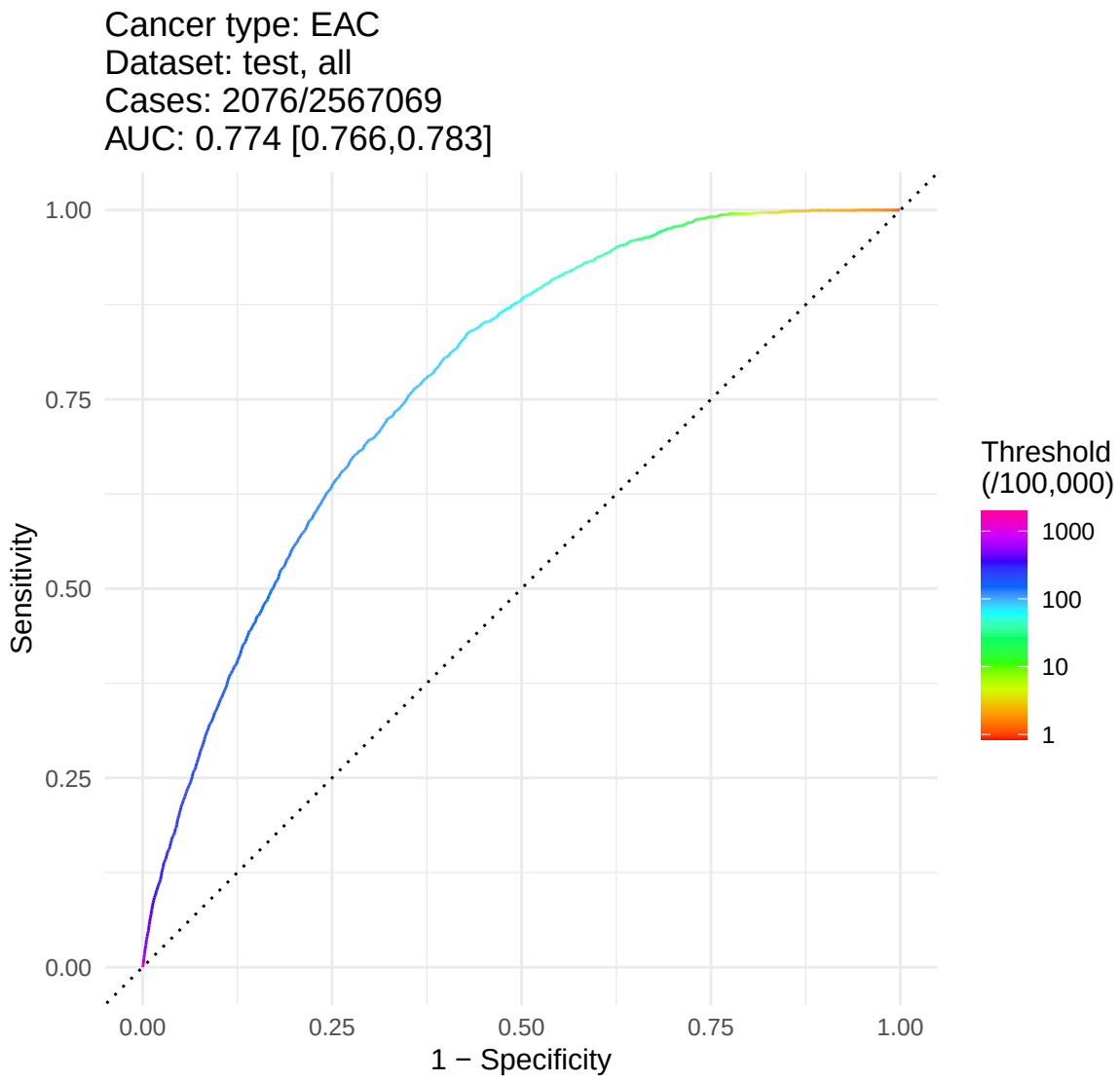
Simon Fontaine

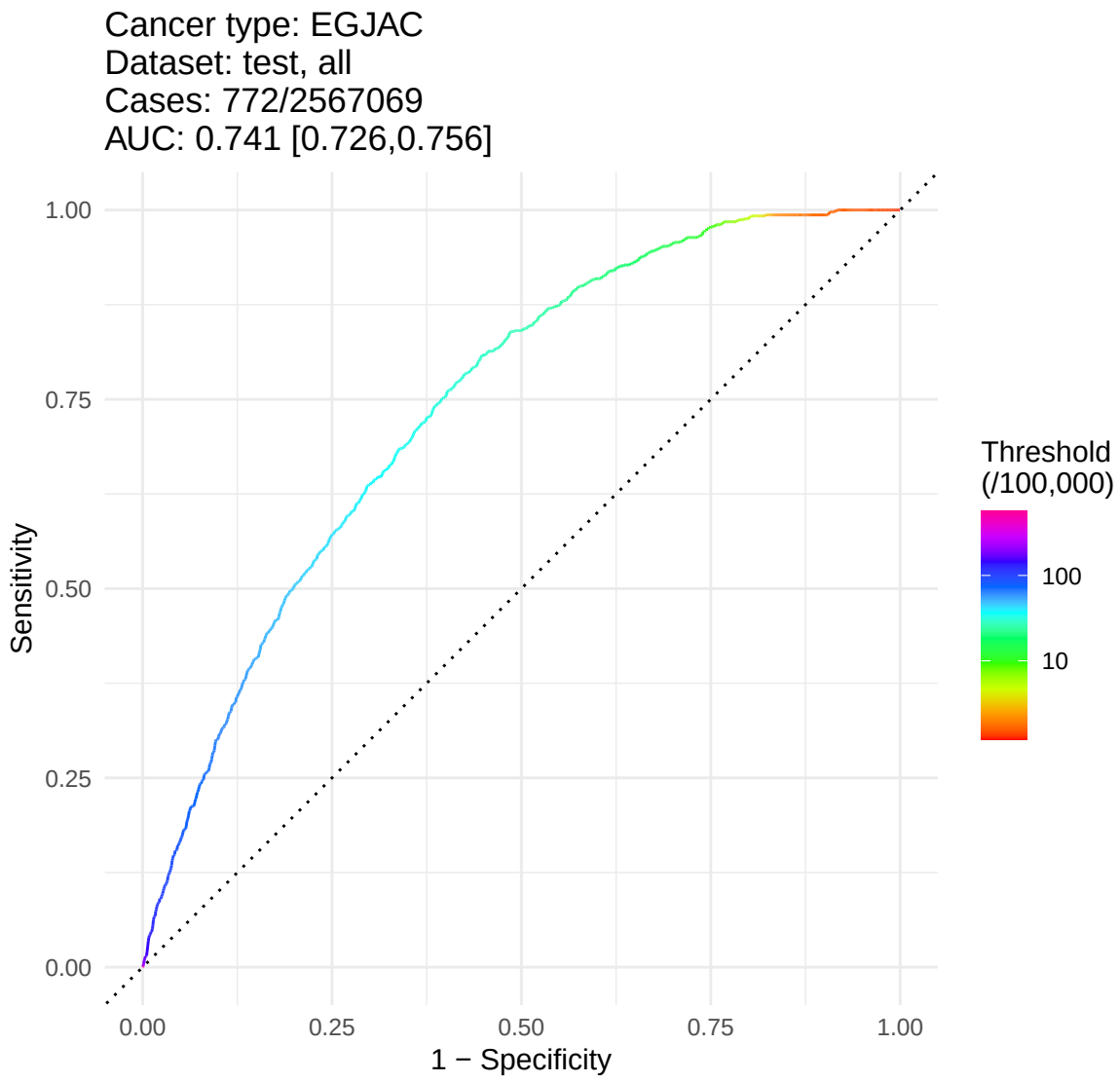
November 8, 2022

Contents

1 Overall performance

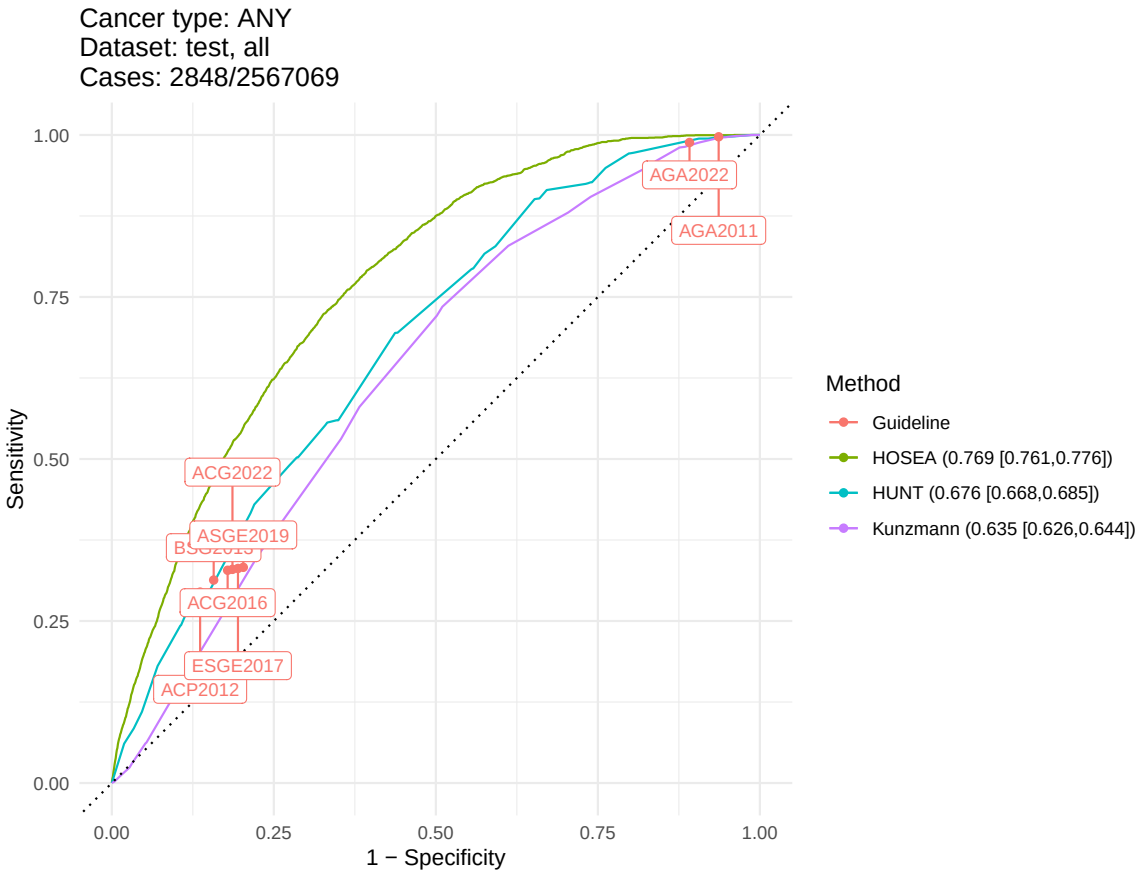


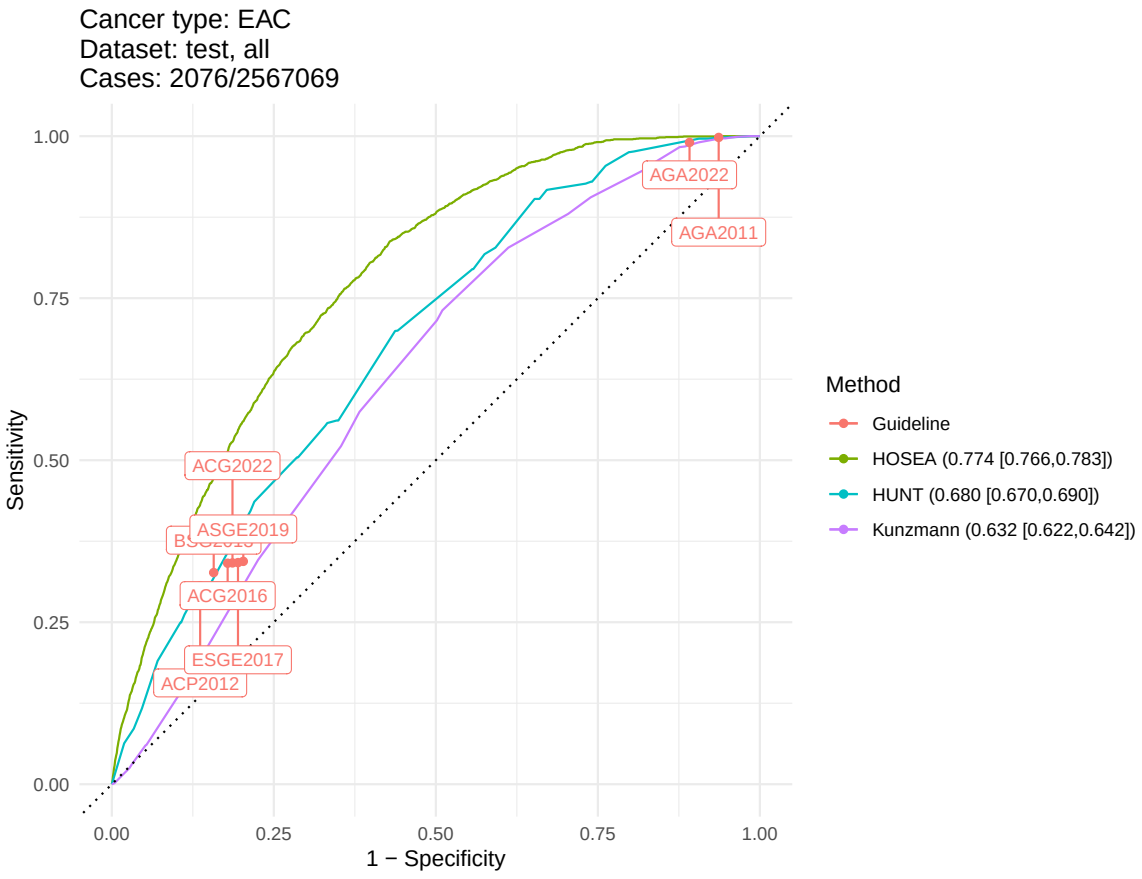

Cancer type: EAC
Dataset: test, all
Cases: 2076/2567069
AUC: 0.774 [0.766,0.783]
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Threshold
(/100,000)
1000
100
10
1

