Preliminary Result of Paper 2; Aggregation Method B

June 26, 2019

## Table 1

Demographics of the ever had relapse vs never had relapse.

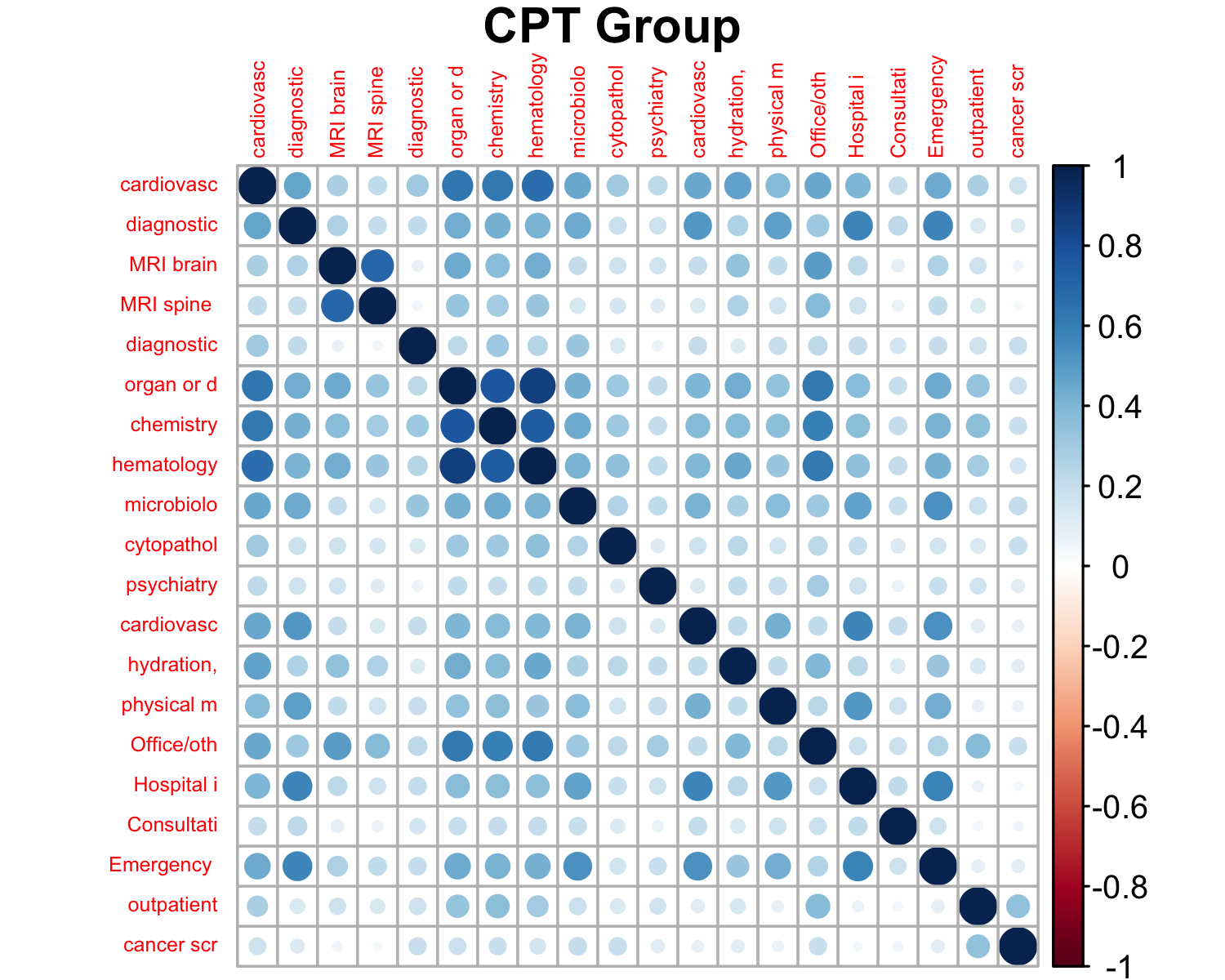
|  |  |  |  |
| --- | --- | --- | --- |
|  | Ever had relapse | Never had relapse | p-value for heter |
| Total number of patients | 1672 | 703 | NA |
| Num belongs to EHR | 1047 | 508 | NA |
| Sex (% female) | 73.8 | 71.4 | 0.233 |
| Race (% white) | 92.5 | 93 | 0.69 |
| Median (IQR) age at first code | 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 24-month. The overall case rate is 0.309.

