Preliminary Result of Paper 2; Aggregation Method B

June 27, 2019

## Table 1

Demographics of the ever had relapse vs never had relapse.

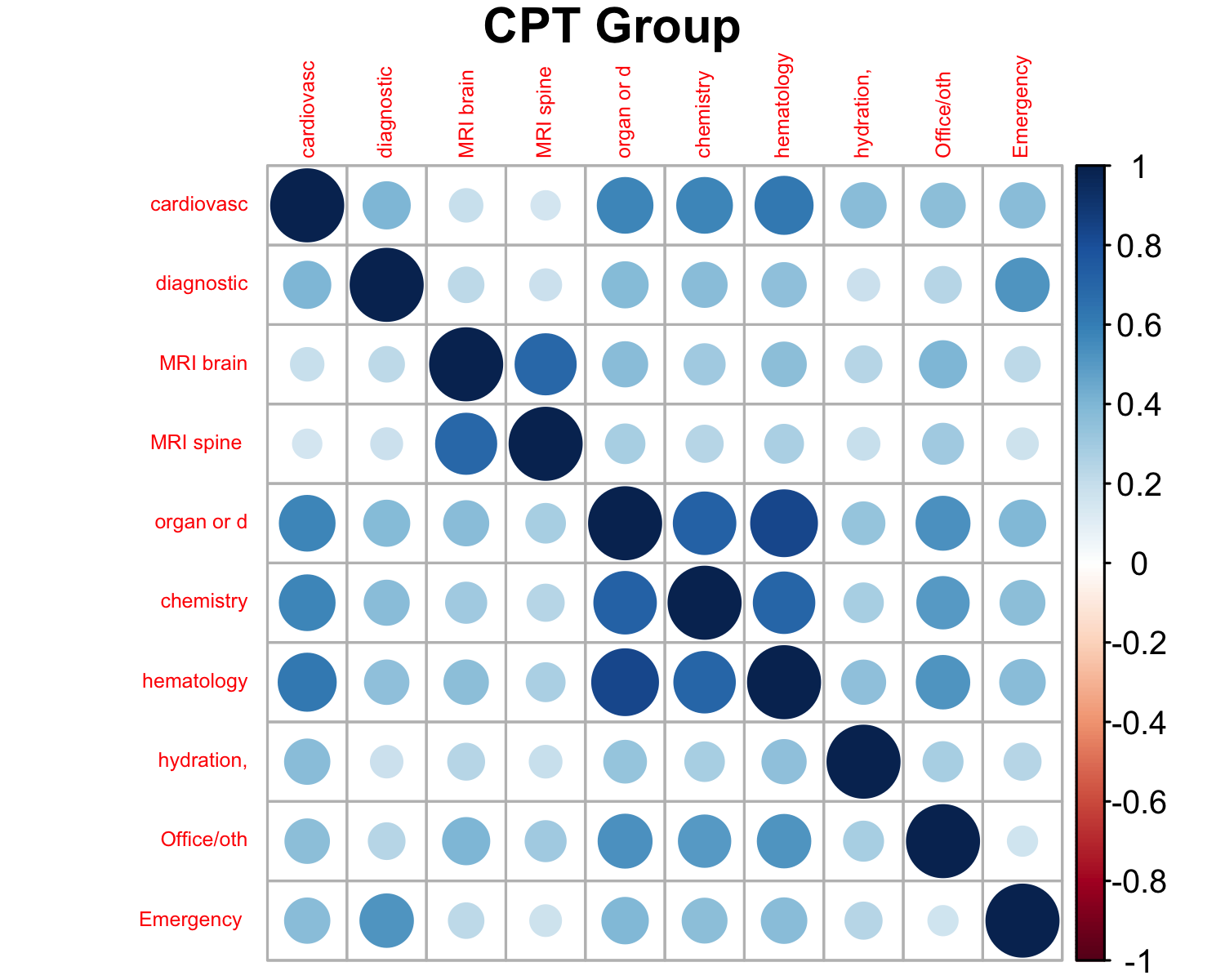
|  |  |  |  |
| --- | --- | --- | --- |
|  | Ever had relapse | Never had relapse | p-value for heter |
| Total number of patients | 1672 | 703 | NA |
| Num belongs to EHR | 1047 | 508 | NA |
| Sex (% female) | 73.8 | 71.4 | 0.233 |
| Race (% white) | 92.5 | 93 | 0.69 |
| Median (IQR) age at first code | 33.4 (15.5) | 39 (15.3) | 4.3e-14 |
| Median (IQR) age at first symptom onset | 32.9 (15) | 33 (13.9) | 0.846 |
| Median (IQR) age at first date | 33.6 (16) | 35.9 (15.2) | 5.8e-05 |
| Median (IQR) duration of follow-up | 12.7 (12.5) | 16.9 (15.6) | 7.1e-24 |
| Median (IQR) number of treatments | 2 (2.2) | 0 (2) | 0.0e+00 |
| % receiving treatments | 95.5 | 44.1 | 7.4e-182 |
| Average Annualized relapse rate 2000-2016 | 0.222 (0.005) | 0 (0) | NA |

