We used the train-test split function from the model selection module implemented in the scikit-learn library to randomly split the dataset into train and test sets. We separate 20% of the data as the test set and the rest as for training a 5-fold cross validation. We conducted the following experiments:

Table 1. Cardiovascular disease classification using the conclusion section from discharge letters

	E11	E78	I10	I21	I25	I42	I48	I50	N18	Z95
Sensitivity	0.8783434	0.7739362	0.6993197	0.9670543	0.8131737	0.9165154	0.8400984	0.8817087	0.9378723	0.7639821
Specificity	0.8055556	0.5970149	0.6804348	0.8650307	0.7333333	0.8494624	0.9005236	0.8475177	0.6000000	0.6378738
Pos Pred Value	0.9931707	0.9700000	0.7776097	0.9784314	0.8761290	0.9863281	0.9472954	0.9492925	0.9927928	0.8623737
Neg Pred Value	0.1705882	0.1355932	0.5861423	0.8057143	0.6285714	0.4619883	0.7257384	0.6887608	0.1411765	0.4764268
F1	0.9322344	0.8609467	0.7363897	0.9727096	0.8434783	0.9501411	0.8904824	0.9142533	0.9645514	0.8102017

Table 2. Cardiovascular disease classification using complete discharge letters

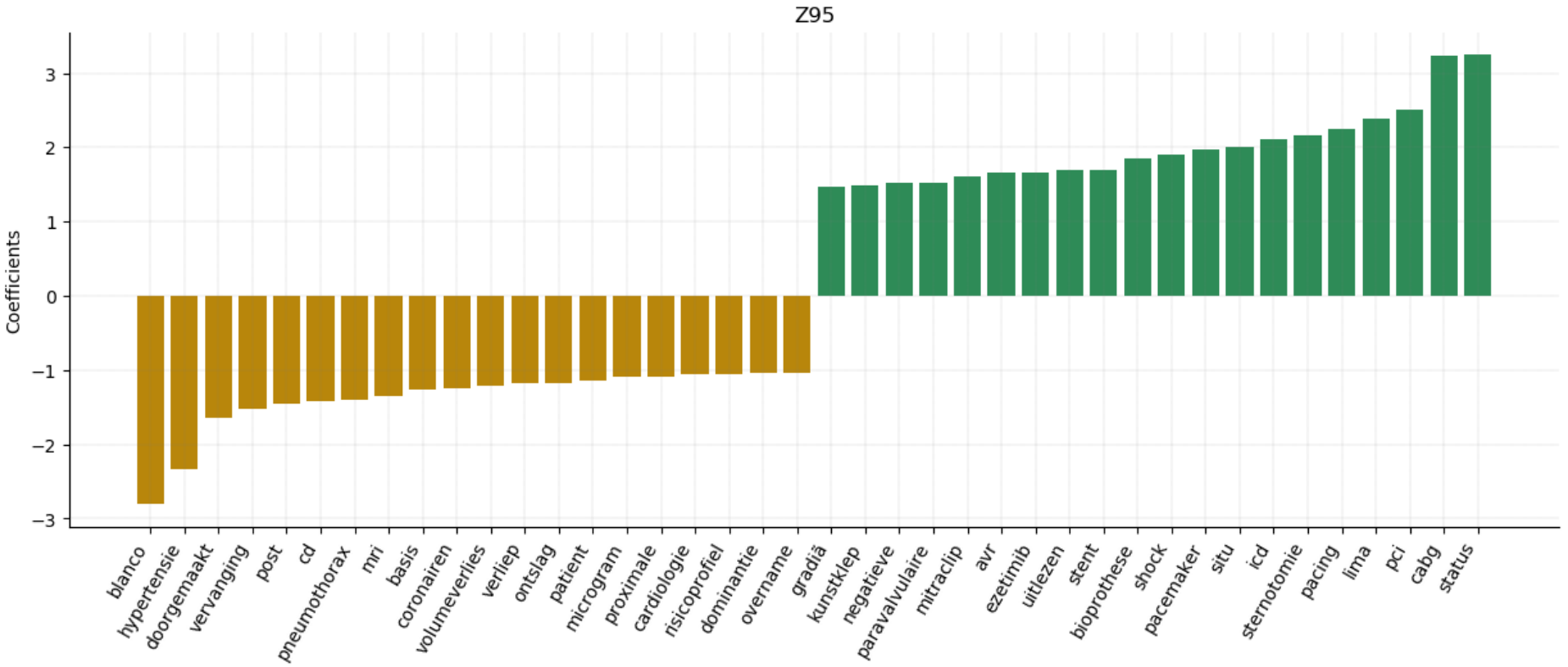
	E11	E78	I10	I21	I25	I42	I48	I50	N18	Z95
Sensitivity	0.9449193	0.7771883	0.7965616	0.9689018	0.8612500	0.9565217	0.8750000	0.9100346	0.9581882	0.8292398
Specificity	0.7887324	0.6718750	0.7887324	0.8614458	0.7822785	0.9124088	0.8997613	0.8201220	0.7872340	0.7558824
Pos Pred Value	0.9707317	0.9766667	0.8411498	0.9774510	0.8890323	0.9882812	0.9417476	0.9304245	0.9909910	0.8952020
Neg Pred Value	0.6588235	0.1457627	0.7340824	0.8171429	0.7357143	0.7309942	0.7953586	0.7752161	0.4352941	0.6377171
F1	0.9576516	0.8655835	0.8182487	0.9731576	0.8749206	0.9721422	0.9071476	0.9201166	0.9743136	0.8609593