## Correlations

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



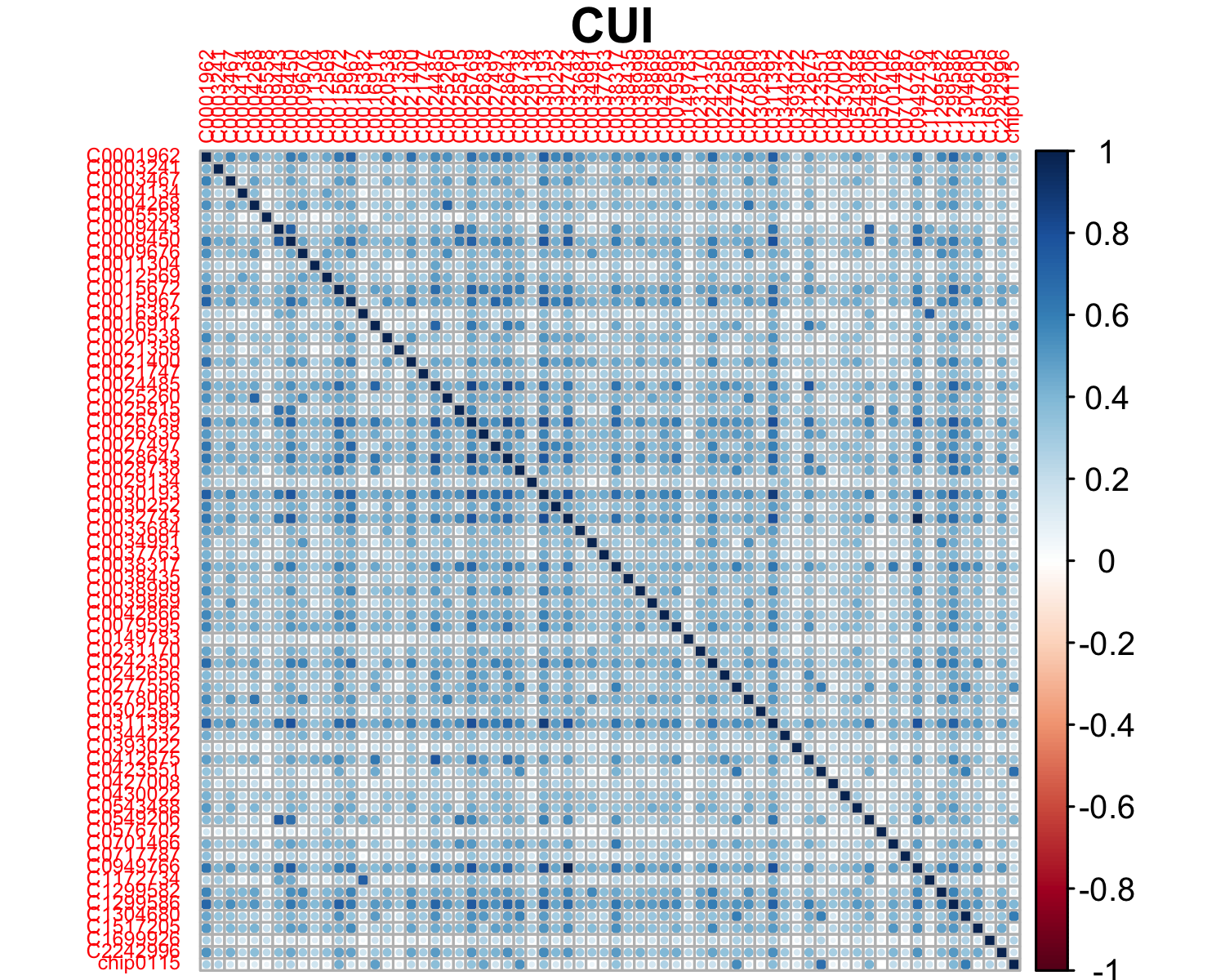
## Among CPT Groups:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.03776 0.17414 0.22820 0.28563 0.39107 0.86498