## Data Manipulation

We split each patient’s follow-up period into 3-month time windows. We then aggregate the ICD/CPT/CUI count for each window start time in the next 6-month. The overall case rate is 0.115.

## Correlations

## Perfect correlation between CUI.C1304680 CUI.chip0104   
## CUI.chip0104 deleted.



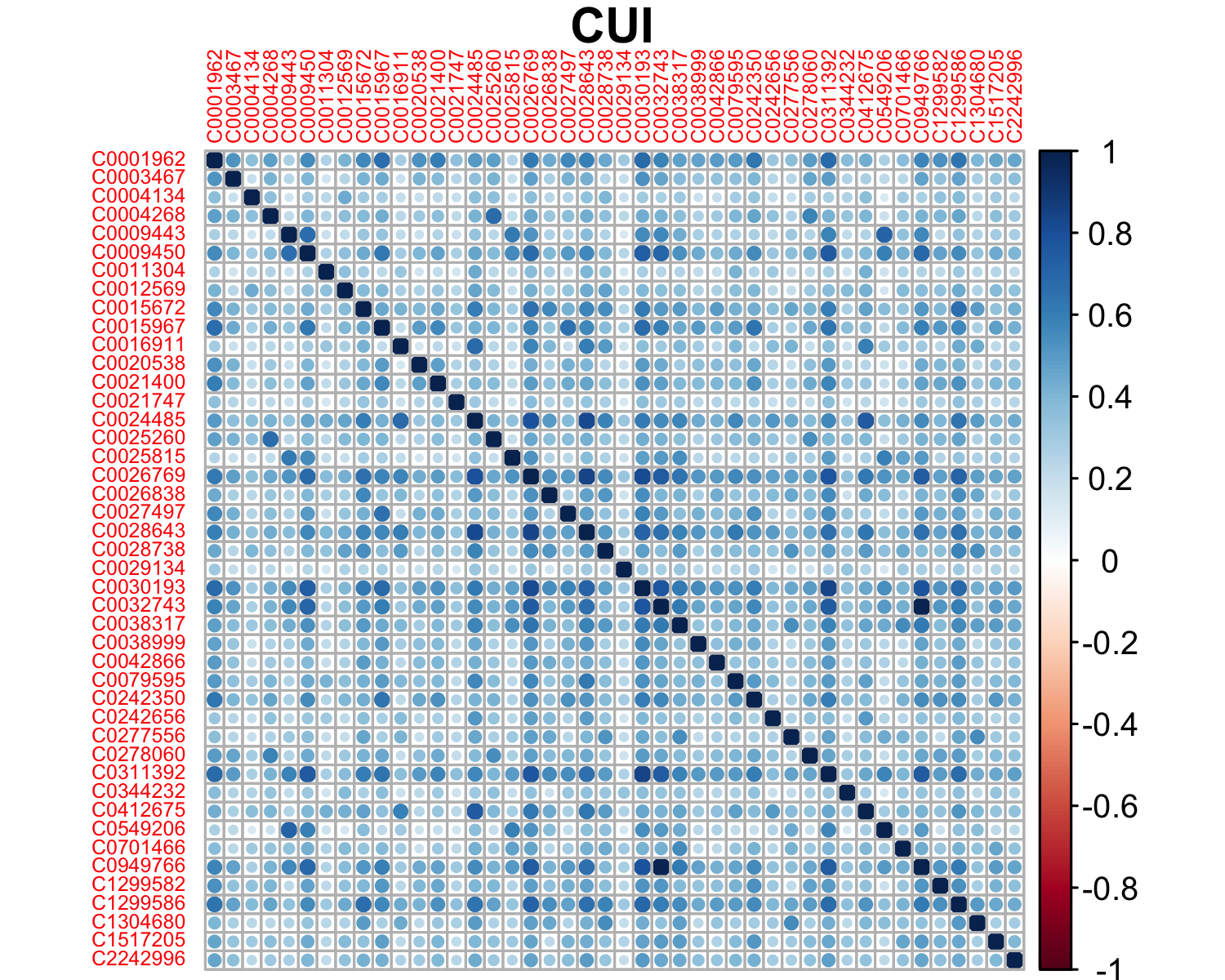
## Among CPT Groups:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.1516 0.2475 0.3608 0.3744 0.4075 0.8315