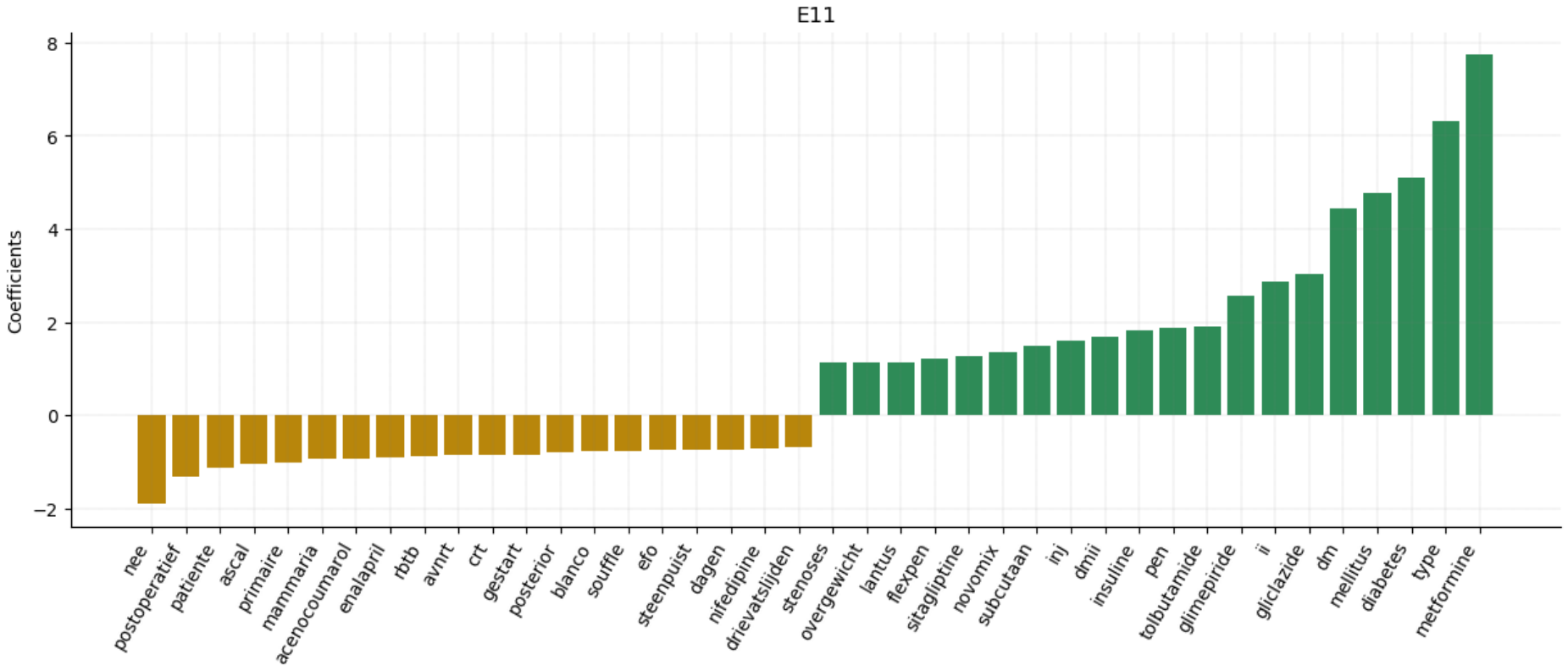
Table 3. Cardiovascular disease classification using discharge letters and variables age and sex