## CPT Group vs others:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.005287 0.180645 0.271625 0.288641 0.380847 0.786906

|  |  |  |
| --- | --- | --- |
| CUI | CUI | Corr |
| C0024485 | C0028643 | 0.855 |
| C0030193 | C0311392 | 0.874 |
| C0026769 | C0028643 | 0.881 |
| C0032743 | C0949766 | 0.985 |

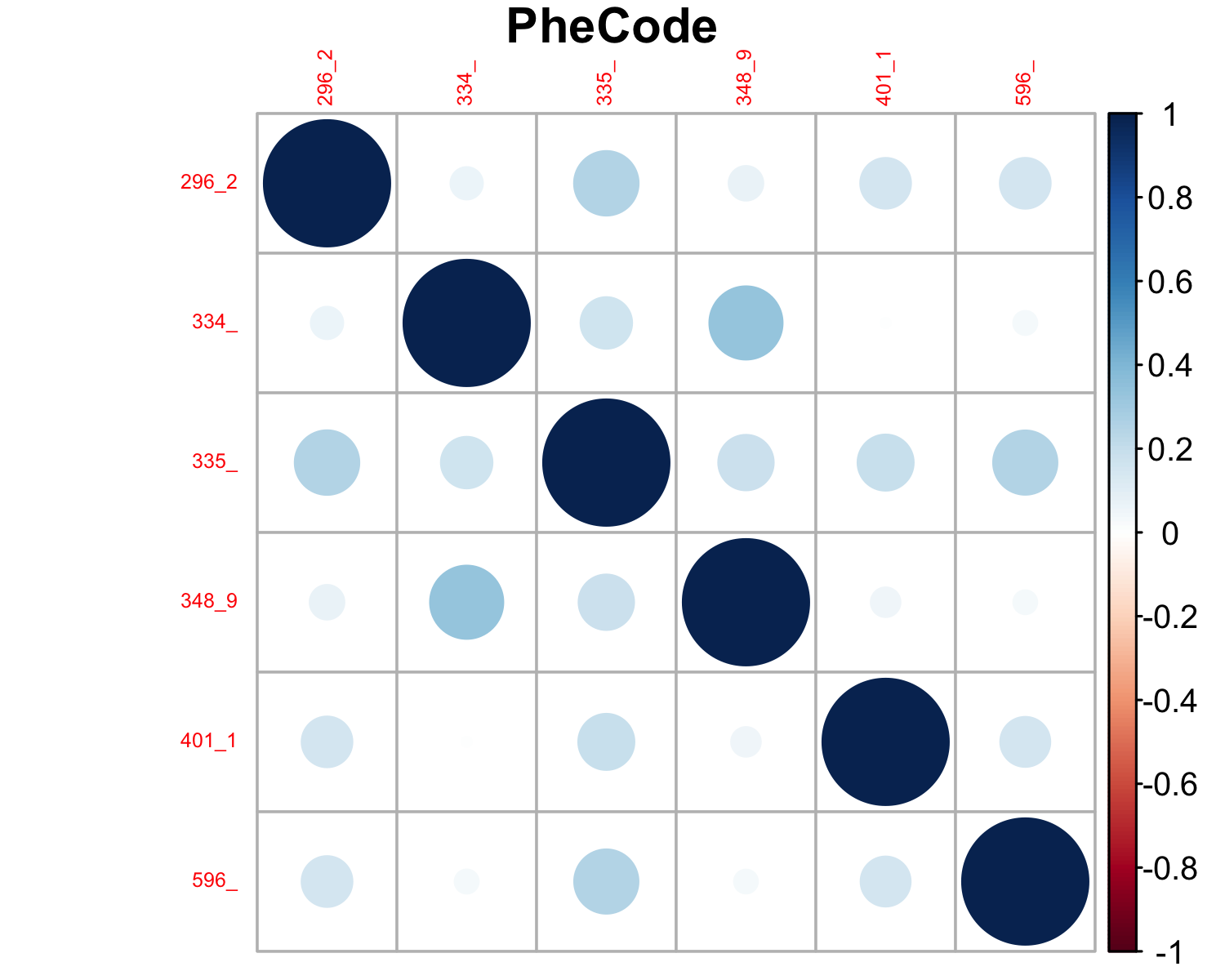