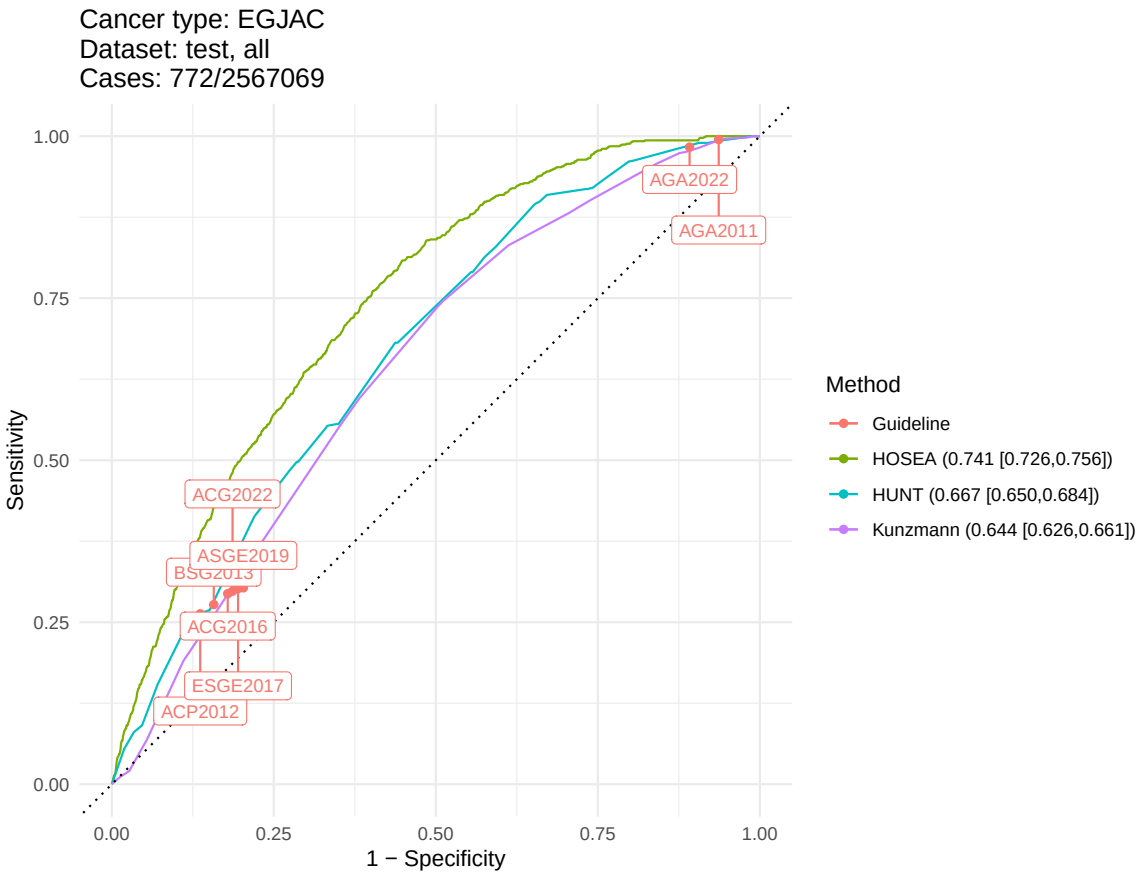

Cancer type: EGJAC
Dataset: test, all
Cases: 772/2567069
AUC: 0.741 [0.726,0.756]
Sensitivity
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
1 − Specificity
Threshold
(/100,000)
100
10

2 Comparison to HUNT, Kunzmann and Guidelines

2.1 Full test data

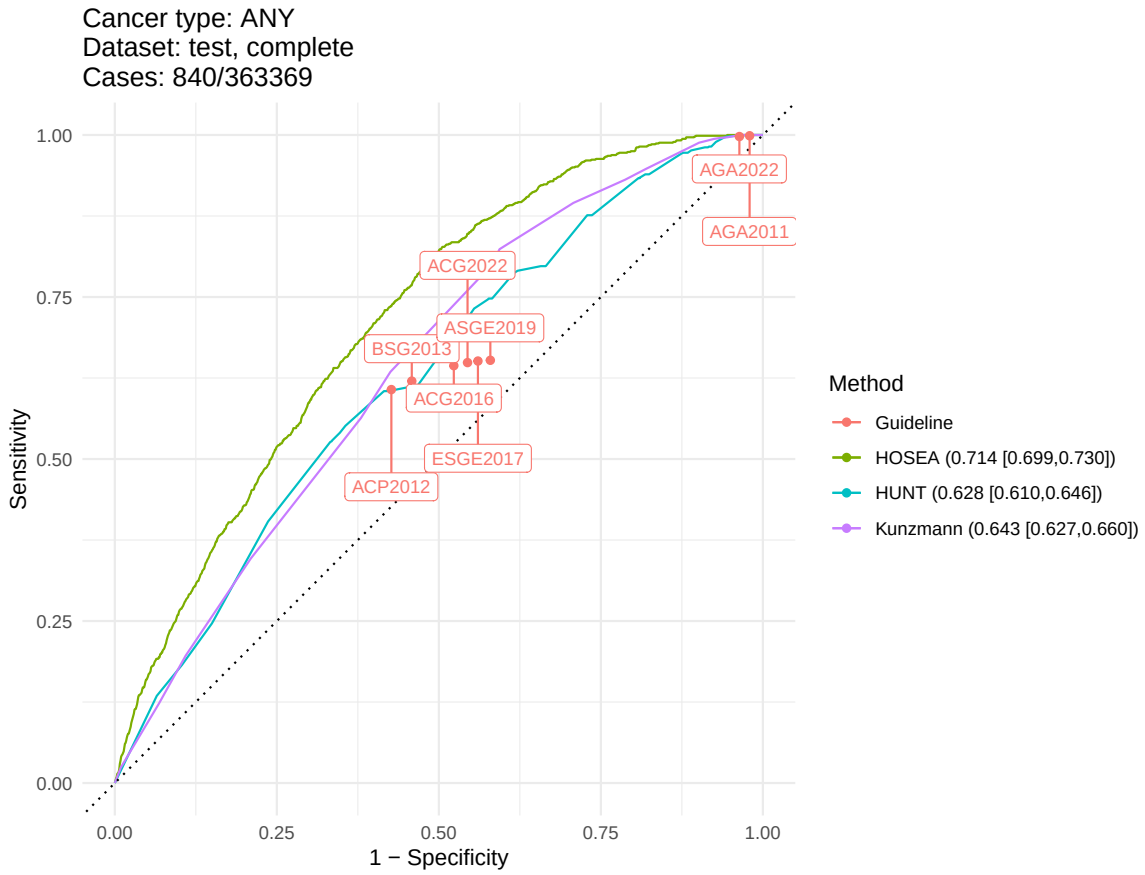


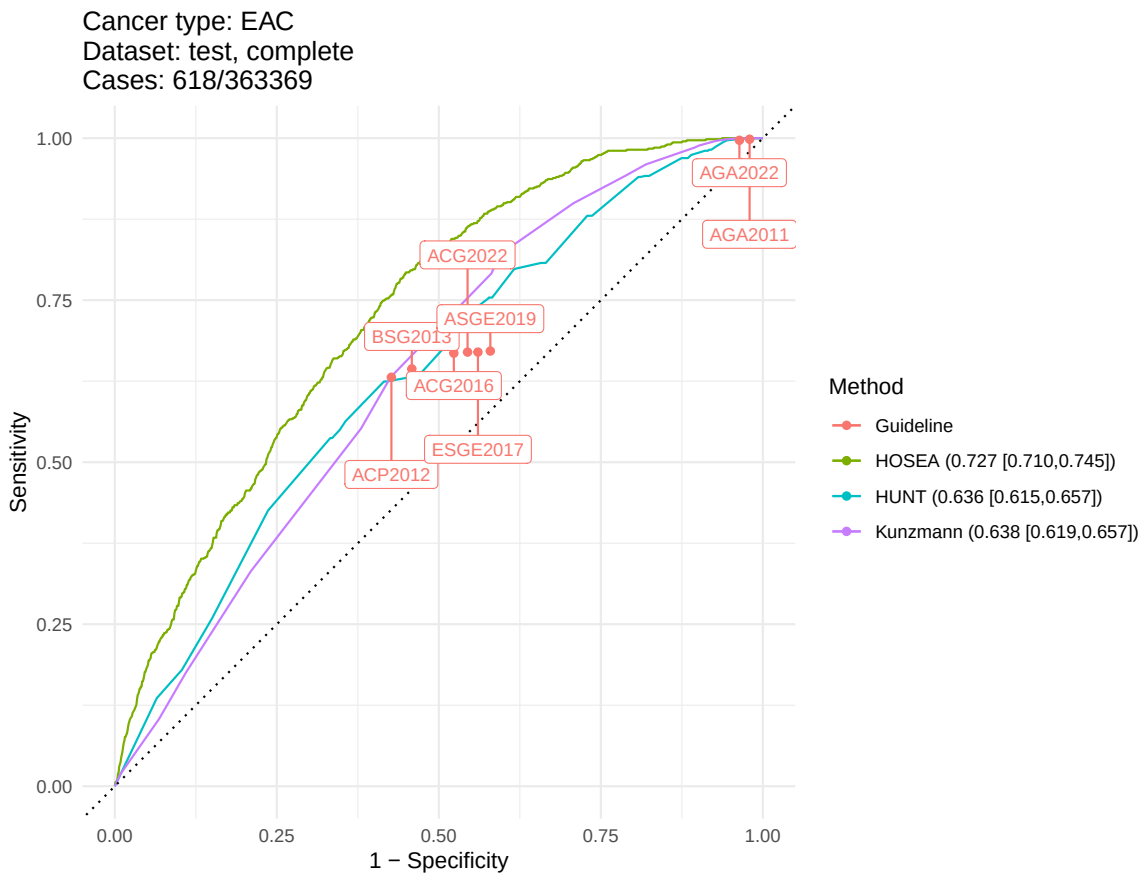

Cancer type: EAC
Dataset: test, all
Cases: 2076/2567069
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
AGA2022
AGA2011
ACG2022
ASGE2019
ACG2016
ESGE2017
ACP2012
Method
Guideline
HOSEA (0.774 [0.766,0.783])
HUNT (0.680 [0.670,0.690])
Kunzmann (0.632 [0.622,0.642])

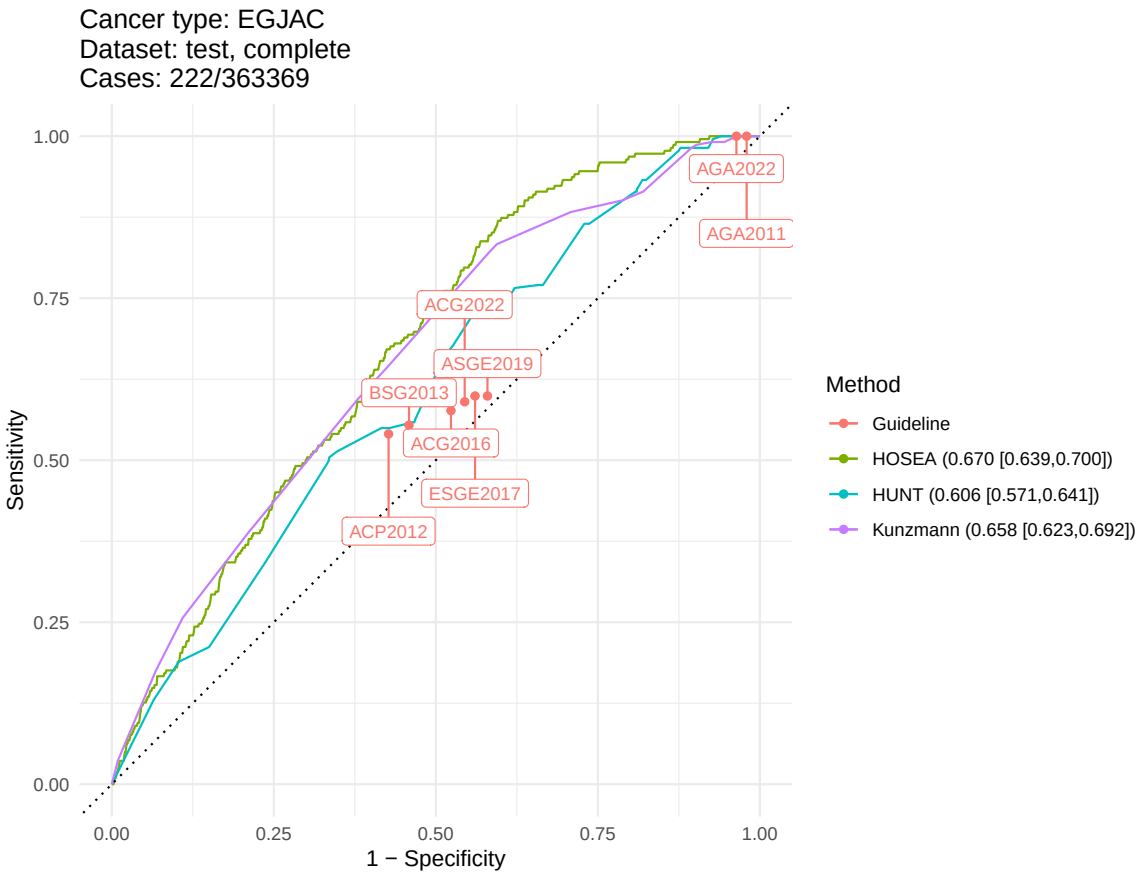

Cancer type: EGJAC
Dataset: test, all
Cases: 772/2567069
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
ACG2022
ASGE2019
BSG2013
ACG2016
ESGE2017
ACP2012
AGA2022
AGA2011
Method
Guideline
HOSEA (0.741 [0.726,0.756])
HUNT (0.667 [0.650,0.684])
Kunzmann (0.644 [0.626,0.661])