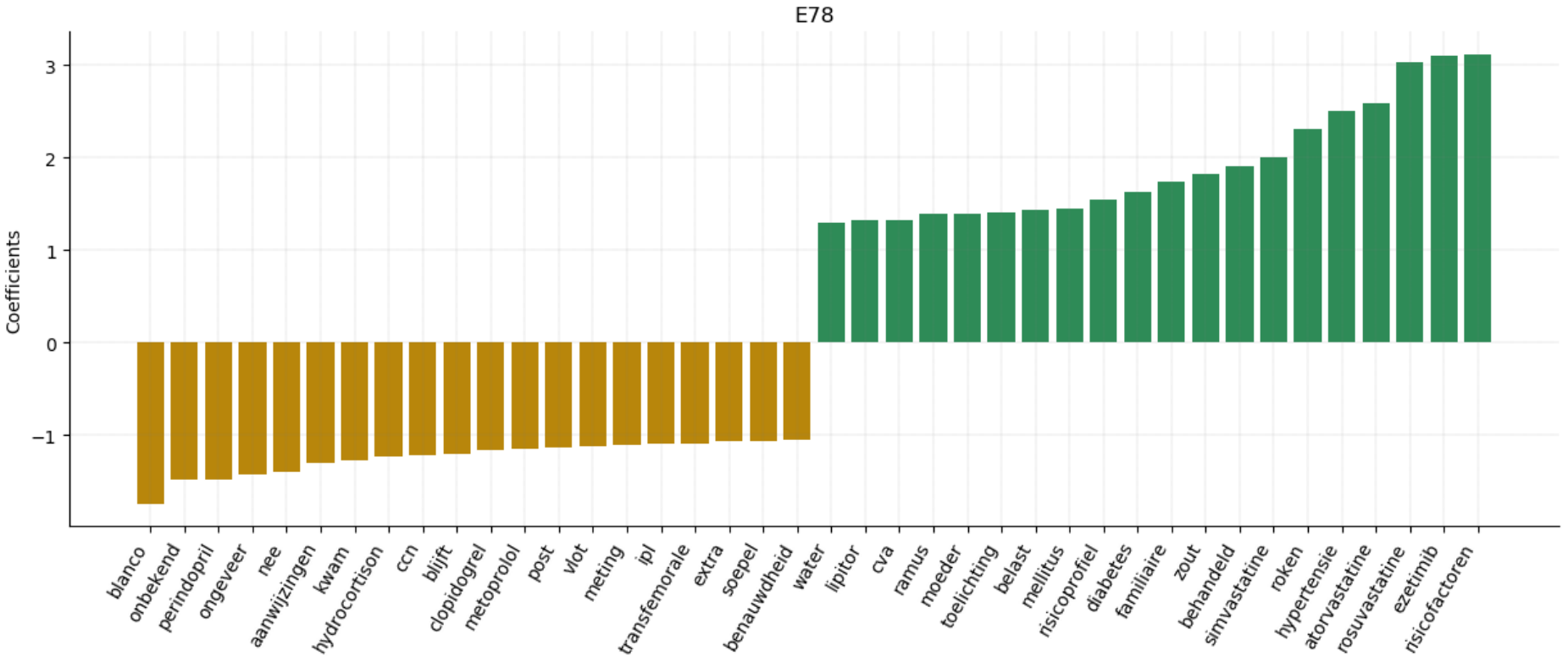
	E11	E78	I10	I21	I25	I42	I48	I50	N18	Z95
Sensitivity	0.9485224	0.7787611	0.7964861	0.9707887	0.8629442	0.9573460	0.8711340	0.9095128	0.9573542	0.8352941
Specificity	0.7945205	0.6923077	0.7714844	0.8630952	0.7665848	0.9000000	0.8926014	0.8078078	0.7826087	0.7623188
Pos Pred Value	0.9707317	0.9777778	0.8229955	0.9774510	0.8774194	0.9863281	0.9375867	0.9245283	0.9909910	0.8964646
Neg Pred Value	0.6823529	0.1525424	0.7397004	0.8285714	0.7428571	0.7368421	0.7890295	0.7752161	0.4235294	0.6526055
F1	0.9594986	0.8669951	0.8095238	0.9741085	0.8701216	0.9716210	0.9031396	0.9169591	0.9738822	0.8647990

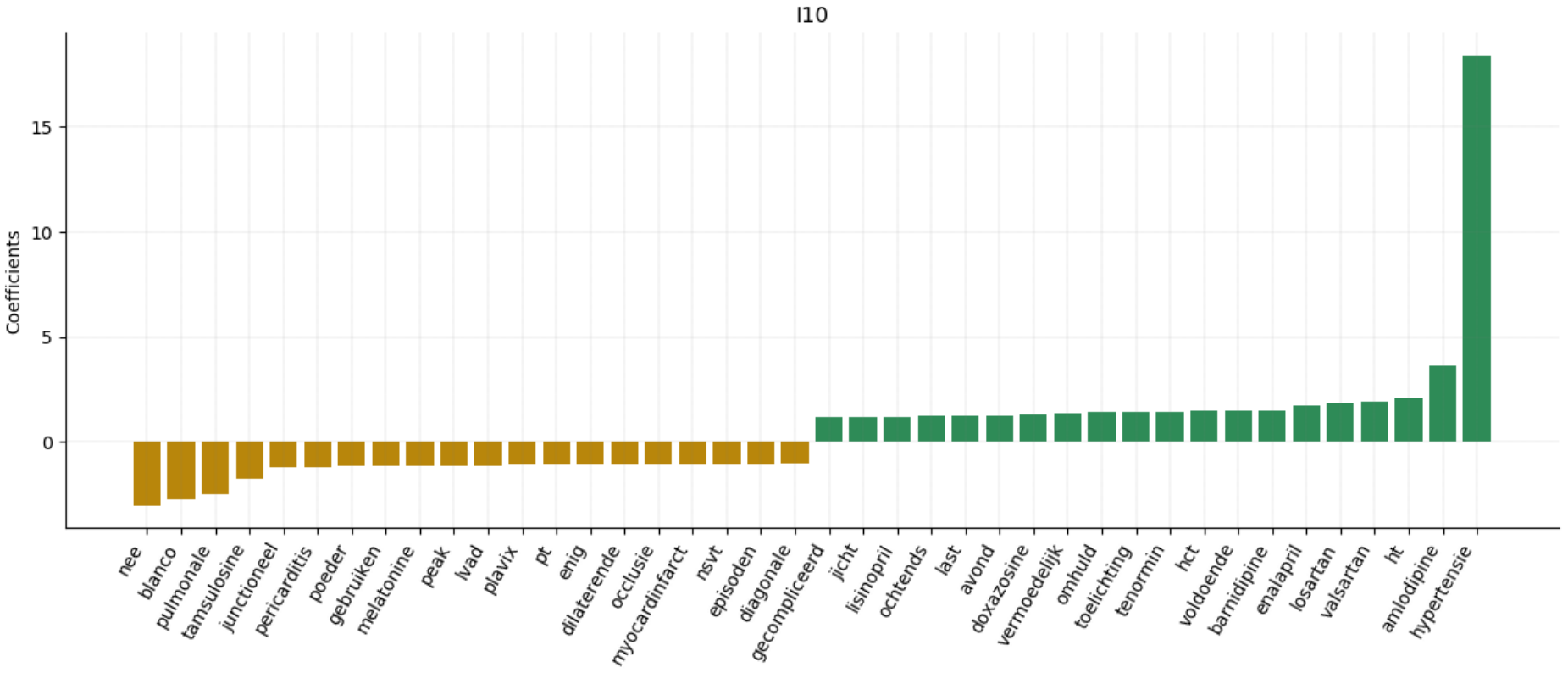
Table 4. Multi-label cardiovascular disease classification