## Among CUIs:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.003276 0.234115 0.328840 0.343236 0.435155 0.984965

## CUI vs others:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0164 0.1686 0.2594 0.2783 0.3679 0.8991



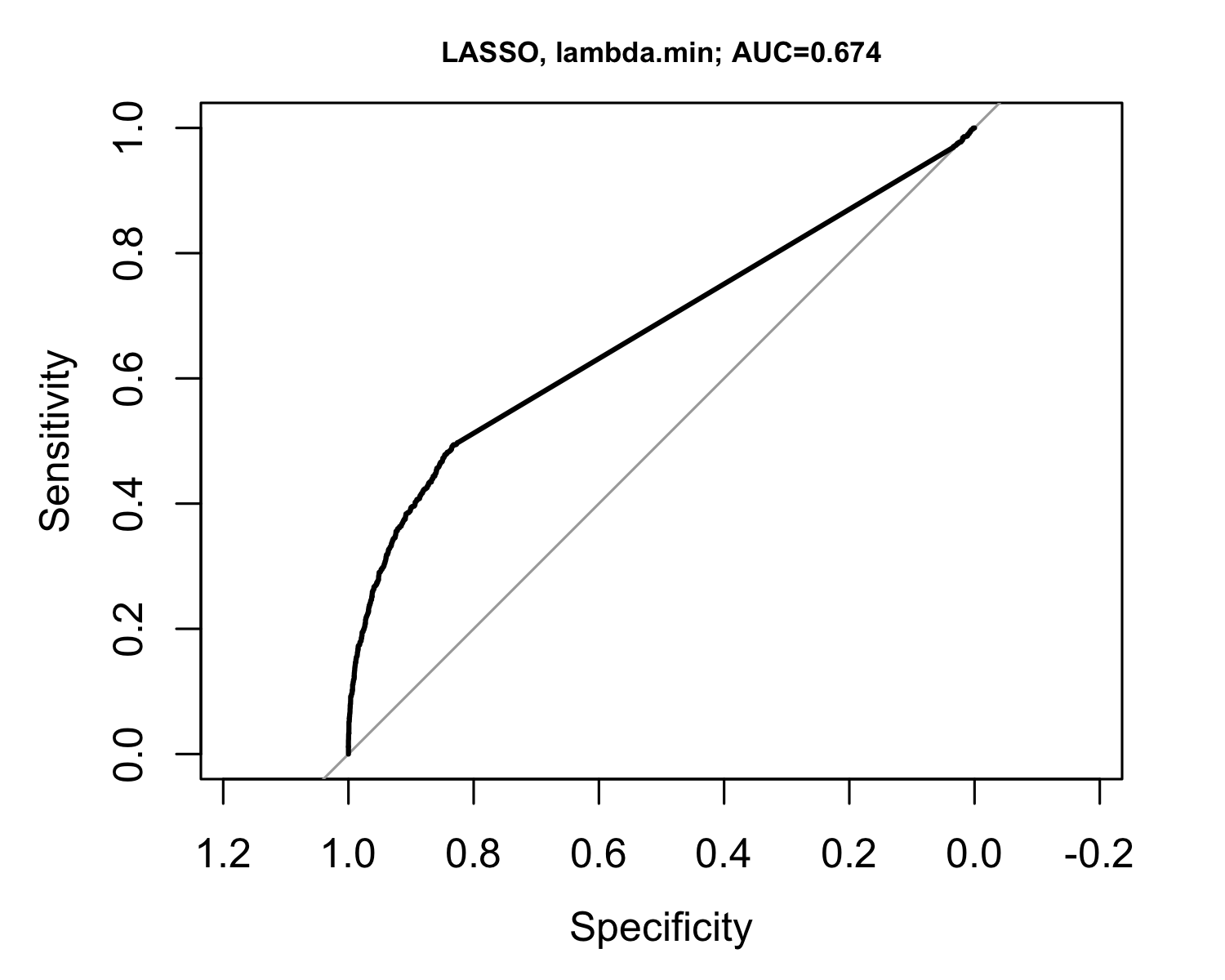
## Among PheCodes:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.005445 0.058328 0.158620 0.142005 0.192166 0.332873