2.2 Complete test patients

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

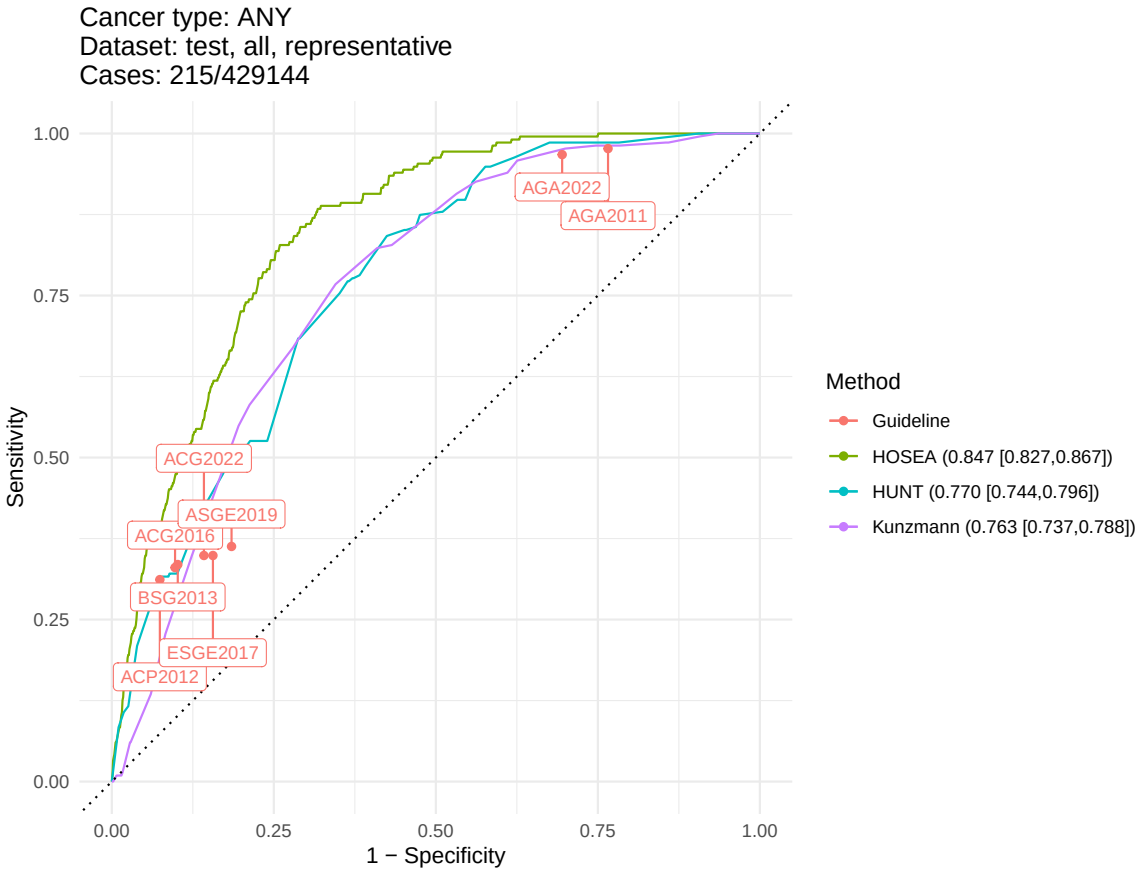


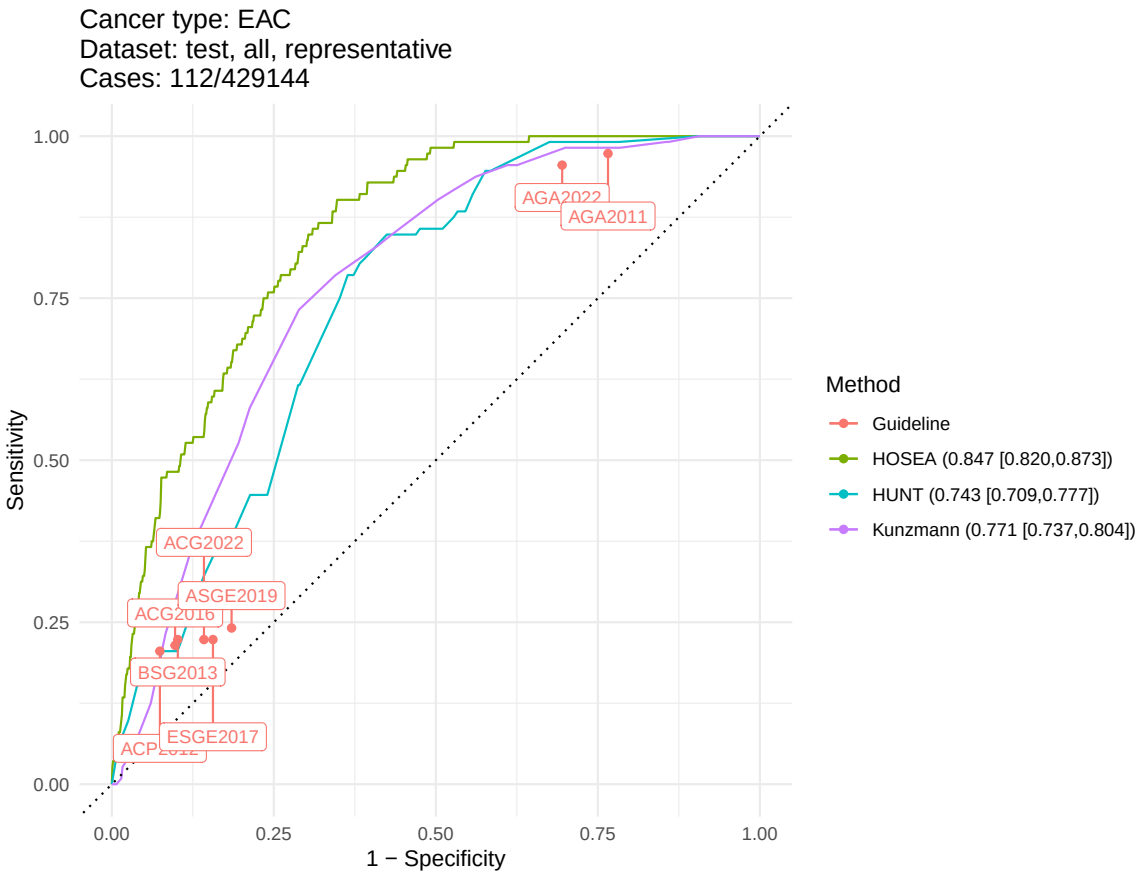

Cancer type: EAC
Dataset: test, complete
Cases: 618/363369
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
AGA2022
AGA2011
ACG2022
ASGE2019
BSG2013
ACG2016
ESGE2017
ACP2012
Method
Guideline
HOSEA (0.727 [0.710,0.745])
HUNT (0.636 [0.615,0.657])
Kunzmann (0.638 [0.619,0.657])

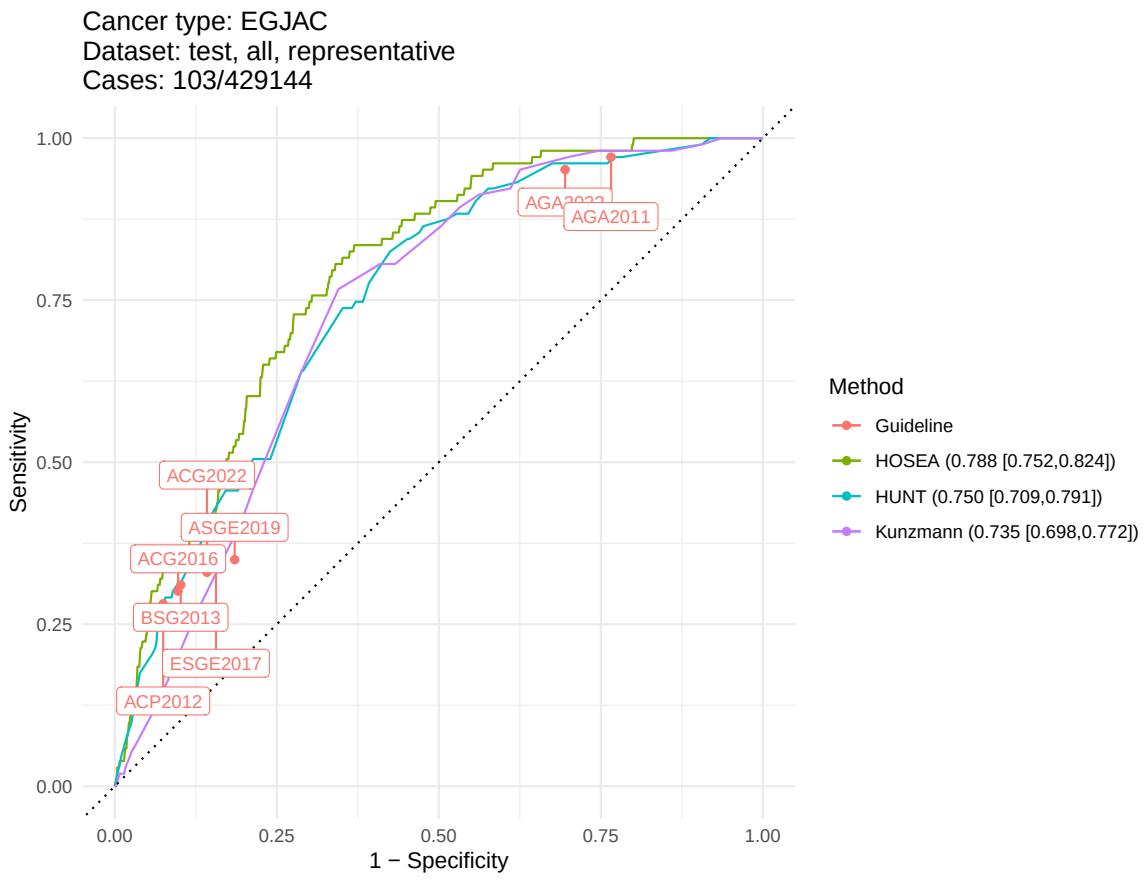

Cancer type: EGJAC
Dataset: test, complete
Cases: 222/363369
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
AGA2022
AGA2011
ACG2022
ASGE2019
BSG2013
ACG2016
ESGE2017
ACP2012
Method
Guideline
HOSEA (0.670 [0.639,0.700])
HUNT (0.606 [0.571,0.641])
Kunzmann (0.658 [0.623,0.692])