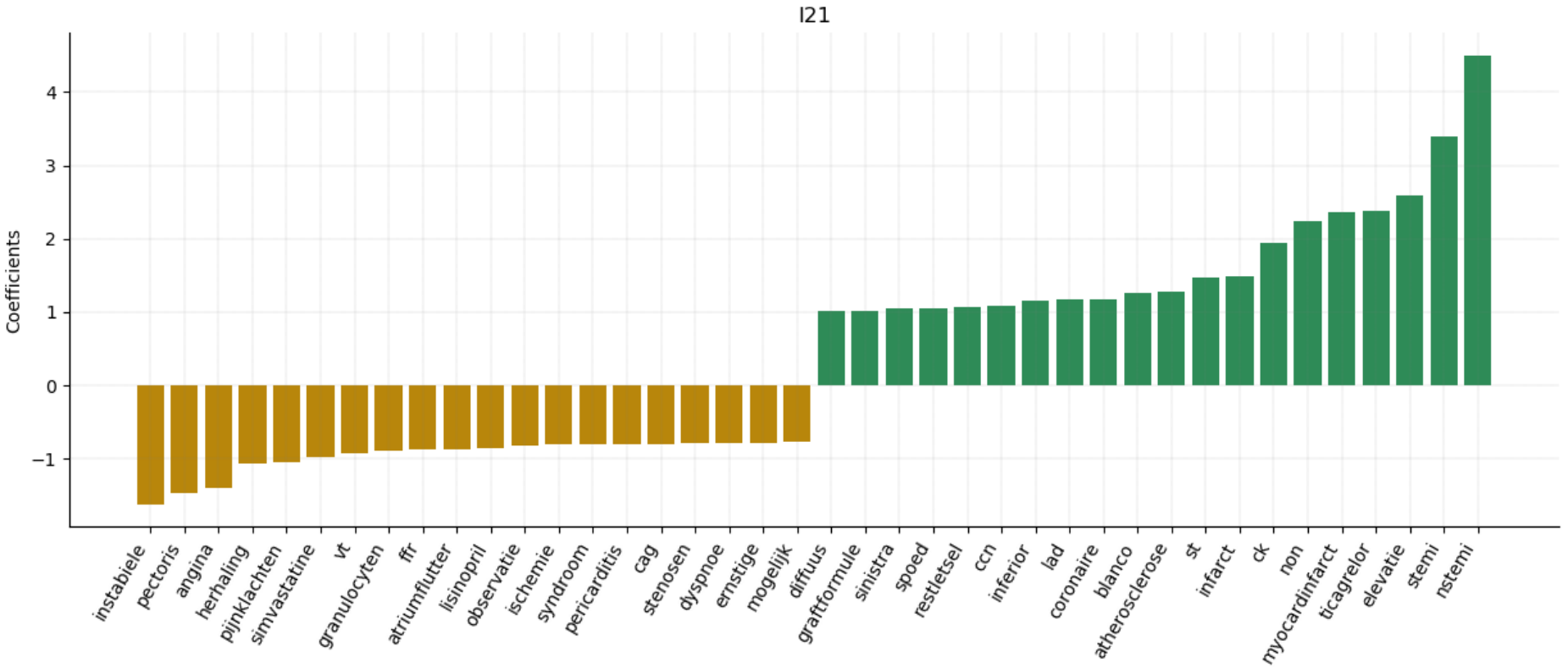
	Sensitivity	Specificity	Pos Pred Value	Neg Pred Value	F1	True cardinality	Prediction cardinality
Multi-label case	0.8798827	0.9813590	0.9274347	0.9551151	0.8893025	2.57	2.39