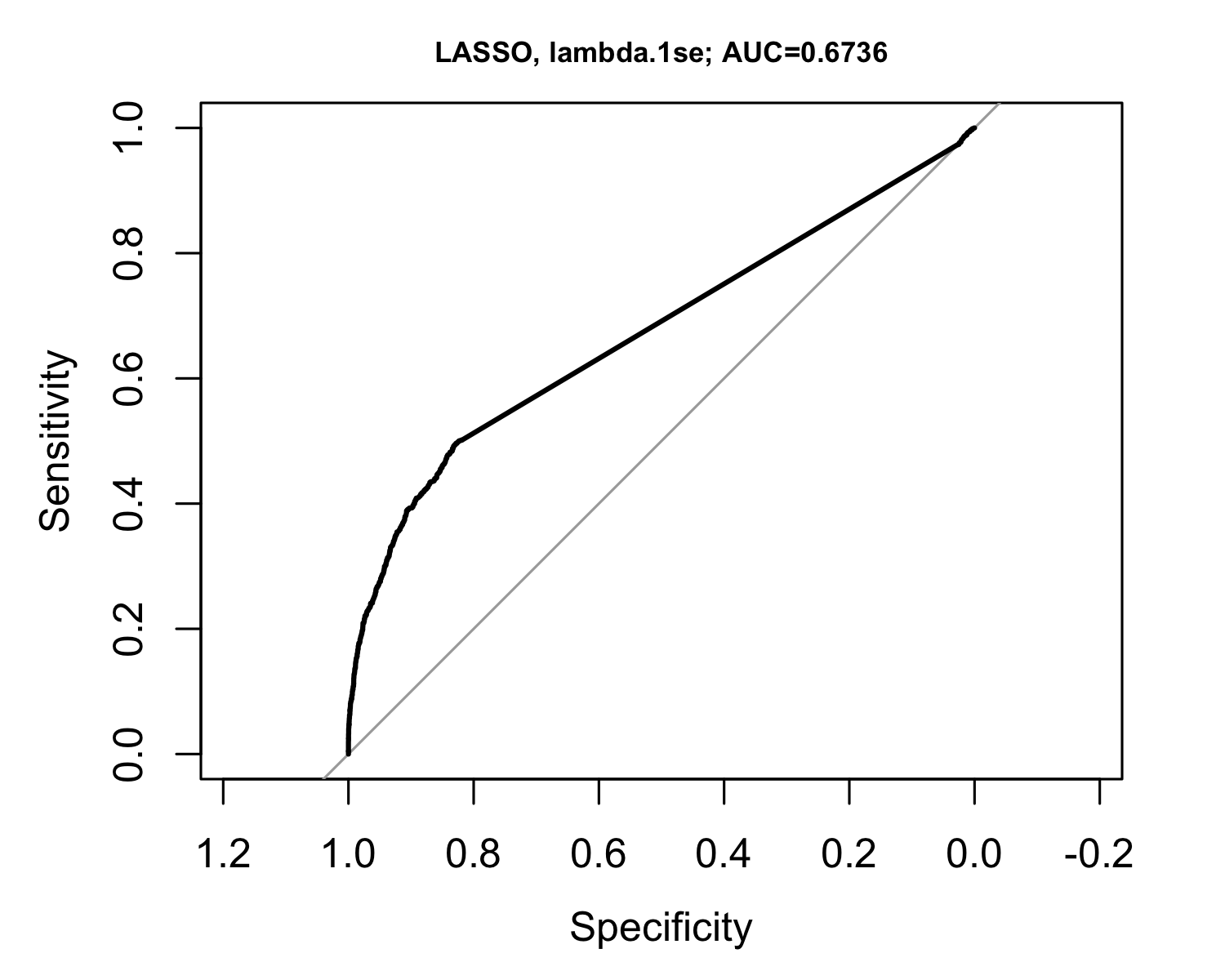
## PheCode vs others:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.005287 0.112567 0.188280 0.219249 0.284872 0.899130