2.3 Representative sample

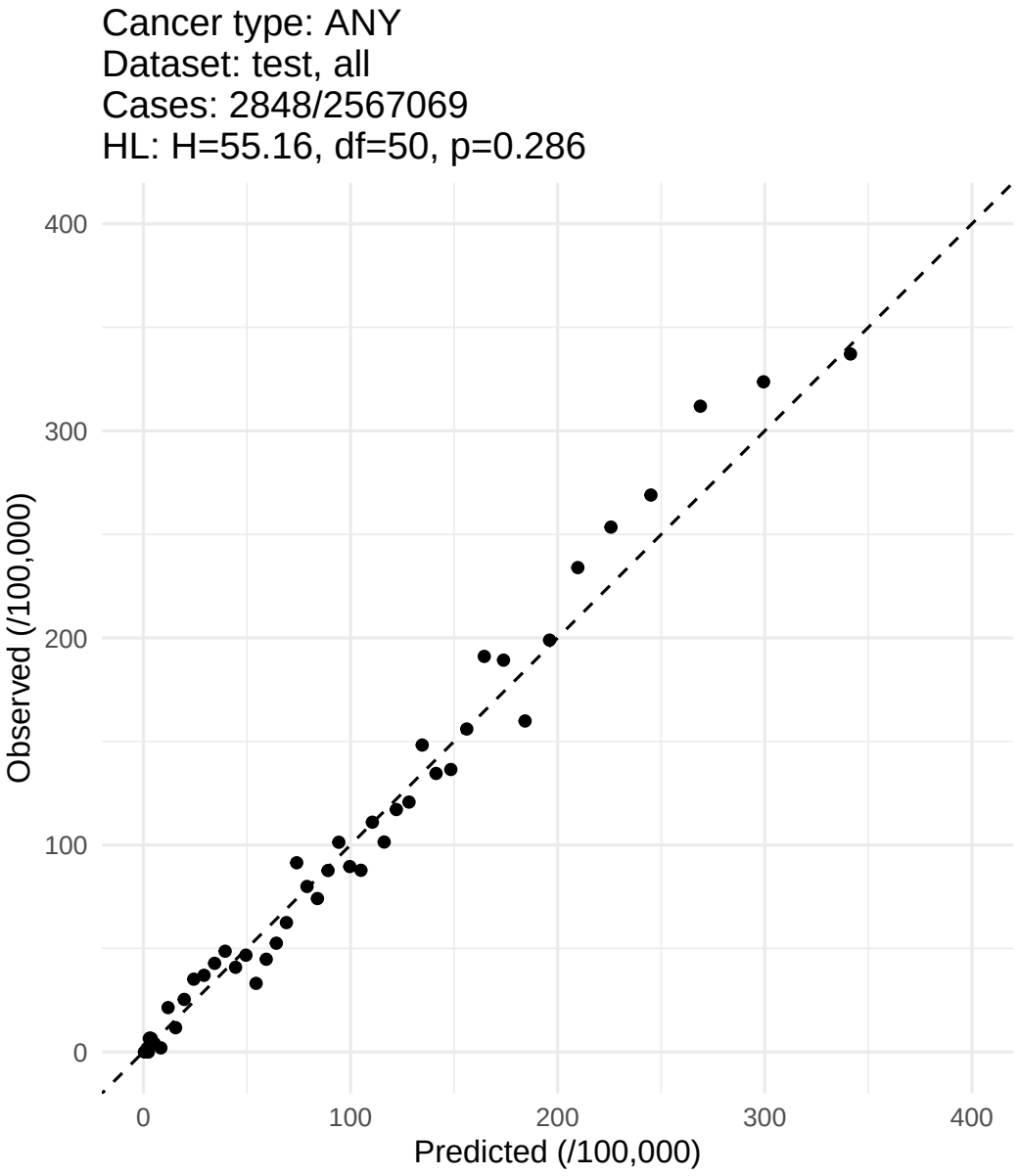
w.r.t. sex and prevalance ratio by sex

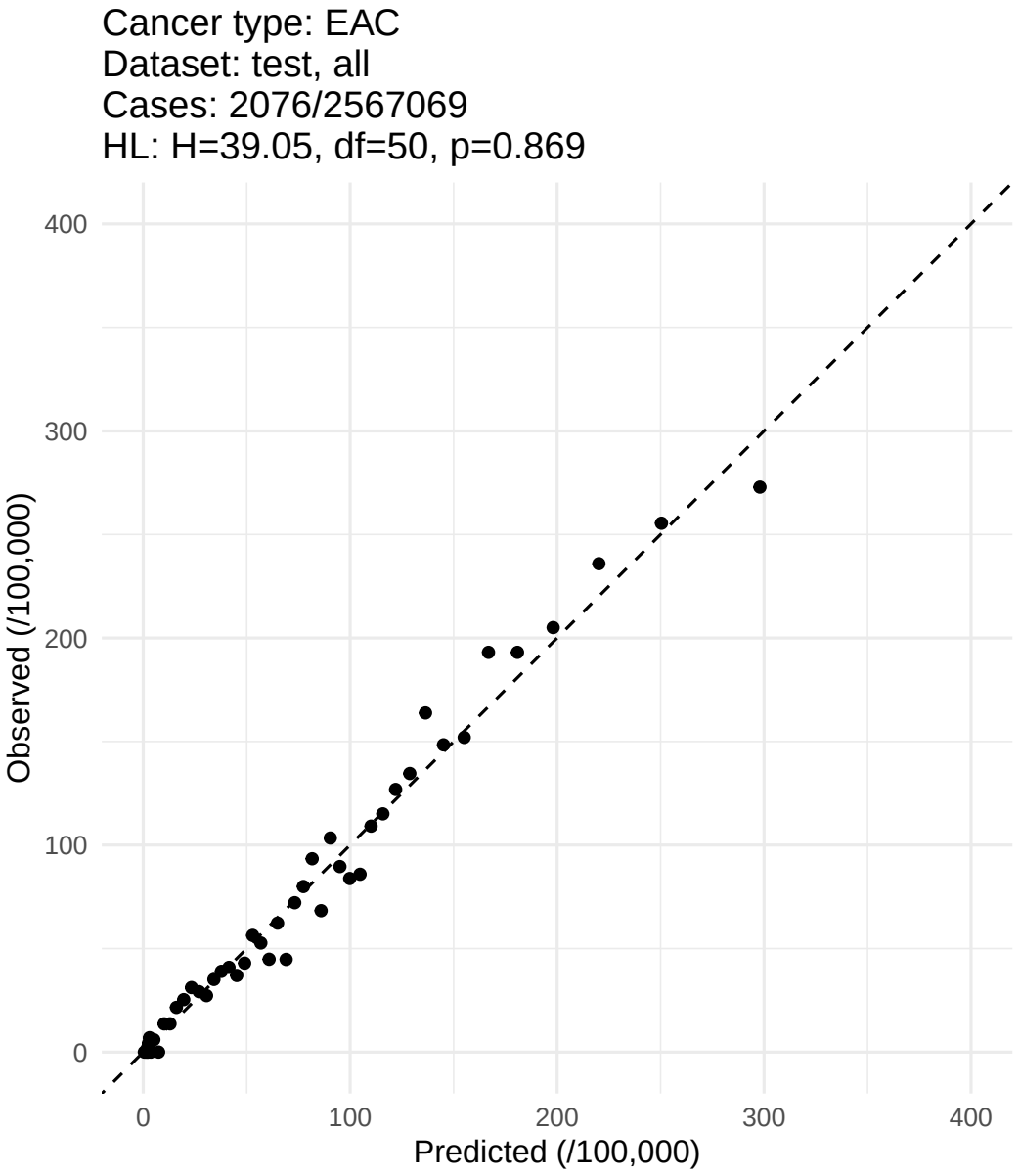


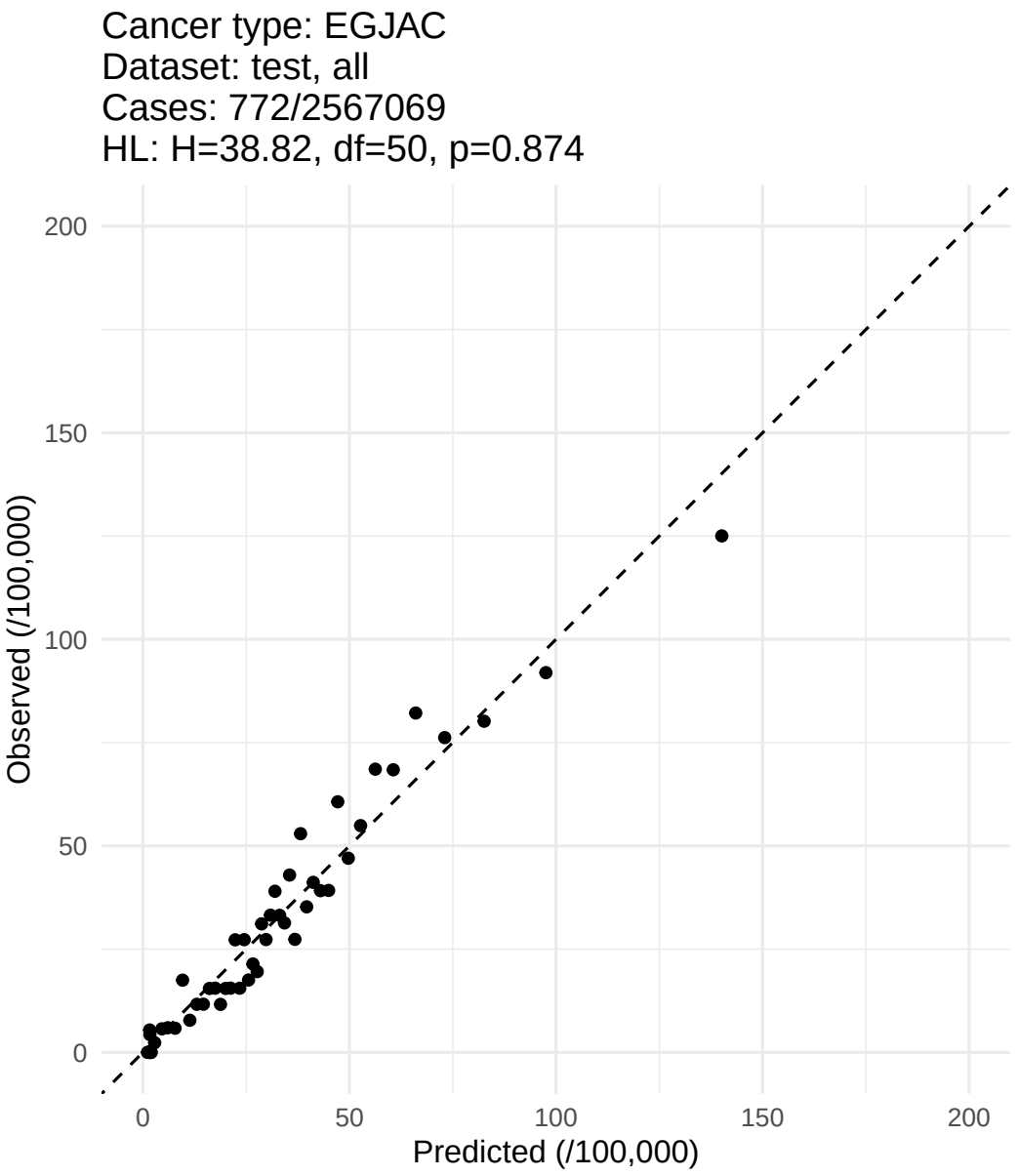

Cancer type: EAC
Dataset: test, all, representative
Cases: 112/429144
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
AGA2022
AGA2011
ACG2022
ASGE2019
ACG2016
BSG2013
ESGE2017
Method
Guideline
HOSEA (0.847 [0.820,0.873])
HUNT (0.743 [0.709,0.777])
Kunzmann (0.771 [0.737,0.804])


Cancer type: EGJAC
Dataset: test, all, representative
Cases: 103/429144
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
ACP2012
ESGE2017
BSG2013
ACG2016
ASGE2019
ACG2022
AGA2011
Method
Guideline
HOSEA (0.788 [0.752,0.824])
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=39.05, df=50, p=0.869
Observed (/100,000)
Predicted (/100,000)
0
100
200
300
400


Cancer type: EGJAC
Dataset: test, all
Cases: 772/2567069
HL: H=38.82, df=50, p=0.874
Observed (/100,000)
Predicted (/100,000)
0
50
100
150
200

4 Thresholds

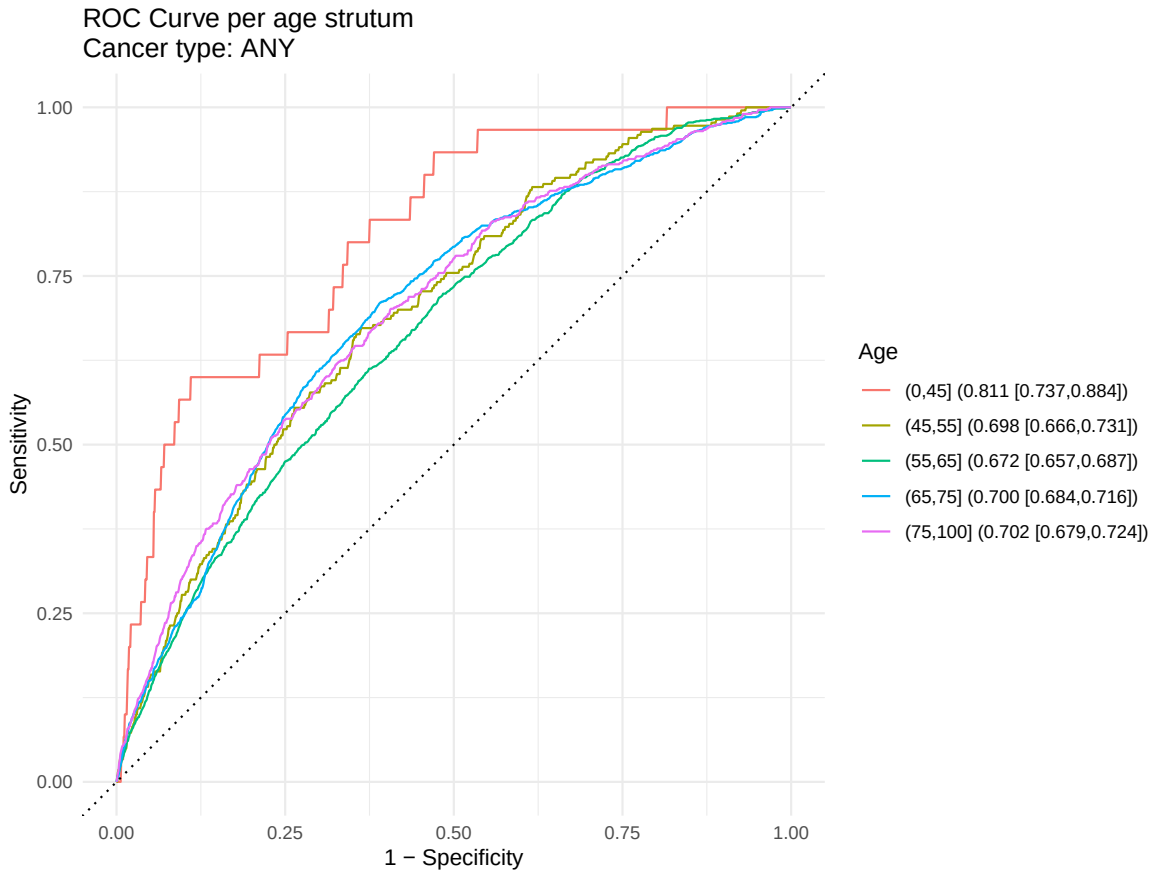
Threshold	TPR	PPV	DetPrevalence
0	100.00	0.11	100.00
5	99.58	0.13	83.11
10	99.51	0.14	80.06
15	99.09	0.14	77.24
20	98.67	0.15	74.83
25	98.03	0.15	72.68
30	97.44	0.15	70.68
35	96.52	0.16	68.73
40	95.68	0.16	66.76
45	95.05	0.16	64.76
50	94.07	0.17	62.76
55	93.61	0.17	60.72
60	92.70	0.18	58.68
65	91.99	0.18	56.63
70	90.77	0.18	54.58
75	89.40	0.19	52.57
80	87.75	0.19	50.56
85	86.31	0.20	48.58
90	84.90	0.20	46.63
95	83.22	0.21	44.72
100	81.85	0.21	42.82
105	80.16	0.22	40.98
110	78.69	0.22	39.16
115	76.86	0.23	37.40
120	75.21	0.23	35.69
125	73.53	0.24	34.02
130	71.88	0.25	32.41
135	69.84	0.25	30.84
140	67.94	0.26	29.35
145	66.08	0.26	27.92
150	64.50	0.27	26.55
155	62.61	0.28	25.25
160	61.06	0.28	24.03
165	58.99	0.29	22.87
170	57.23	0.29	21.77
175	55.62	0.30	20.75
180	53.86	0.30	19.77
185	52.77	0.31	18.84
190	51.23	0.32	17.98
195	49.82	0.32	17.15
200	48.31	0.33	16.37
220	42.80	0.35	13.65
240	37.68	0.36	11.46
260	32.79	0.38	9.67
280	28.90	0.39	8.18
300	24.89	0.40	6.94
320	22.47	0.42	5.89
340	20.01	0.44	5.01
360	17.45	0.45	4.28
380	15.59	0.47	3.66
400	13.97	0.50	3.13
420	12.11	0.50	2.69
440	10.85	0.52	2.31

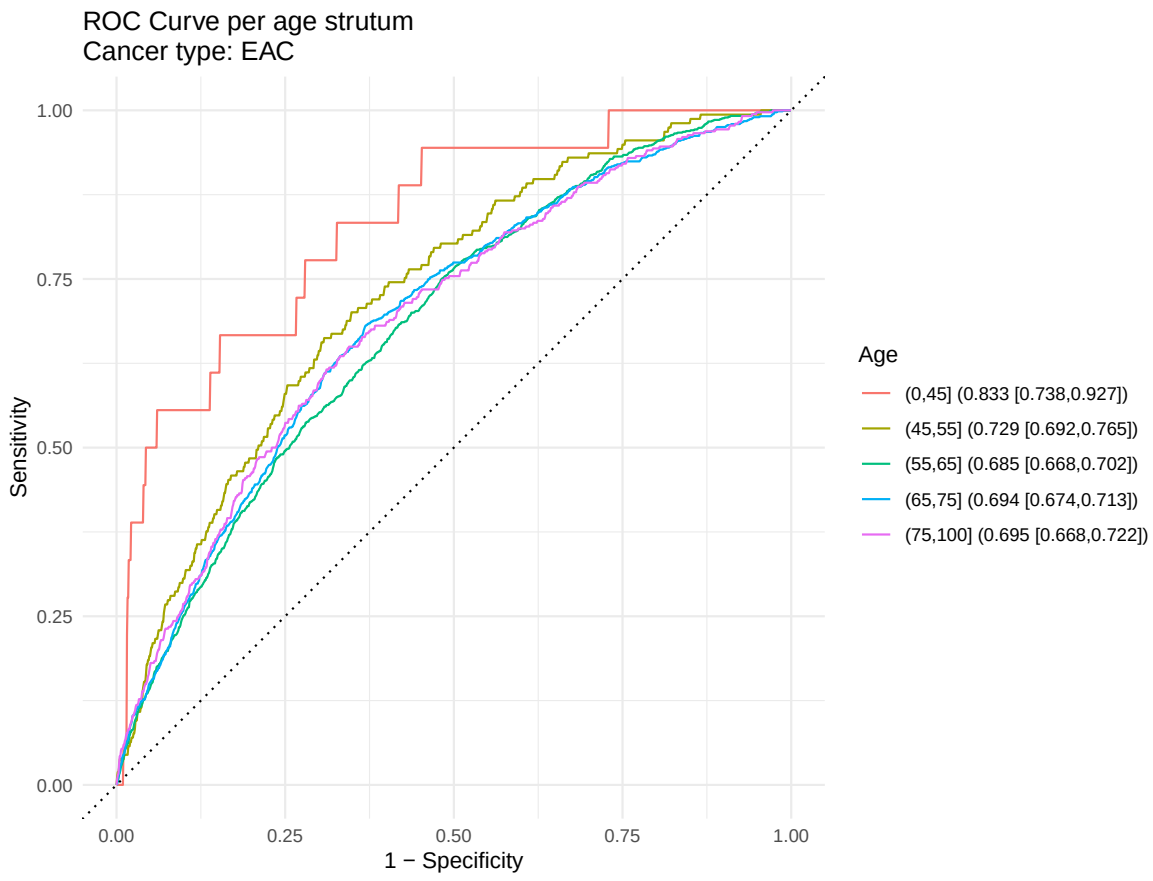
Threshold	TPR	PPV	DetPrevalence
0	100.00	0.08	100.00
5	99.61	0.10	80.99
10	99.42	0.10	77.02
15	98.80	0.11	73.56
20	97.88	0.11	70.72
25	97.01	0.12	68.05
30	96.05	0.12	65.27
35	95.04	0.12	62.48
40	93.50	0.13	59.73
45	92.24	0.13	57.05
50	90.99	0.14	54.42
55	89.31	0.14	51.88
60	87.76	0.14	49.37
65	86.08	0.15	46.91
70	84.63	0.15	44.50
75	82.61	0.16	42.09
80	80.25	0.16	39.73
85	77.75	0.17	37.39
90	75.39	0.17	35.13
95	72.74	0.18	32.97
100	70.33	0.18	30.86
105	68.26	0.19	28.88
110	66.09	0.20	27.00
115	63.97	0.21	25.23
120	61.22	0.21	23.58
125	58.96	0.22	22.06
130	56.65	0.22	20.64
135	54.14	0.23	19.31
140	52.02	0.23	18.10
145	49.71	0.24	16.99
150	47.74	0.24	15.95
155	45.91	0.25	14.99
160	44.36	0.25	14.09
165	42.44	0.26	13.26
170	40.37	0.26	12.48
175	38.92	0.27	11.75
180	37.09	0.27	11.08
185	35.55	0.28	10.44
190	34.15	0.28	9.85
195	32.90	0.29	9.29
200	31.89	0.29	8.77
220	26.35	0.30	6.99
240	22.54	0.33	5.57
260	18.59	0.34	4.46
280	15.70	0.35	3.58
300	13.87	0.39	2.88
320	11.37	0.40	2.31
340	10.16	0.44	1.85
360	8.96	0.49	1.49
380	7.56	0.51	1.20
400	6.21	0.52	0.97
420	4.87	0.50	0.78
440	4.14	0.53	0.63