## CPT Group vs others:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.1103 0.2578 0.3355 0.3552 0.4408 0.7233

|  |  |  |
| --- | --- | --- |
| CUI | CUI | Corr |
| C0032743 | C0949766 | 0.982 |



## Among CUIs:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.1336 0.2787 0.3709 0.3851 0.4700 0.9821

## CUI vs others:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.1103 0.2658 0.3459 0.3640 0.4491 0.8720

## Only one phecode left: 335\_ ; no corrplot returned.

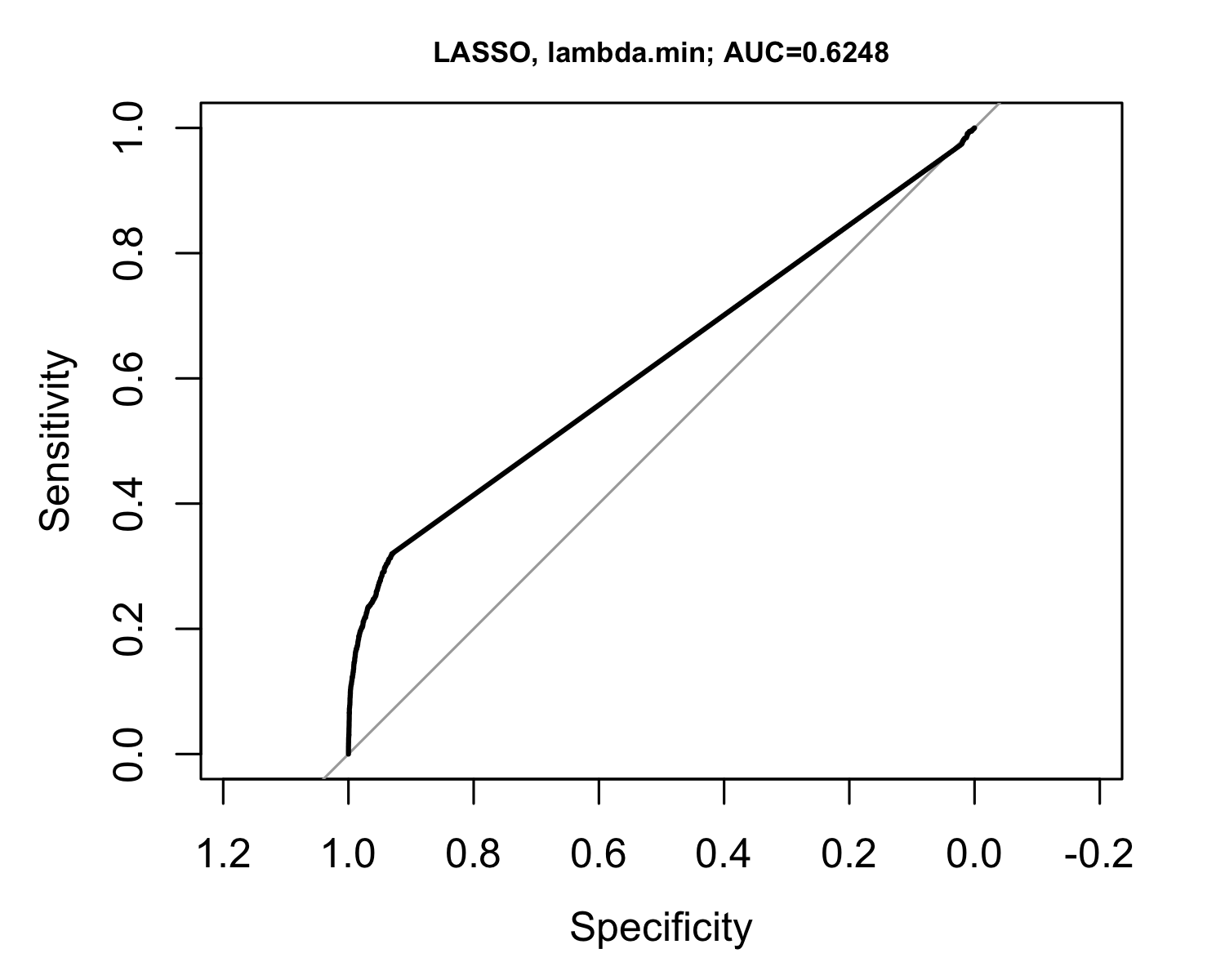
## Among PheCodes:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
##