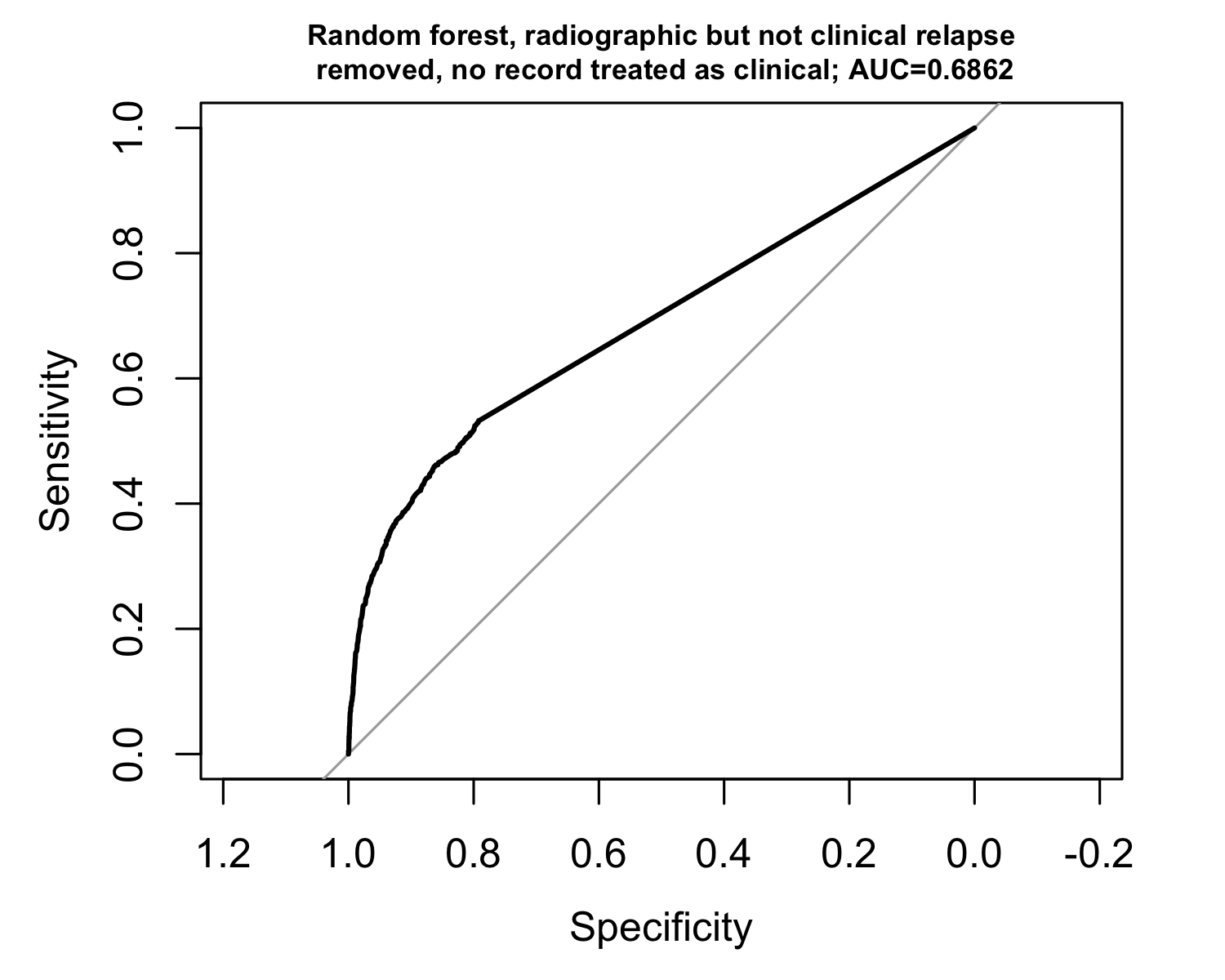