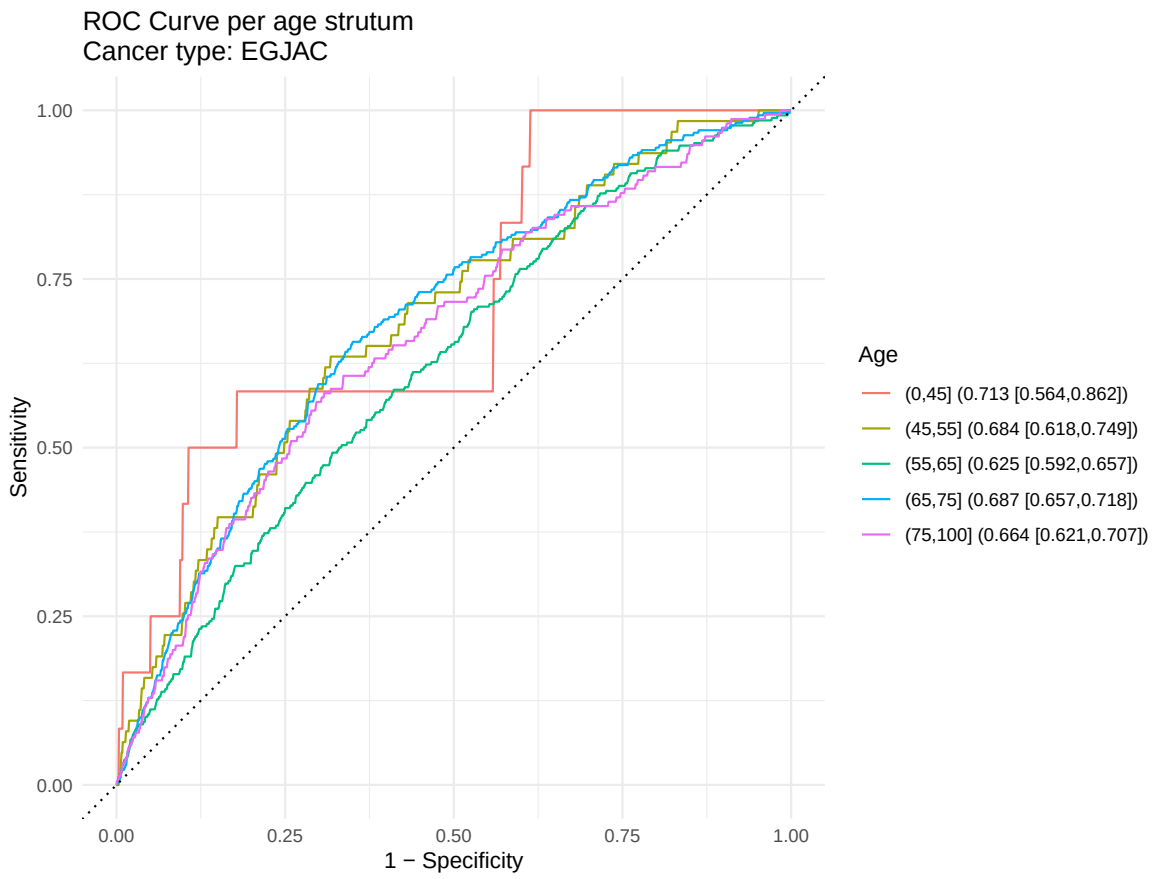
Threshold	TPR	PPV	DetPrevalence
0	100.00	0.03	100.00
5	99.09	0.04	80.31
10	97.28	0.04	74.43
15	94.95	0.04	68.37
20	91.32	0.05	60.86
25	85.23	0.05	51.88
30	78.11	0.06	42.42
35	67.88	0.06	33.64
40	58.55	0.07	26.39
45	51.30	0.07	20.84
50	44.30	0.08	16.70
55	38.08	0.08	13.59
60	32.77	0.09	11.20
65	28.37	0.09	9.29
70	24.48	0.10	7.75
75	21.24	0.10	6.48
80	18.13	0.10	5.43
85	15.54	0.10	4.54
90	13.47	0.11	3.82
95	11.27	0.11	3.20
100	9.97	0.11	2.69
105	9.07	0.12	2.26
110	8.03	0.13	1.90
115	6.74	0.13	1.61
120	5.44	0.12	1.35
125	4.66	0.12	1.14
130	4.27	0.13	0.97
135	3.89	0.14	0.82
140	3.11	0.13	0.70
145	2.07	0.10	0.60
150	1.55	0.09	0.51
155	1.42	0.10	0.44
160	1.42	0.11	0.37
165	1.30	0.12	0.32
170	1.04	0.11	0.28
175	0.91	0.11	0.24
180	0.78	0.11	0.21
185	0.65	0.11	0.18
190	0.52	0.10	0.16
195	0.52	0.11	0.14
200	0.52	0.13	0.12
220	0.13	0.05	0.07
240	0.13	0.09	0.04
260	0.13	0.14	0.03
280	0.00	0.00	0.02
300	0.00	0.00	0.01
320	0.00	0.00	0.01
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

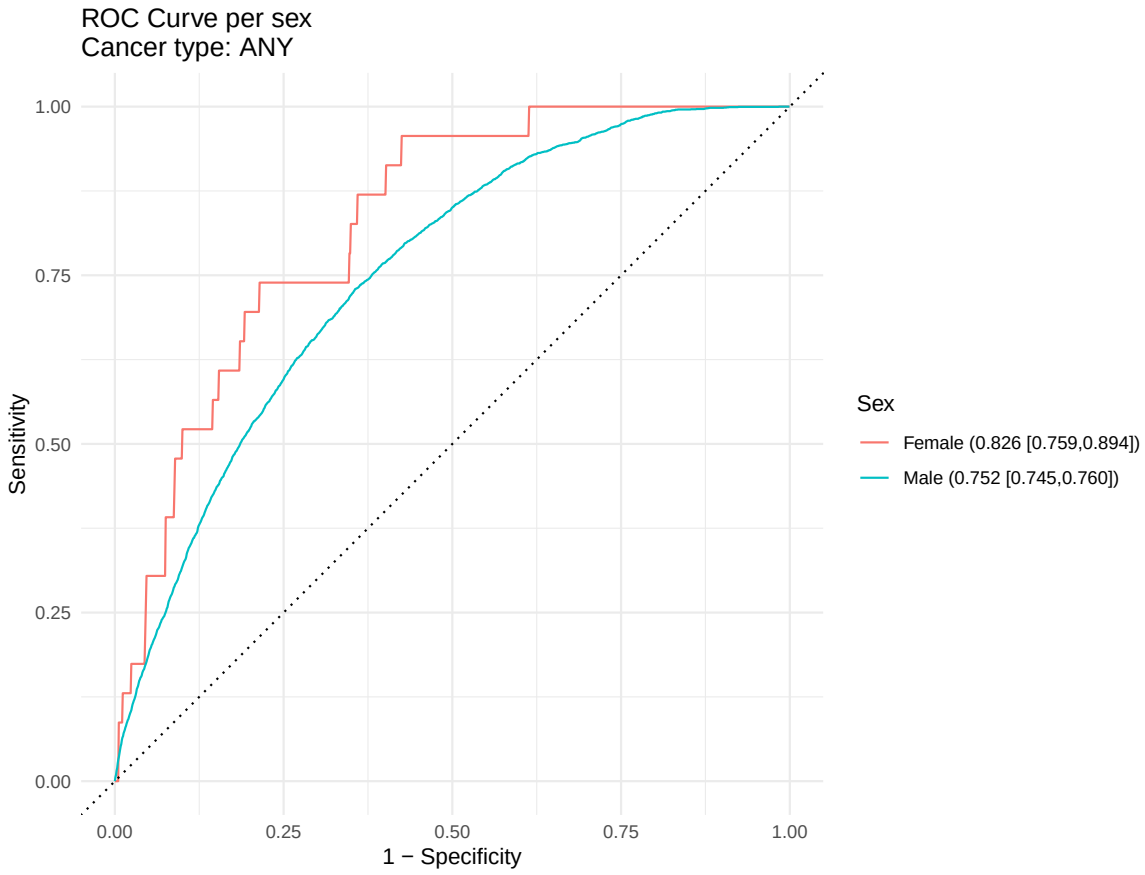
5.1 Age

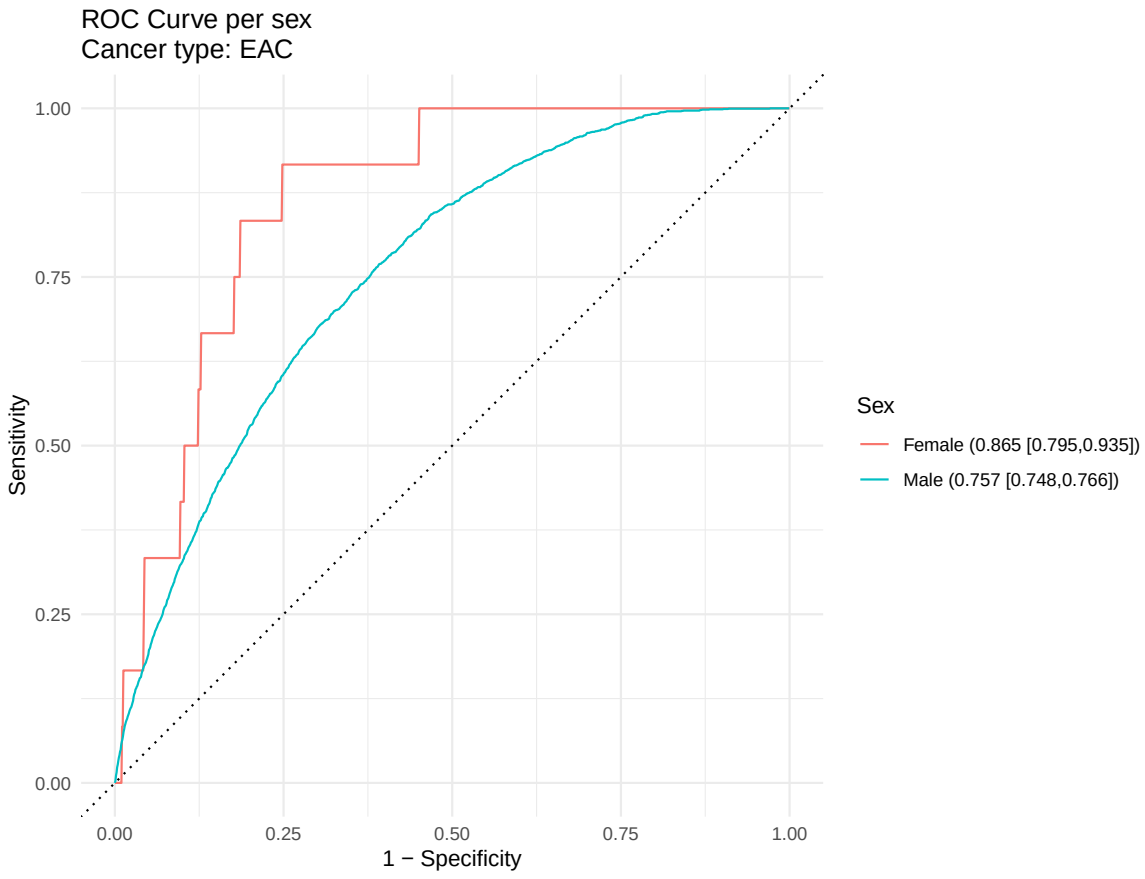


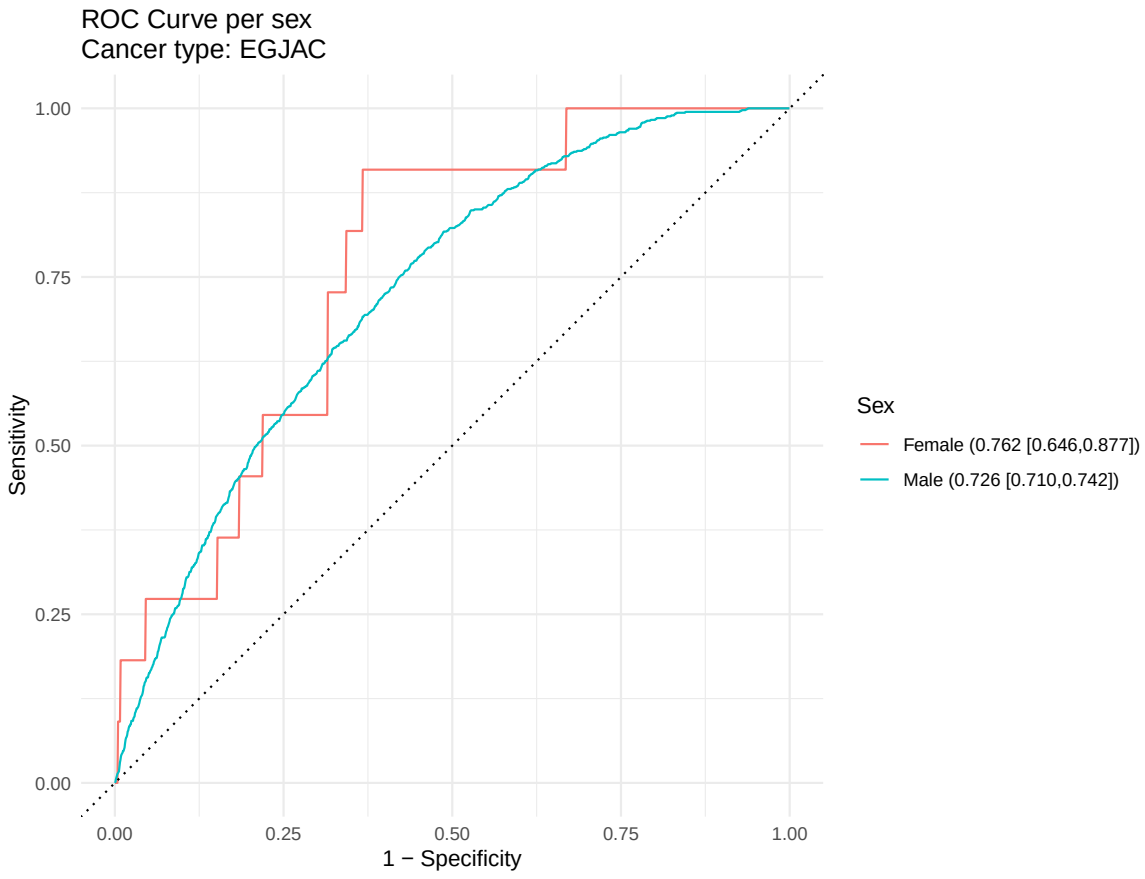

ROC Curve per age strutum
Cancer type: EAC
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Age
(0,45] (0.833 [0.738,0.927])
(45,55] (0.729 [0.692,0.765])
(55,65] (0.685 [0.668,0.702])
(65,75] (0.694 [0.674,0.713])
(75,100] (0.695 [0.668,0.722])


ROC Curve per age strutum
Cancer type: EGJAC
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Age
(0,45] (0.713 [0.564,0.862])
(45,55] (0.684 [0.618,0.749])
(55,65] (0.625 [0.592,0.657])
(65,75] (0.687 [0.657,0.718])
(75,100] (0.664 [0.621,0.707])