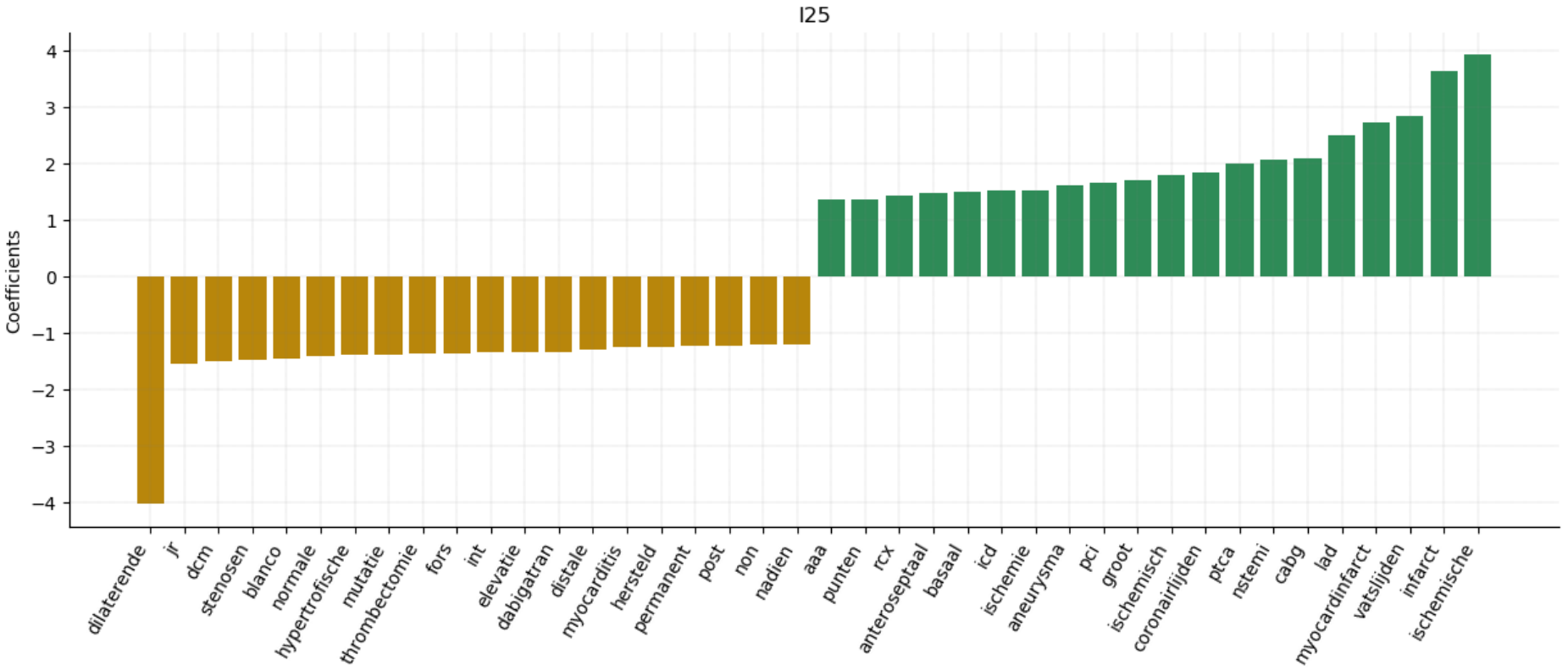

Z95
Coefficients
3
2
1
0
−1
−2
−3
blanco
hypertensie
doorgemaakt
vervanging
post
cd
pneumothorax
mri
basis
coronairen
volumeverlies
verliep
ontslag
patient
microgram
proximale
cardiologie
risicoprofiel
dominantie
overname
gradiä
kunstklep
negatieve
paravalvulaire
mitraclip
avr
ezetimib
uitlezen
stent
bioprothese
shock
pacemaker
situ
icd
sternotomie
pacing
lima
pci
cabg
status