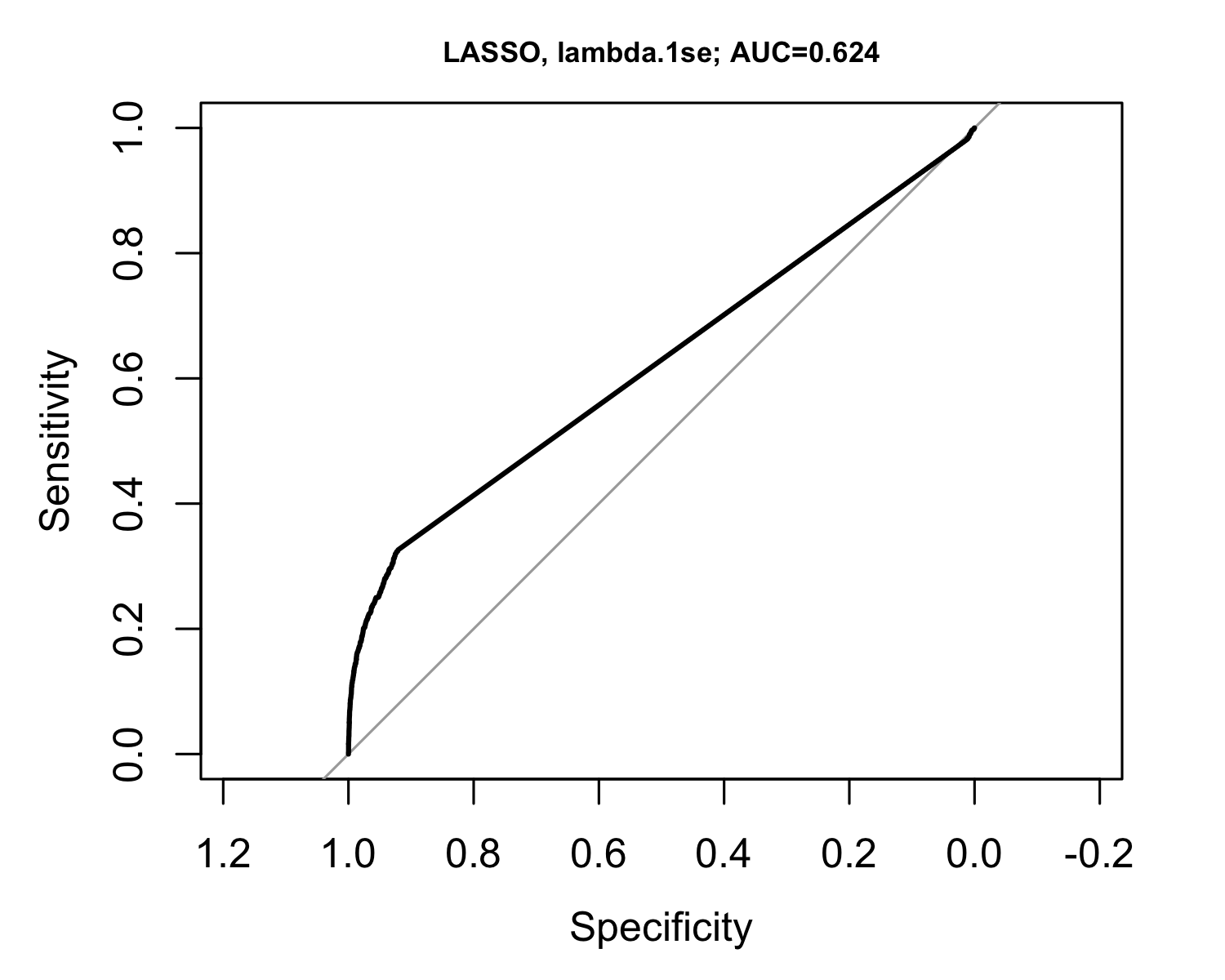
## PheCode vs others:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.3007 0.3985 0.4768 0.4998 0.6076 0.8720