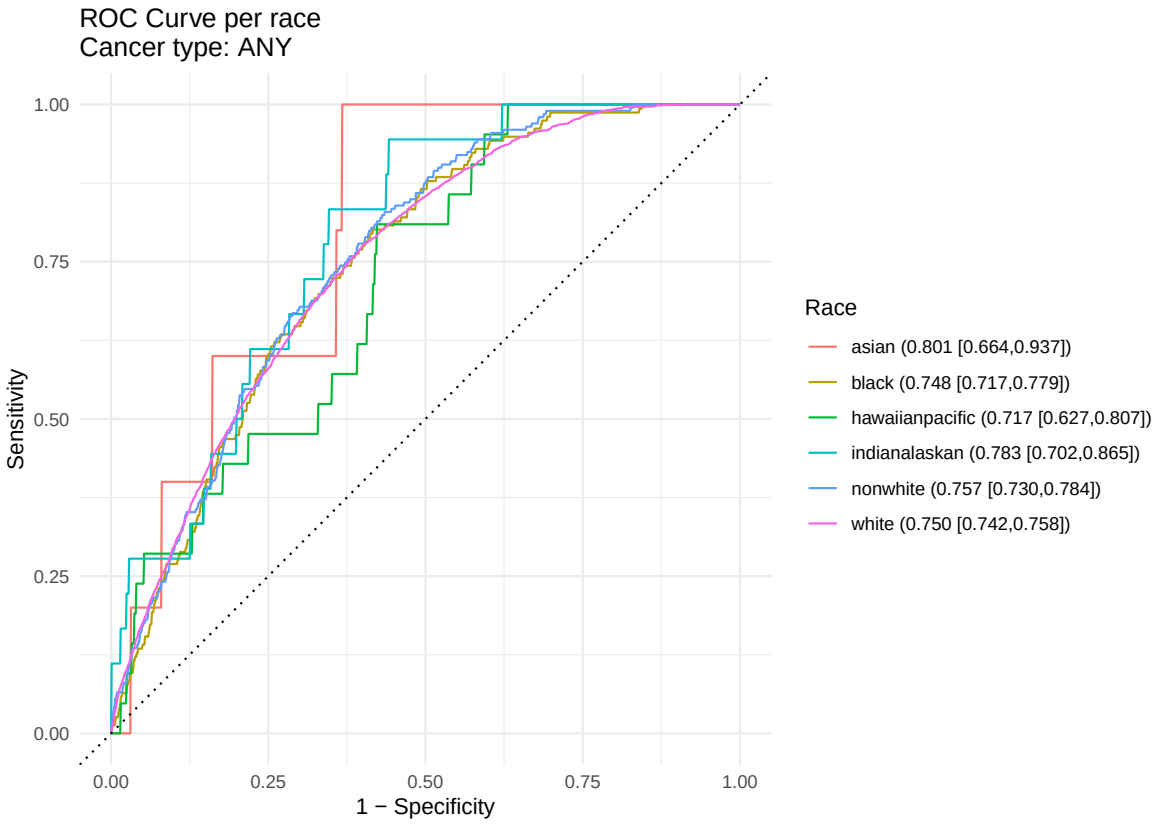
5.2 Sex

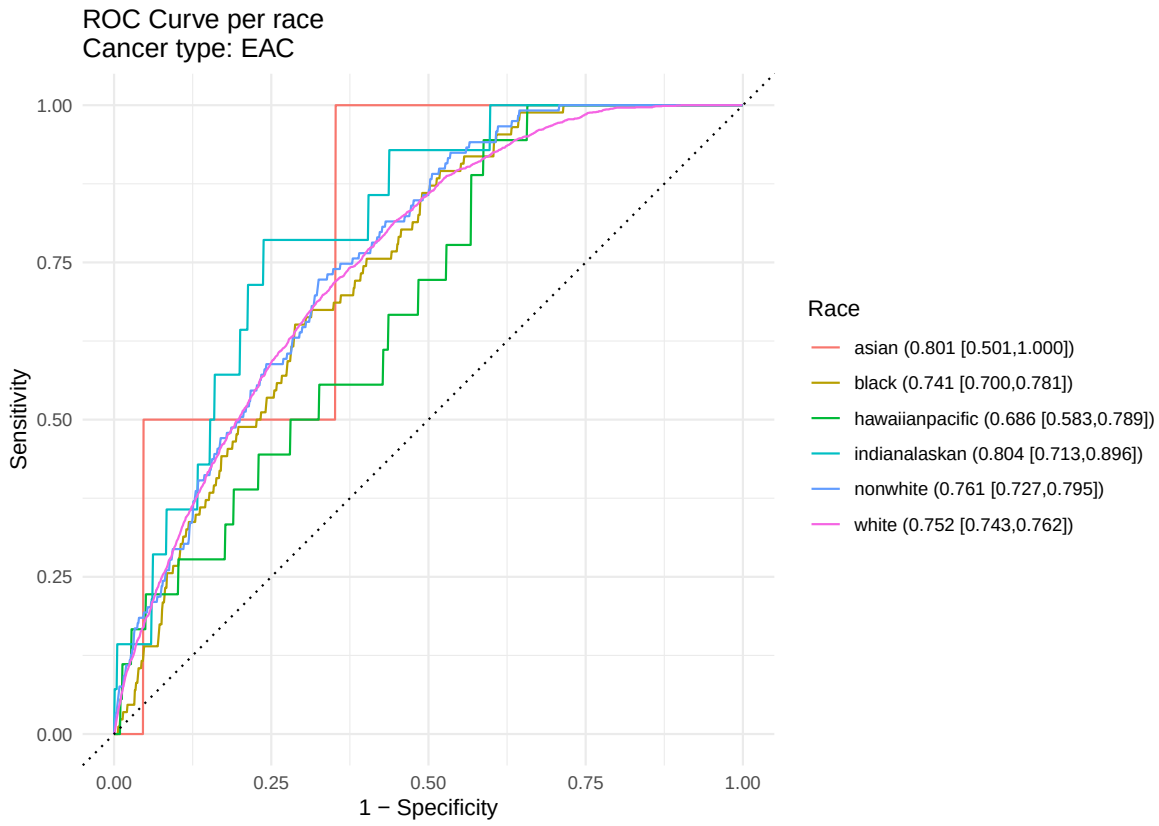

ROC Curve per sex
Cancer type: ANY
Sensitivity
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
1 − Specificity
Sex
Female (0.826 [0.759,0.894])
Male (0.752 [0.745,0.760])

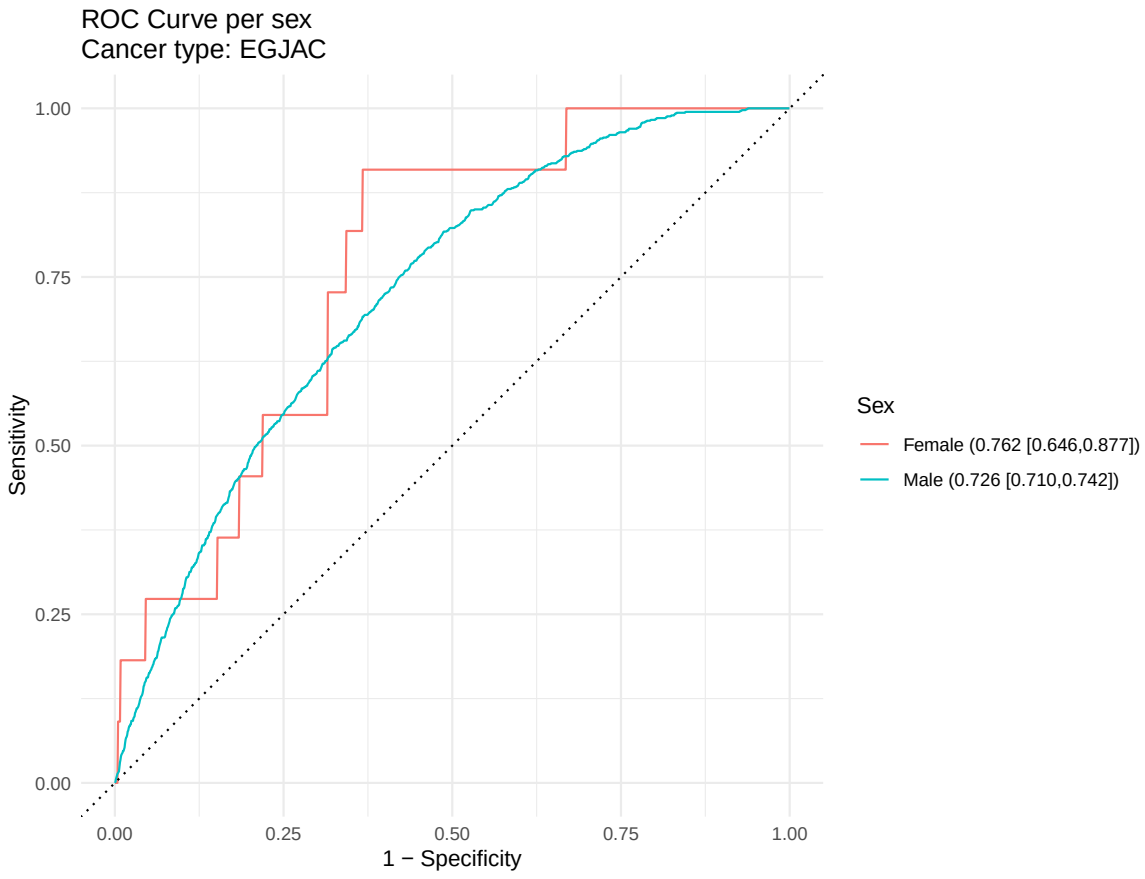

ROC Curve per sex
Cancer type: EAC
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Sex
Female (0.865 [0.795,0.935])
Male (0.757 [0.748,0.766])


ROC Curve per sex
Cancer type: EGJAC
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Sex
Female (0.762 [0.646,0.877])
Male (0.726 [0.710,0.742])

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 (0.801 [0.664,0.937])
black (0.748 [0.717,0.779])
hawaiianpacific (0.717 [0.627,0.807])
indianalaskan (0.783 [0.702,0.865])
nonwhite (0.757 [0.730,0.784])
white (0.750 [0.742,0.758])


ROC Curve per race
Cancer type: EAC
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Race
asian (0.801 [0.501,1.000])
black (0.741 [0.700,0.781])
hawaiianpacific (0.686 [0.583,0.789])
indianalaskan (0.804 [0.713,0.896])
nonwhite (0.761 [0.727,0.795])
white (0.752 [0.743,0.762])


ROC Curve per sex
Cancer type: EGJAC
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Sex
Female (0.762 [0.646,0.877])
Male (0.726 [0.710,0.742])

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.

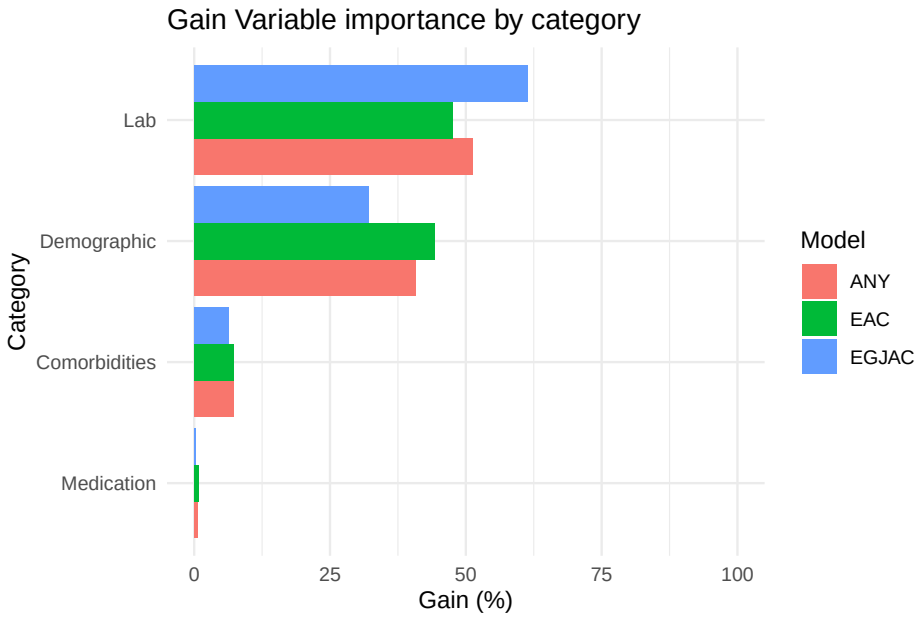
Stage	Test. AUC	Nb. cases
Any	0.768 [0.761,0.776]	2818
I	0.807 [0.790,0.824]	495
II	0.787 [0.768,0.806]	435
III	0.774 [0.761,0.787]	878
IV	0.762 [0.750,0.773]	1244
I+	0.772 [0.764,0.780]	2448
II+	0.765 [0.757,0.774]	2169
III+	0.763 [0.754,0.772]	1950
IV+	0.762 [0.750,0.773]	1244

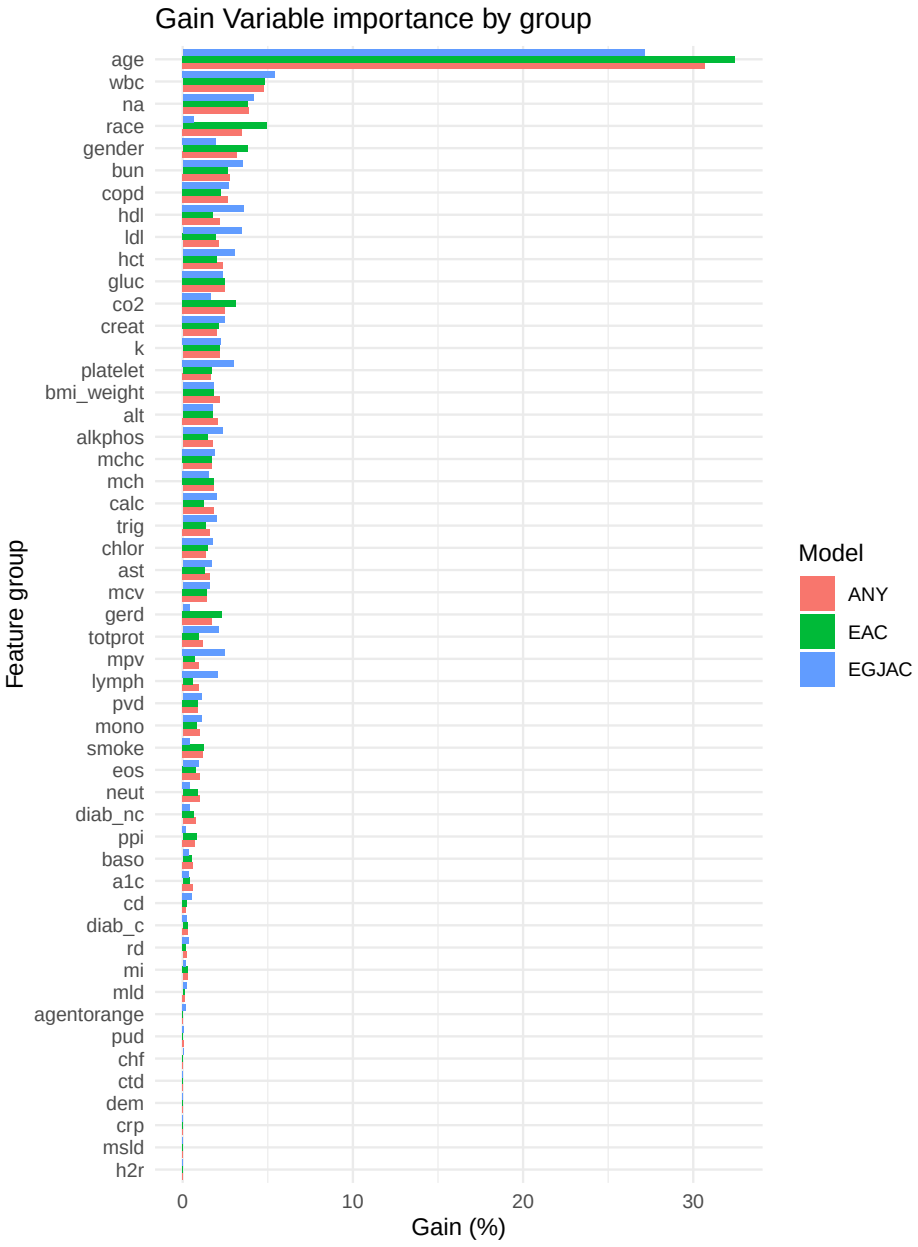
Stage	Test. AUC	Nb. cases
Any	0.774 [0.766,0.783]	2054
I	0.816 [0.797,0.835]	350
II	0.790 [0.768,0.813]	302
III	0.774 [0.759,0.790]	650
IV	0.768 [0.755,0.781]	913
I+	0.775 [0.766,0.784]	1813
II+	0.768 [0.758,0.778]	1608
III+	0.767 [0.757,0.777]	1451
IV+	0.768 [0.755,0.781]	913