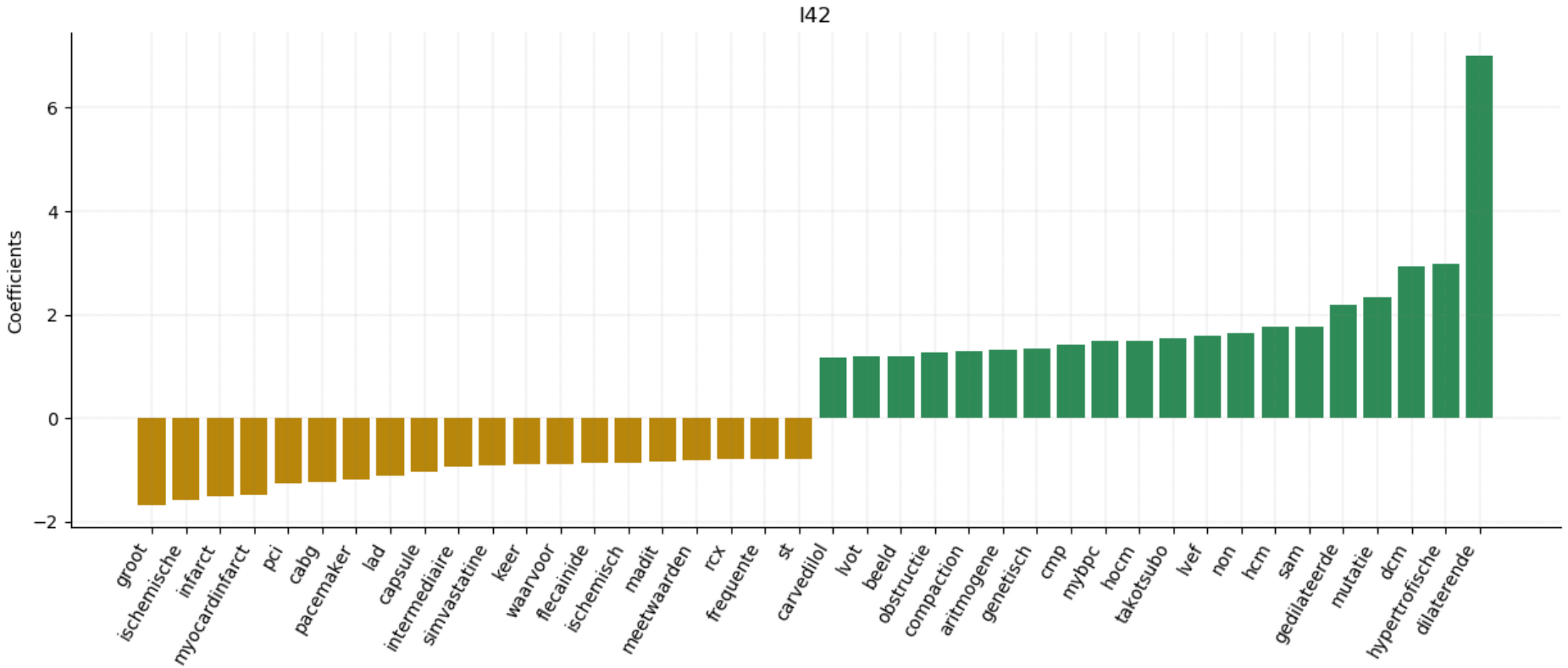

E11
Coefficients
8
6
4
2
0
−2
nee
postoperatief
patiente
ascal
primaire
mammaria
acenocoumarol
enalapril
rbtb
avnrt
crt
gestart
posterior
blanco
souffle
efo
steenpuist
dagen
nifedipine
drievatslijden
stenoses
overgewicht
lantus
flexpen
sitagliptine
novomix
subcutaan
inj
dmii
insuline
pen
tolbutamide
glimepiride
ii
gliclazide
dm
mellitus
diabetes
type
metformine

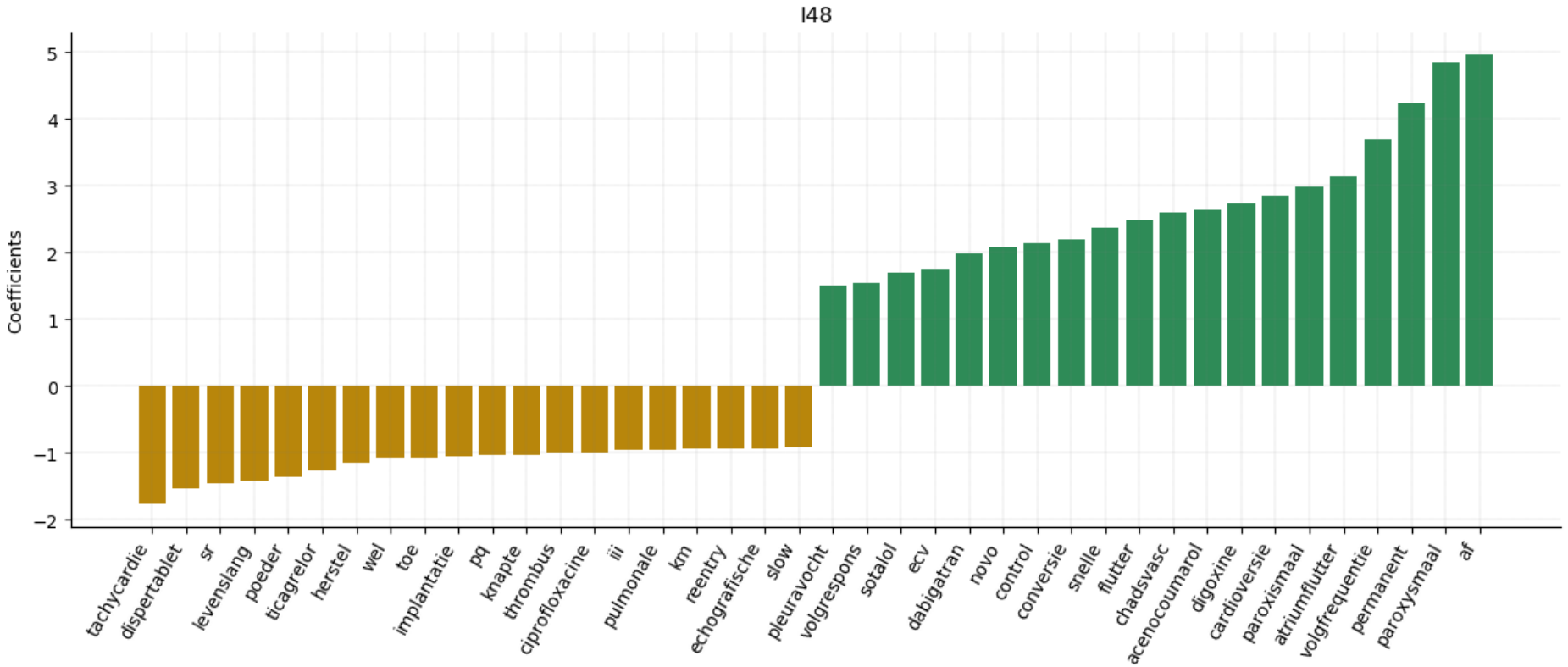

E78
Coefficients
3
2
1
0
−1
blanco
onbekend
perindopril
ongeveer
nee
aanwijzingen
kwam
hydrocortison
ccn
blijft
clopidogrel
metoprolol
post
vlot
meting
ipl
transfemorale
extra
soepel
benauwdheid
water
lipitor
cva
ramus
moeder
toelichting
belast
mellitus
risicoprofiel
diabetes
familiaire
zout
behandeld
simvastatine
roken
hypertensie
atorvastatine
rosuvastatine
ezetimib
risicofactoren