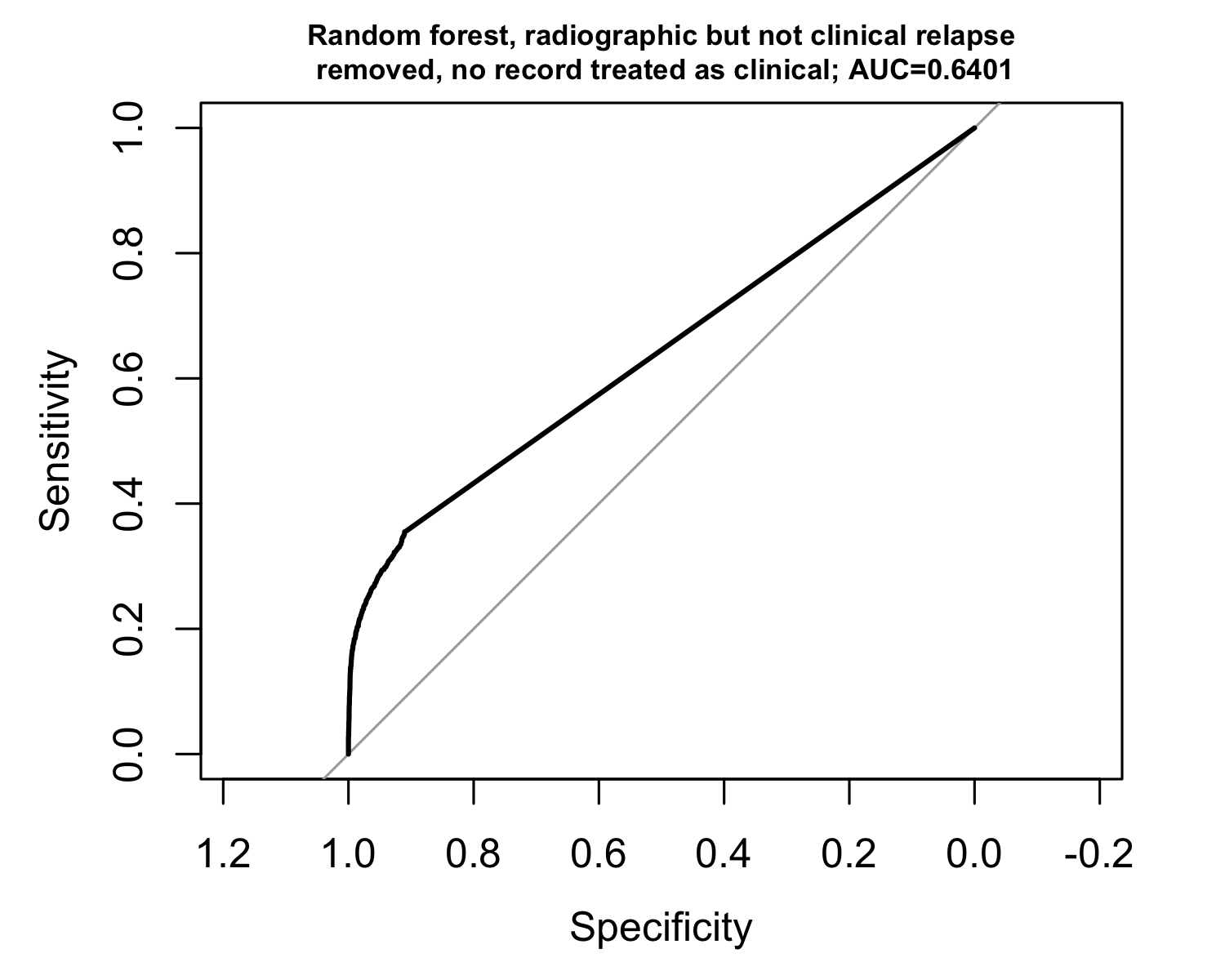
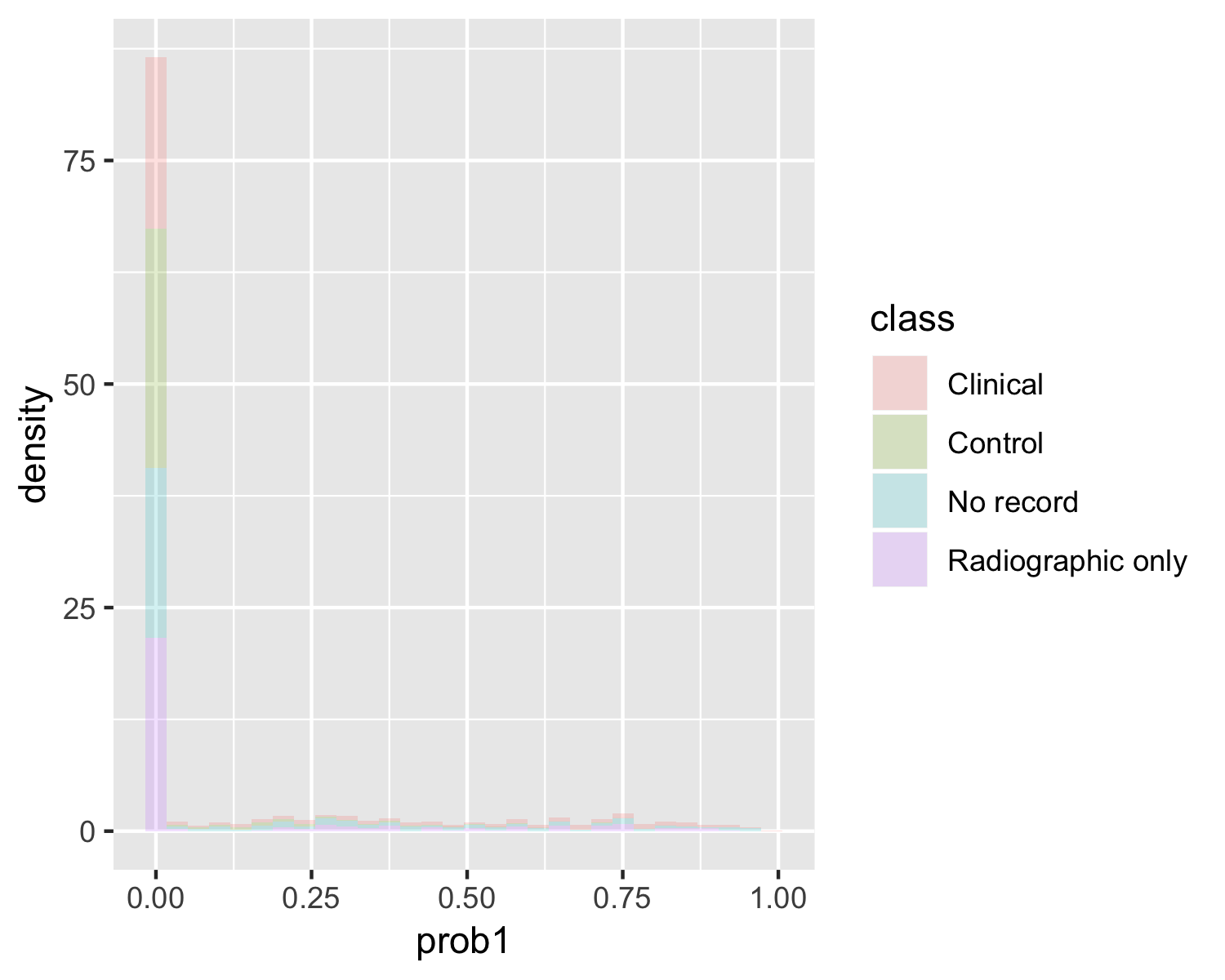
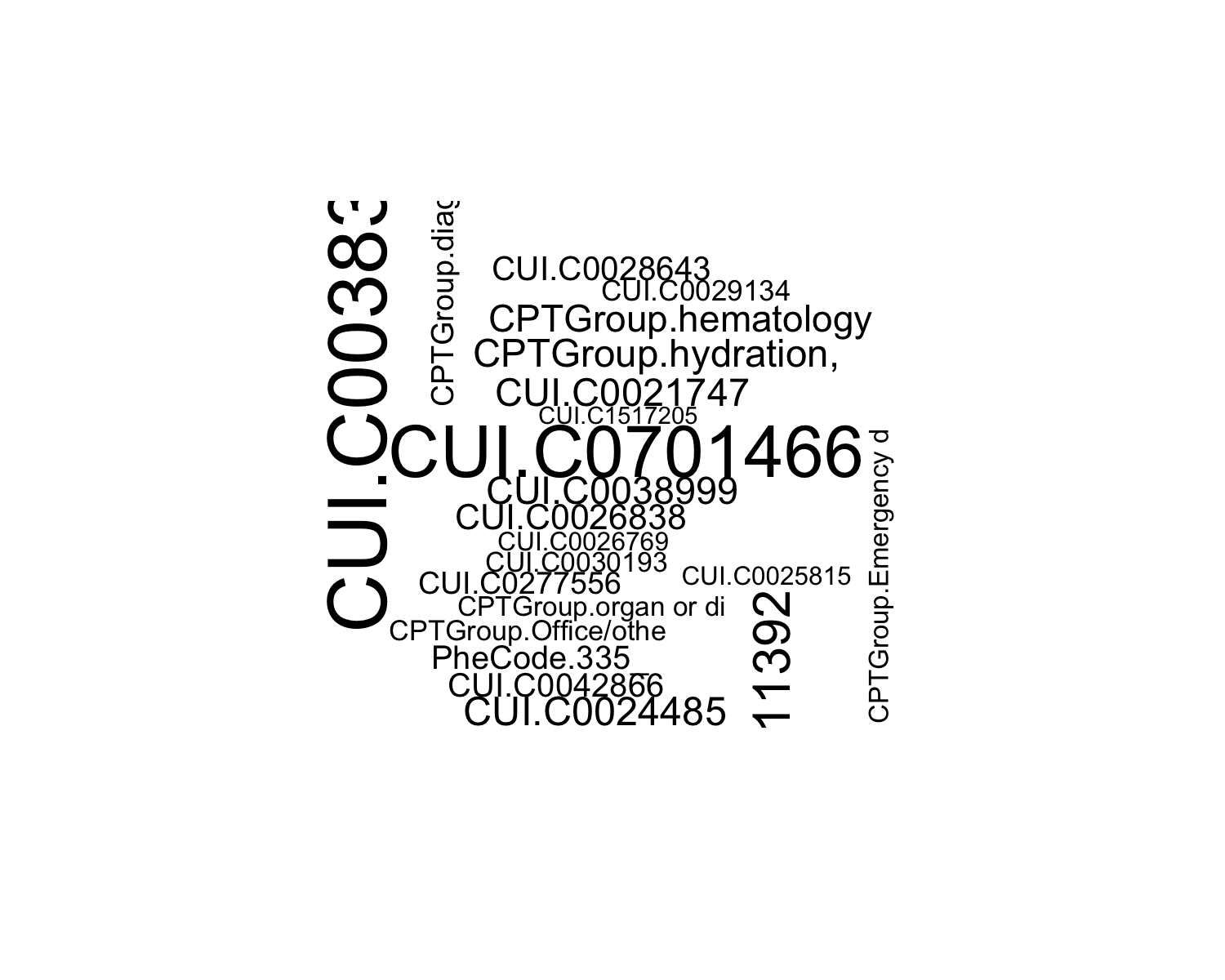