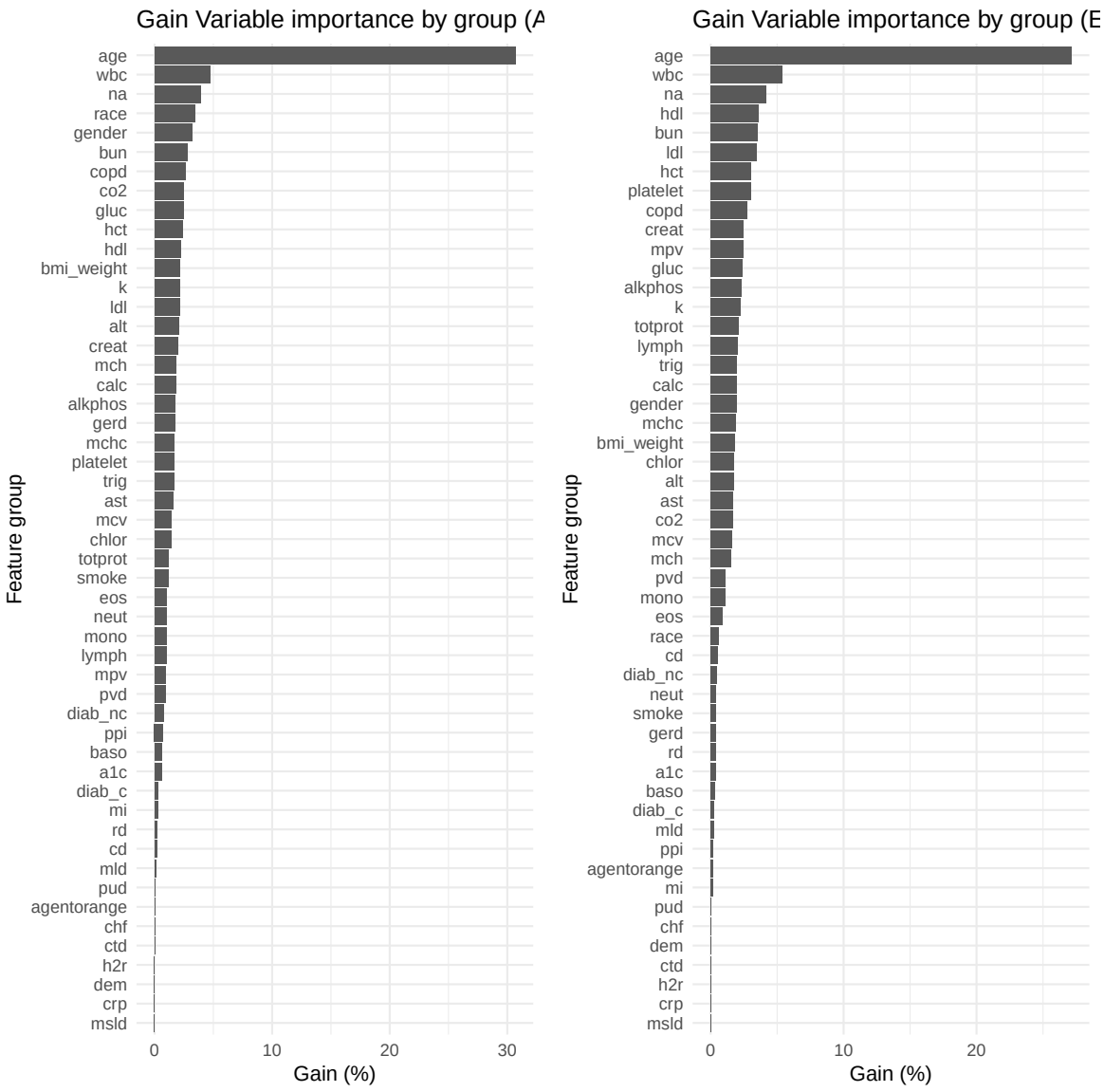
Stage	Test. AUC	Nb. cases
Any	0.741 [0.725,0.756]	764
I	0.768 [0.731,0.804]	145
II	0.757 [0.719,0.794]	133
III	0.754 [0.727,0.781]	228
IV	0.729 [0.705,0.753]	331
I+	0.743 [0.727,0.760]	635
II+	0.738 [0.720,0.756]	561
III+	0.735 [0.716,0.754]	499
IV+	0.729 [0.705,0.753]	331

7 Variable importance

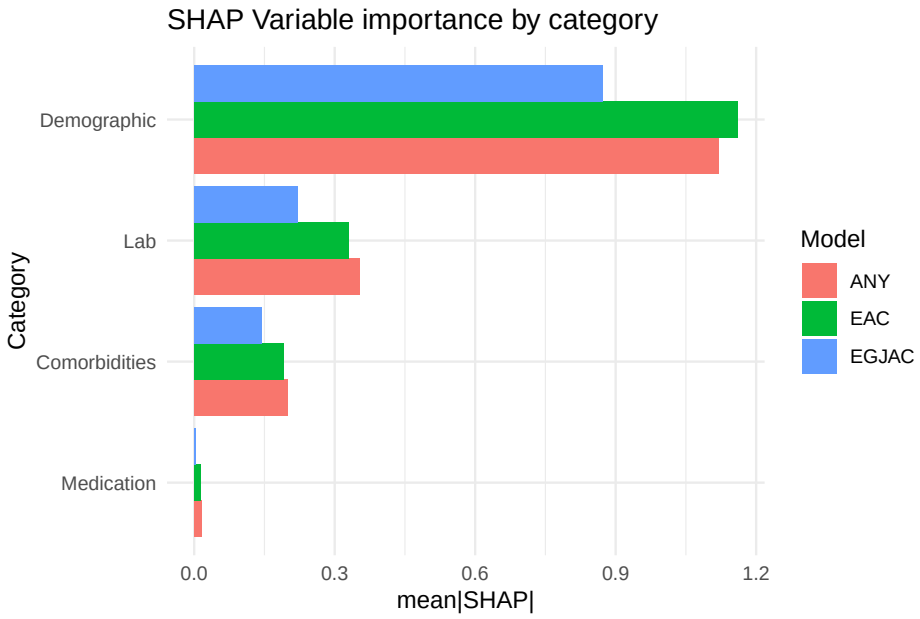
7.1 Gain VI

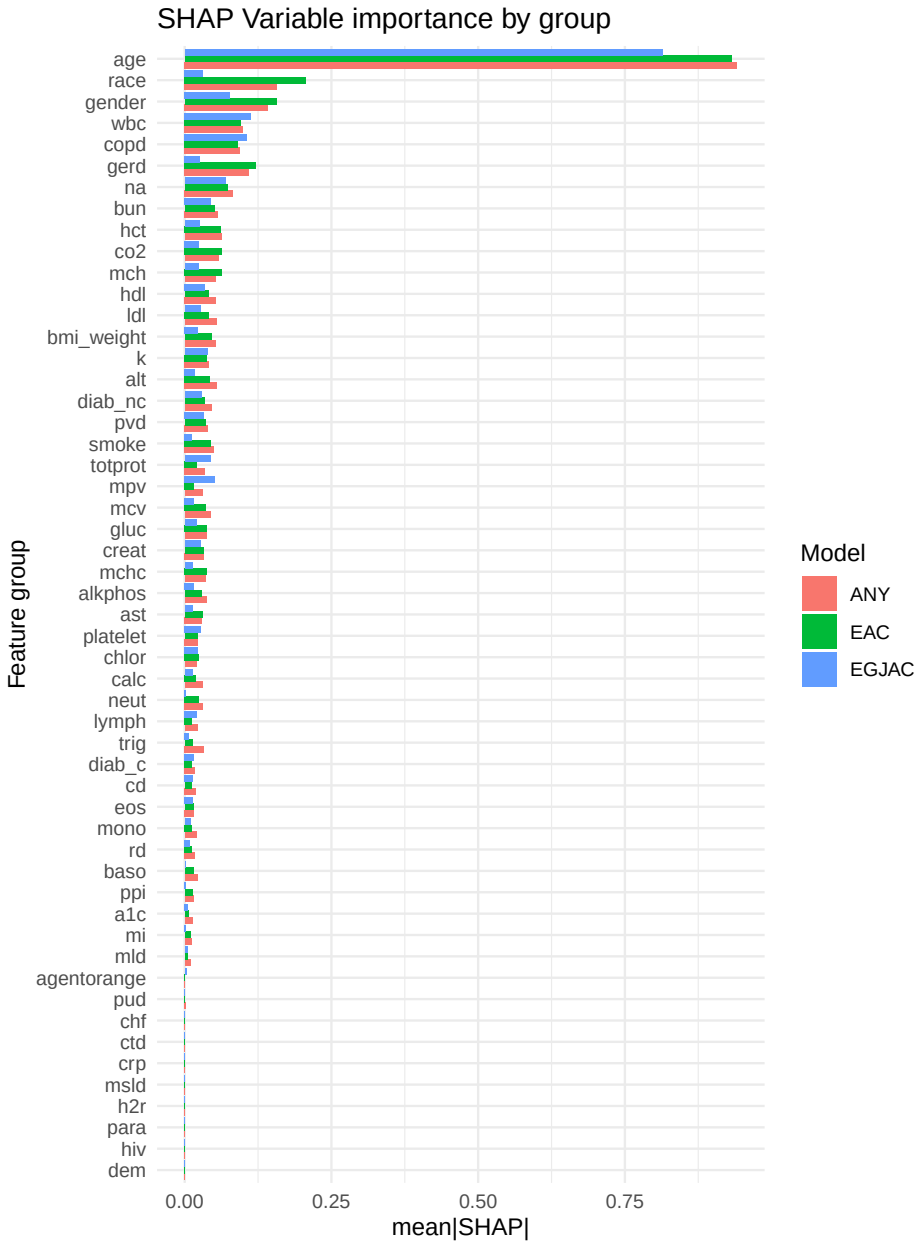


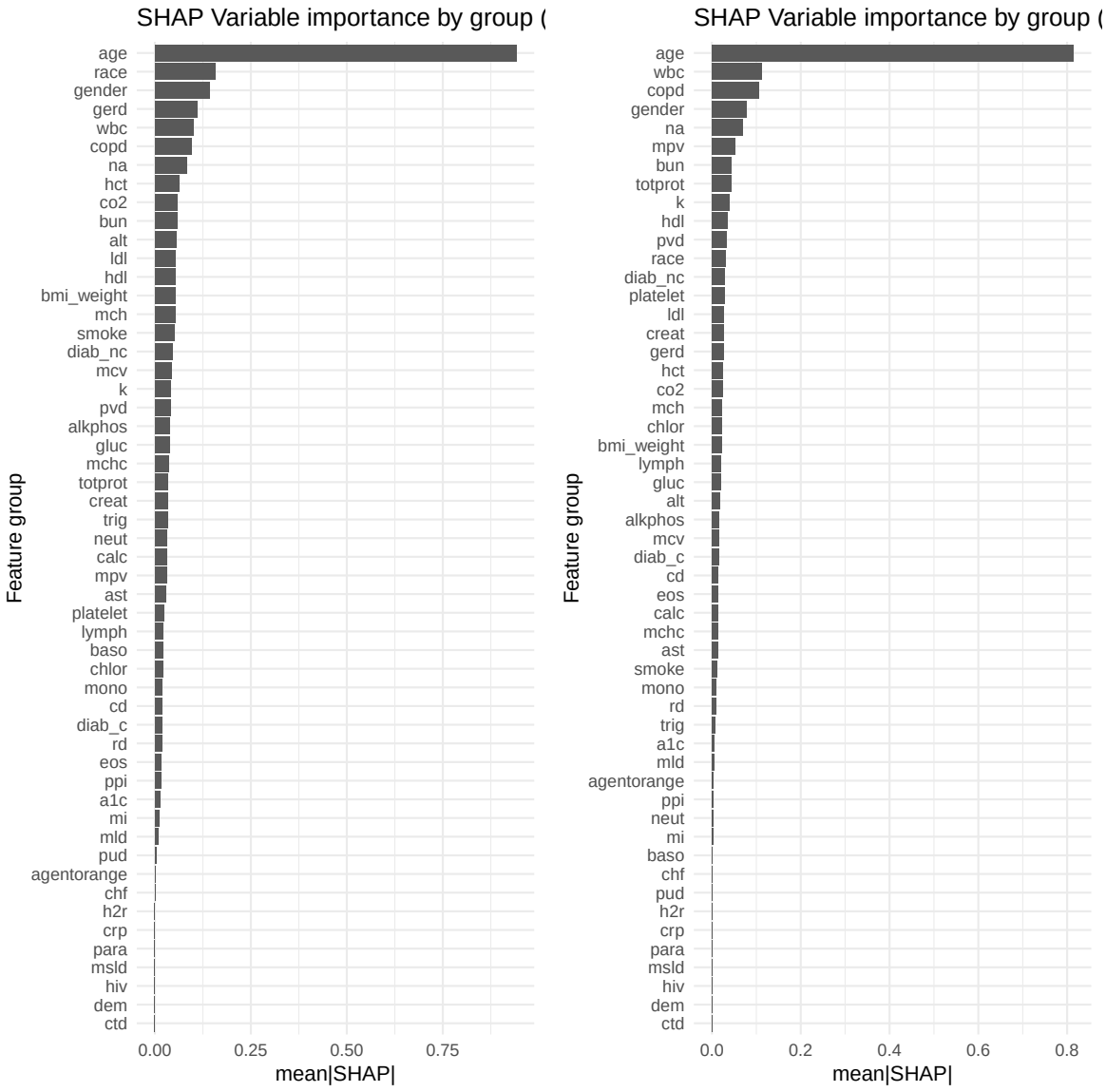

Gain Variable importance by group
age
wbc
na
race
gender
bun
copd
hdl
ldl
hct
gluc
co2
creat
k
platelet
bmi_weight
alt
alkphos
mchc
mch
calc
trig
chlor
ast
mcv
gerd
totprot
mpv
lymph
pvd
mono
smoke
eos
neut
diab_nc
ppi
baso
a1c
cd
diab_c
rd
mi
mld
agentorange
pud
chf
ctd
dem
crp
msld
h2r
Feature group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