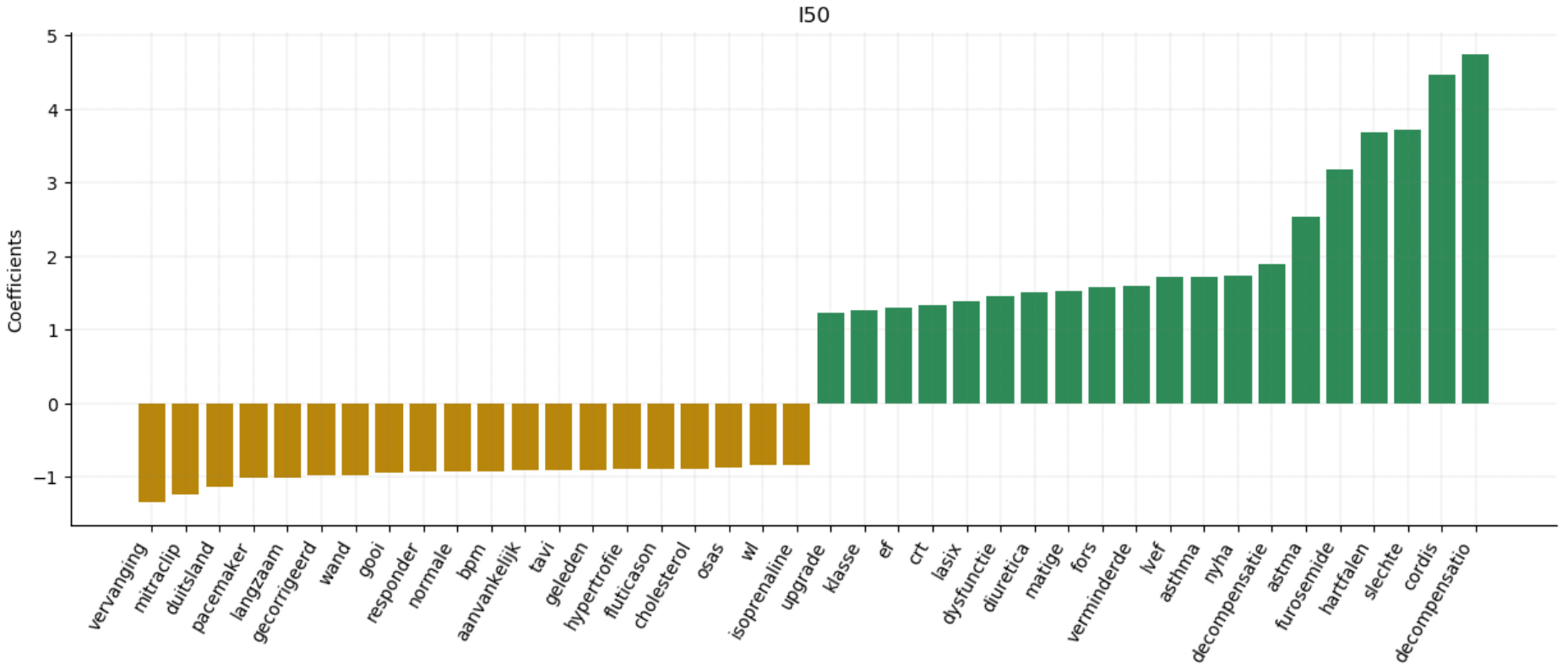

I10
Coefficients
15
10
5
0
nee
blanco
pulmonale
tamsulosine
junctioneel
pericarditis
poeder
gebruiken
melatonine
peak
lvad
plavix
pt
enig
dilaterende
occlusie
myocardinfarct
nsvt
episoden
diagonale
gecompliceerd
jicht
lisinopril
ochtends
last
avond
doxazosine
vermoedelijk
omhuld
toelichting
tenormin
hct
voldoende
barnidipine
enalapril
losartan
valsartan
ht
amlodipine
hypertensie