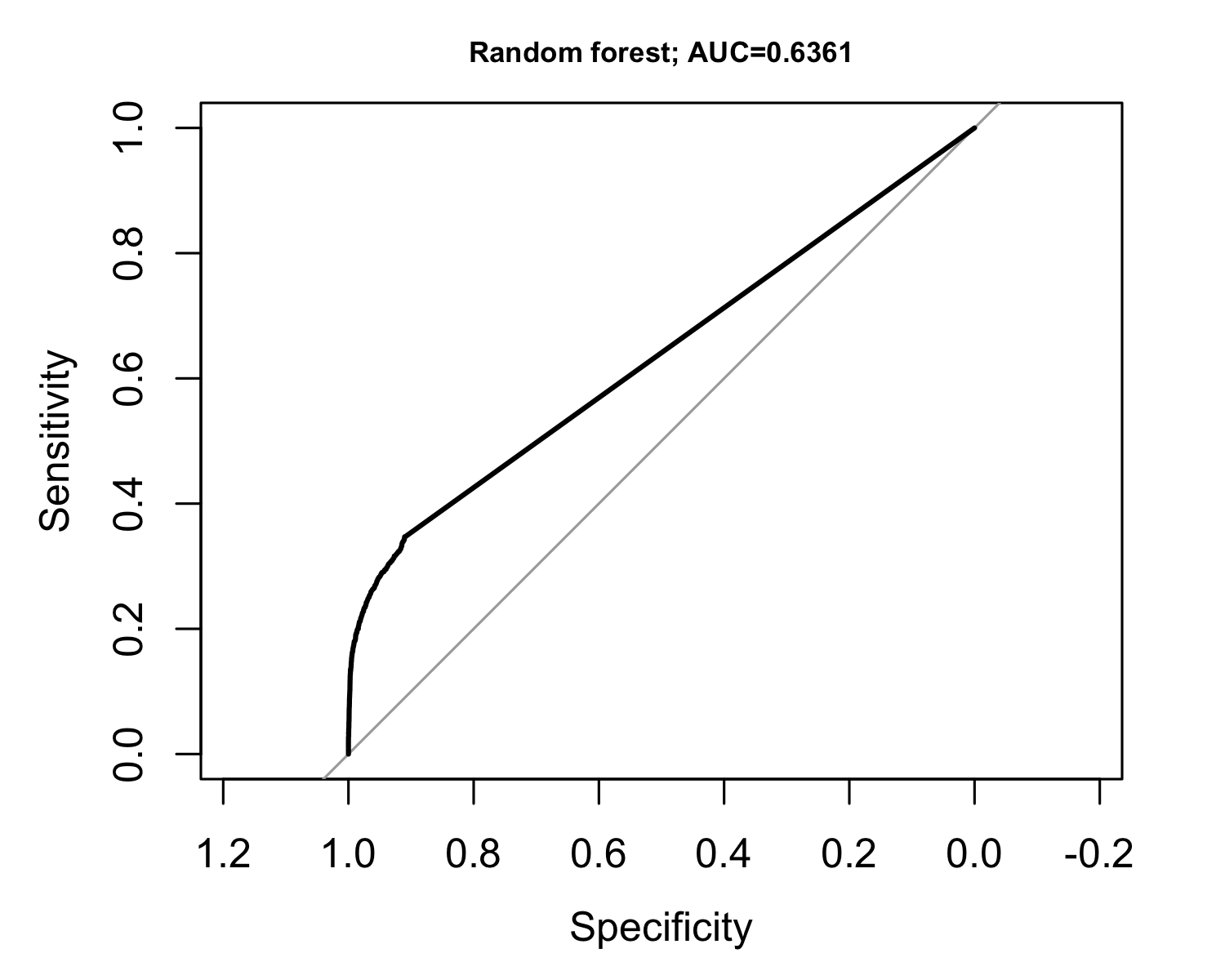
## LASSO model with optimal (in the sense of minimize deviance) lambda (tuning parameter)



## LASSO model with optimal (in the sense of minimize deviance) lambda (tuning parameter) + one standard error



## Random forest



## Ensemble logistic regression (LASSO) with random forest