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

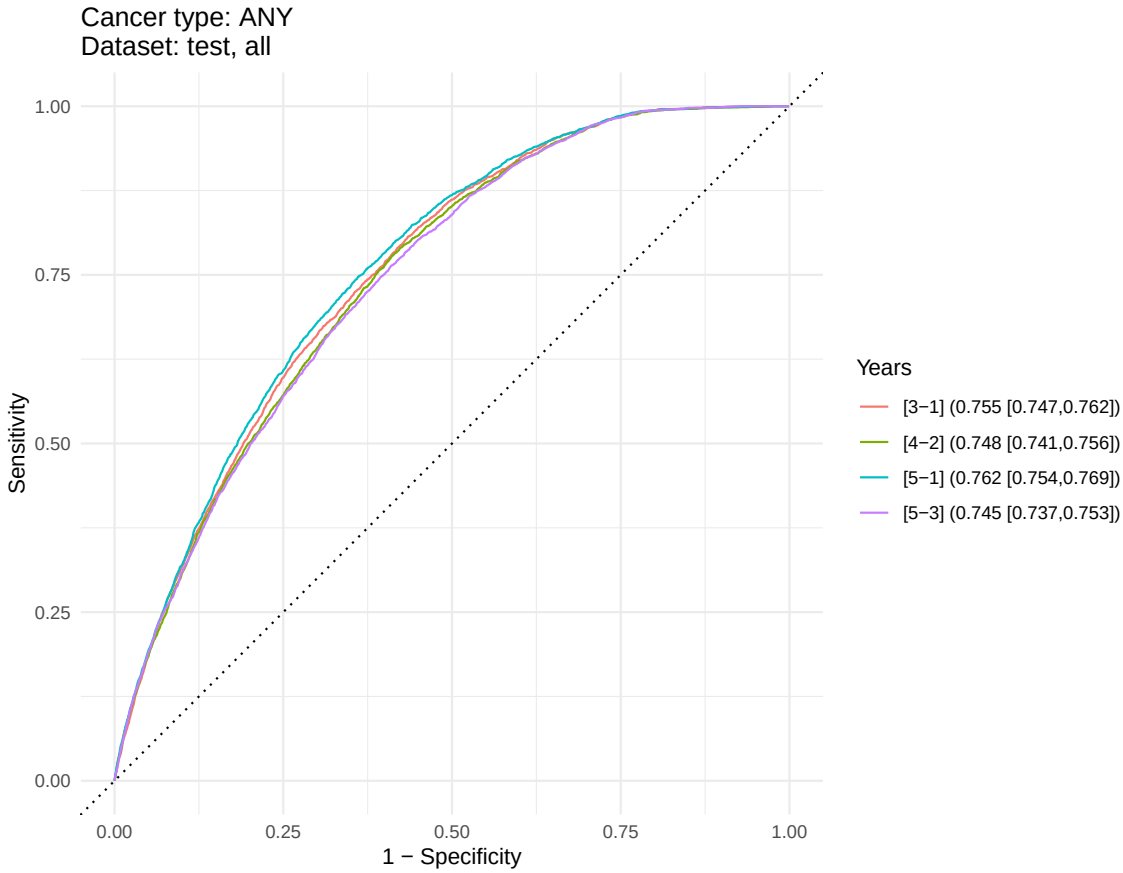
7.2 SHAP VI

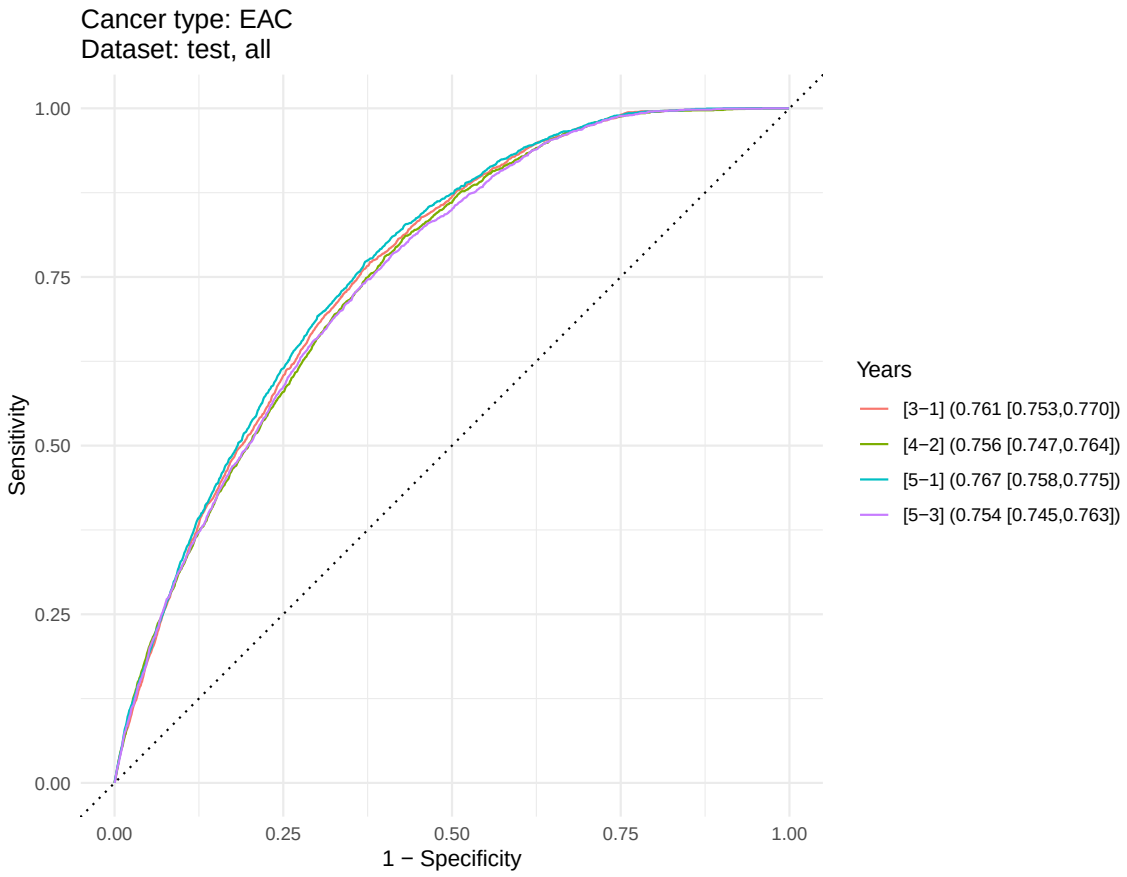

SHAP Variable importance by category
Demographic
Lab
Comorbidities
Medication
Category
0.0
0.3
0.6
0.9
1.2
mean|SHAP|
Model
ANY
EAC
EGJAC

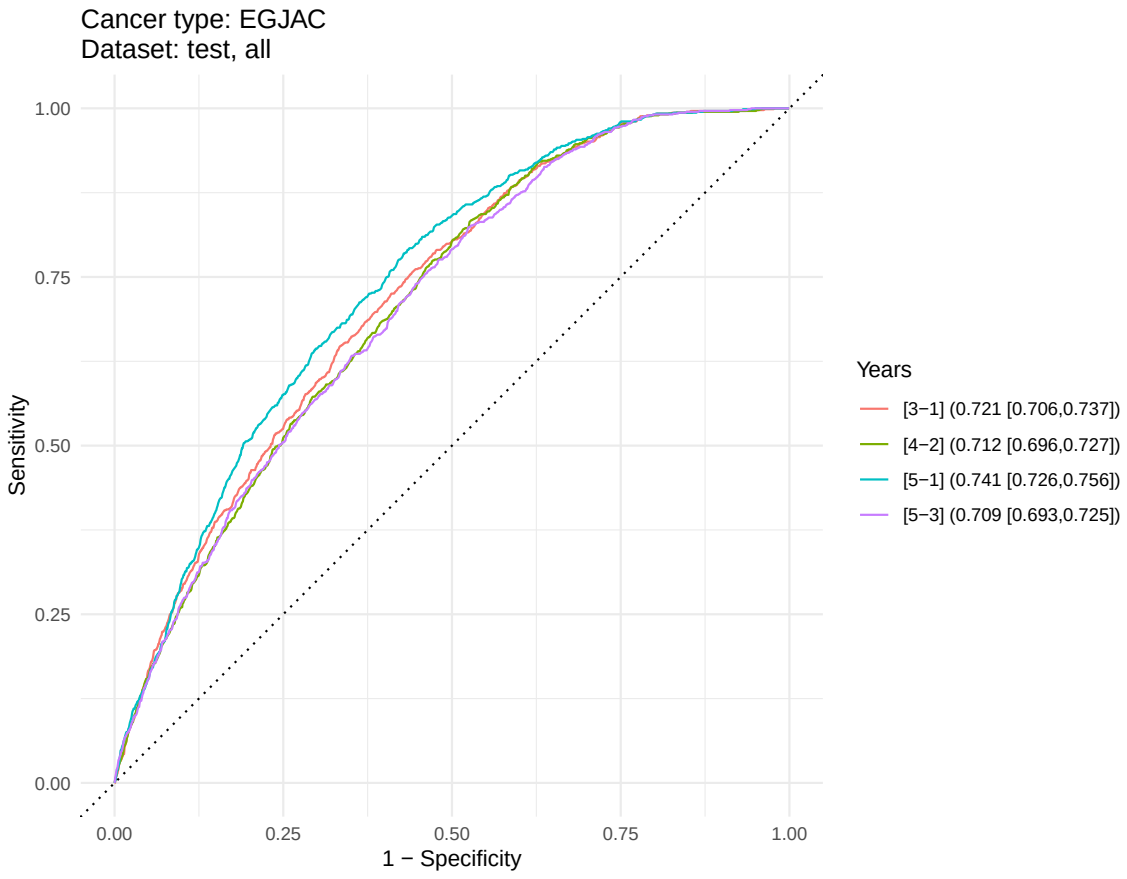

SHAP Variable importance by group
Feature group
age
race
gender
wbc
copd
gerd
na
bun
hct
co2
mch
hdl
ldl
bmi_weight
k
alt
diab_nc
pvd
smoke
totprot
mpv
mcv
gluc
creat
mchc
alkphos
ast
platelet
chlor
calc
neut
lymph
trig
diab_c
cd
eos
mono
rd
baso
ppi
a1c
mi
mld
agentorange
pud
chf
ctd
crp
msld
h2r
para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


SHAP Variable importance by group (
Feature group
age
race
gender
gerd
wbc
copd
na
hct
co2
bun
alt
ldl
hdl
bmi_weight
mch
smoke
diab_nc
mcv
k
pvd
alkphos
gluc
mchc
totprot
creat
trig
neut
calc
mpv
ast
platelet
lymph
baso
chlor
mono
cd
diab_c
rd
eos
ppi
a1c
mi
mld
pud
agentorange
chf
h2r
crp
para
msld
hiv
dem
ctd
0.00
0.25
0.50
0.75
mean|SHAP|
SHAP Variable importance by group (
Feature group
age
wbc
copd
gender
na
mpv
bun
totprot
k
hdl
pvd
race
diab_nc
platelet
ldl
creat
gerd
hct
co2
mch
chlor
bmi_weight
lymph
gluc
alt
alkphos
mcv
diab_c
cd
eos
calc
mchc
ast
smoke
mono
rd
trig
a1c
mld
agentorange
ppi
neut
mi
baso
chf
pud
h2r
crp
para
msld
hiv
dem
ctd
0.0
0.2
0.4
0.6
0.8
mean|SHAP|

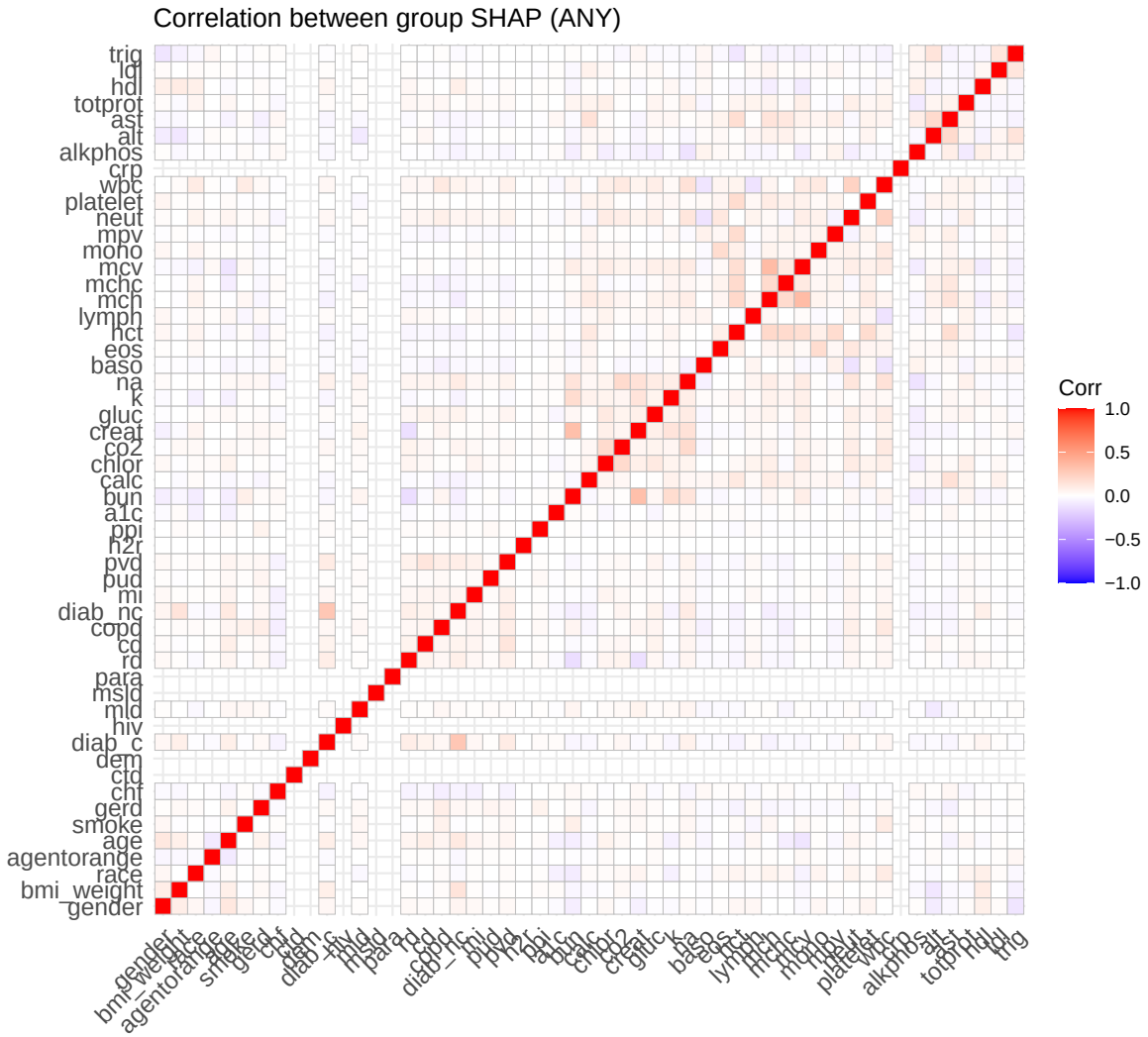
8 Years prior

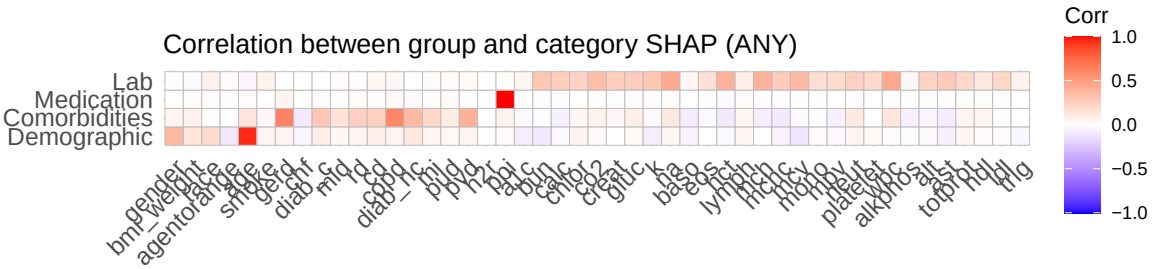


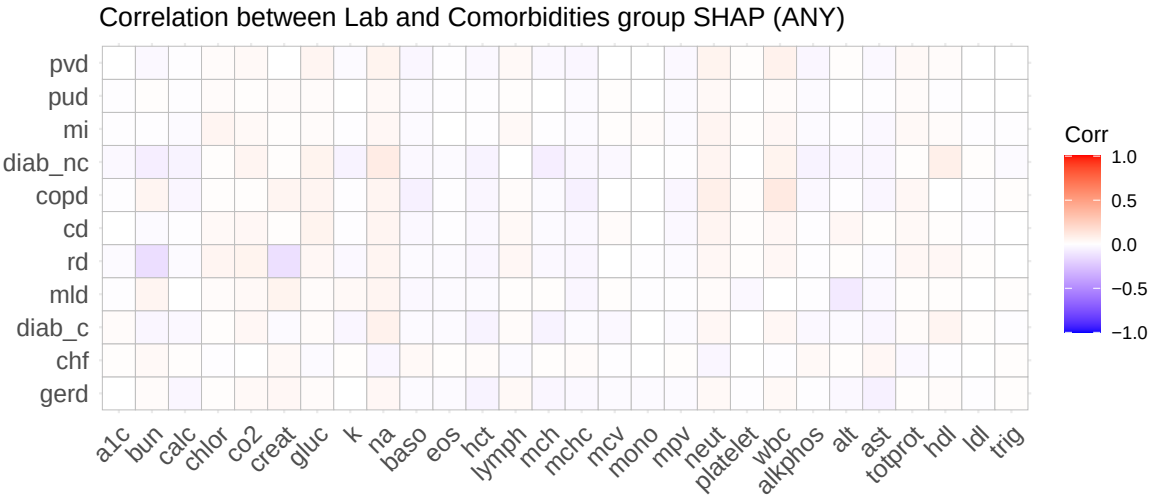

Cancer type: EAC
Dataset: test, all
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Years
[3−1] (0.761 [0.753,0.770])
[4−2] (0.756 [0.747,0.764])
[5−1] (0.767 [0.758,0.775])
[5−3] (0.754 [0.745,0.763])


Cancer type: EGJAC
Dataset: test, all
Sensitivity
1 − Specificity
0.00
0.25
0.50
0.75
1.00
Years
[3−1] (0.721 [0.706,0.737])
[4−2] (0.712 [0.696,0.727])
[5−1] (0.741 [0.726,0.756])
[5−3] (0.709 [0.693,0.725])

9 SHAP correlation and imputation model




Correlation between group and category SHAP (ANY)
Lab
Medication
Comorbidities
Demographic
gender
bmi_weight
race
agentorange
age
smoke
gerd
chf
diab_c
mld
rd
cd
copd
diab_nc
mi
pud
pvd
h2
ppi
a1c
bun
calc
chlor
co2
creat
gluc
k
na
baso
eos
hct
lymph
mch
mchc
mcv
mono
mpv
neut
platelet
wbc
alkphos
alt
ast
totprot
hdl
ldl
trig
Corr
1.0
0.5
0.0
−0.5
−1.0


Correlation between Lab and Comorbidities group SHAP (ANY)
pvd
pud
mi
diab_nc
copd
cd
rd
mld
diab_c
chf
gerd
a1c
bun
calc
chlor
co2
creat
gluc
k
na
baso
eos
hct
lymph
mch
mchc
mcv
mono
mpv
neut
platelet
wbc
alkphos
alt
ast
totprot
hdl
ldl
trig
Corr
1.0
0.5
0.0
−0.5
−1.0