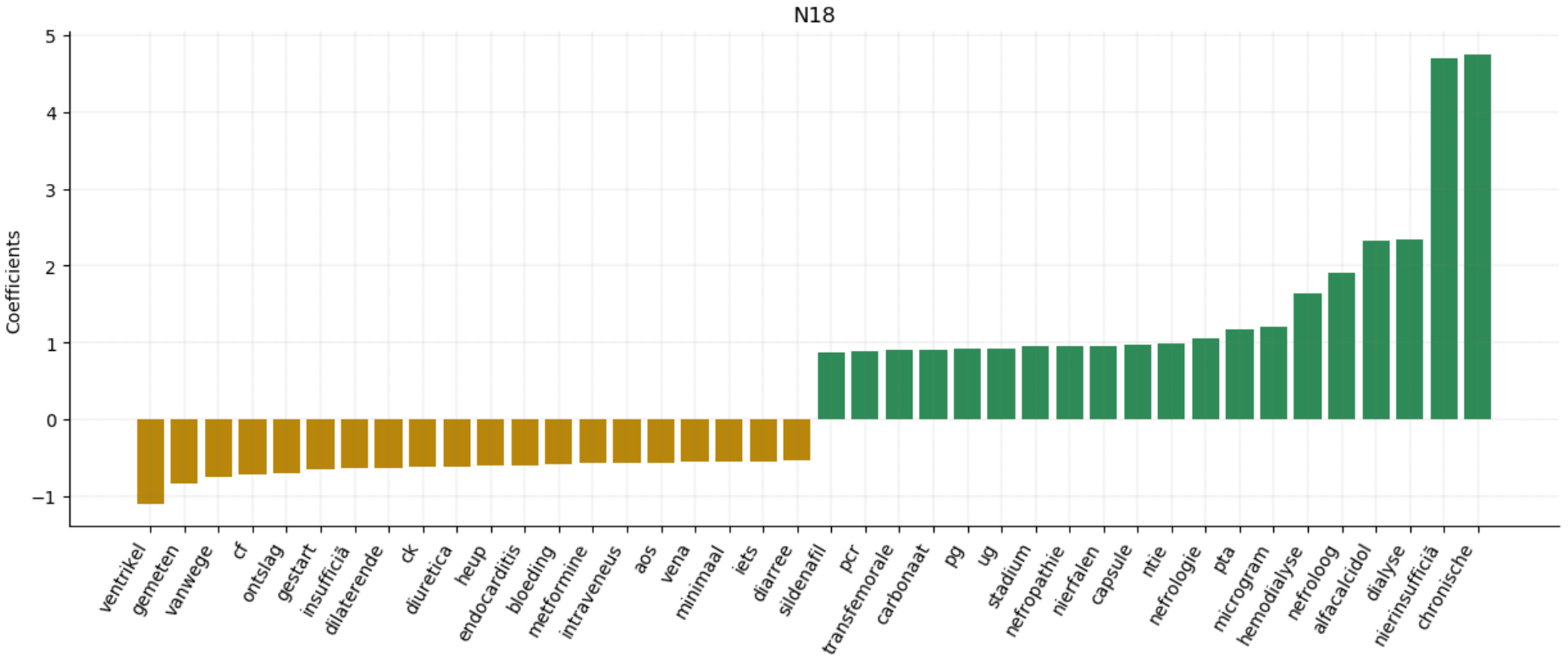

I21
Coefficients
4
3
2
1
0
−1
instabiele
pectoris
angina
herhaling
pijnklachten
simvastatine
vt
granulocyten
ffr
atriumflutter
lisinopril
observatie
ischemie
syndroom
pericarditis
cag
stenosen
dyspnoe
ernstige
mogelijk
diffuus
graftformule
sinistra
spoed
restletsel
ccn
inferior
lad
coronaire
blanco
atherosclerose
st
infarct
ck
non
myocardinfarct
ticagrelor
elevatie
stemi
nstemi


I25
Coefficients
4
3
2
1
0
−1
−2
−3
−4
dilaterende
jr
dcm
stenosen
blanco
normale
hypertrofische
mutatie
thrombectomie
fors
int
elevatie
dabigatran
distale
myocarditis
hersteld
permanent
post
non
nadien
aaa
punten
rcx
anteroseptaal
basaal
icd
ischemie
aneurysma
pci
groot
ischemisch
coronairlijden
ptca
nstemi
cabg
lad
myocardinfarct
vatslijden
infarct
ischemische


I42
Coefficients
6
4
2
0
−2
groot
ischemische
infarct
myocardinfarct
pci
cabg
pacemaker
lad
capsule
intermediaire
simvastatine
keer
waarvoor
flecainide
ischemisch
madit
meetwaarden
rcx
frequente
st
carvedilol
lvot
beeld
obstructie
compaction
aritmogene
genetisch
cmp
mybpc
hocm
takotsubo
lvef
non
hcm
sam
gedilateerde
mutatie
dcm
hypertrofische
dilaterende


I48
Coefficients
5
4
3
2
1
0
−1
−2
tachycardie
dispertablet
sr
levenslang
poeder
ticagrelor
herstel
wel
toe
implantatie
pq
knapte
thrombus
ciprofloxacine
iii
pulmonale
km
reentry
echografische
slow
pleuravocht
volgrespons
sotalol
ecv
dabigatran
novo
control
conversie
snelle
flutter
chadsvasc
acenocoumarol
digoxine
cardioversie
paroxismaal
atriumflutter
volgfrequentie
permanent
paroxysmaal
af


I50
Coefficients
5
4
3
2
1
0
−1
vervanging
nitraclip
duitsland
pacemaker
langzaam
gecorrigeerd
wand
gooi
responder
normale
bpm
aanvankelijk
tavi
geleden
hypertrofie
fluticason
cholesterol
osas
wl
isoprenaline
upgrade
klasse
ef
crt
lasix
dysfunctie
diuretica
matige
fors
verminderde
lvef
asthma
nyha
decompensatie
astma
furosemide
hartfalen
slechte
cordis
decompensatio


N18
Coefficients
5
4
3
2
1
0
−1
ventrikel
gemeten
vanwege
cr
ontslag
gestart
insufficiä
dilaterende
ck
diuretica
heup
endocarditis
bloeding
metformine
intraveneus
aos
vena
minimaal
iets
diarree
sildenafil
pcr
transfemorale
carbonaat
pg
ug
stadium
nefropathie
nierfalen
capsule
ntie
nefrologie
pta
microgram
hemodialyse
nefroloog
alfacalcidol
dialyse
nierinsufficiä
chronische