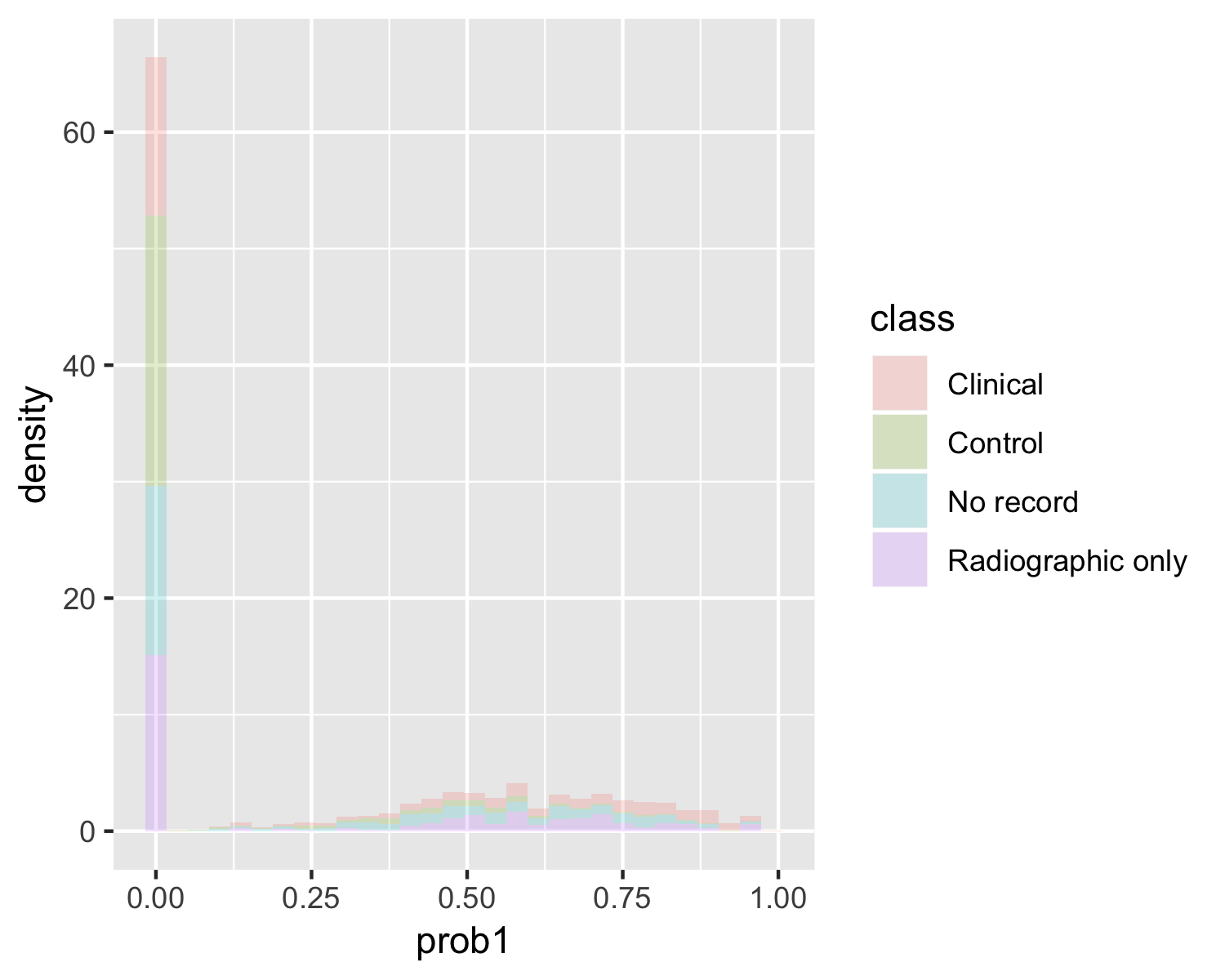
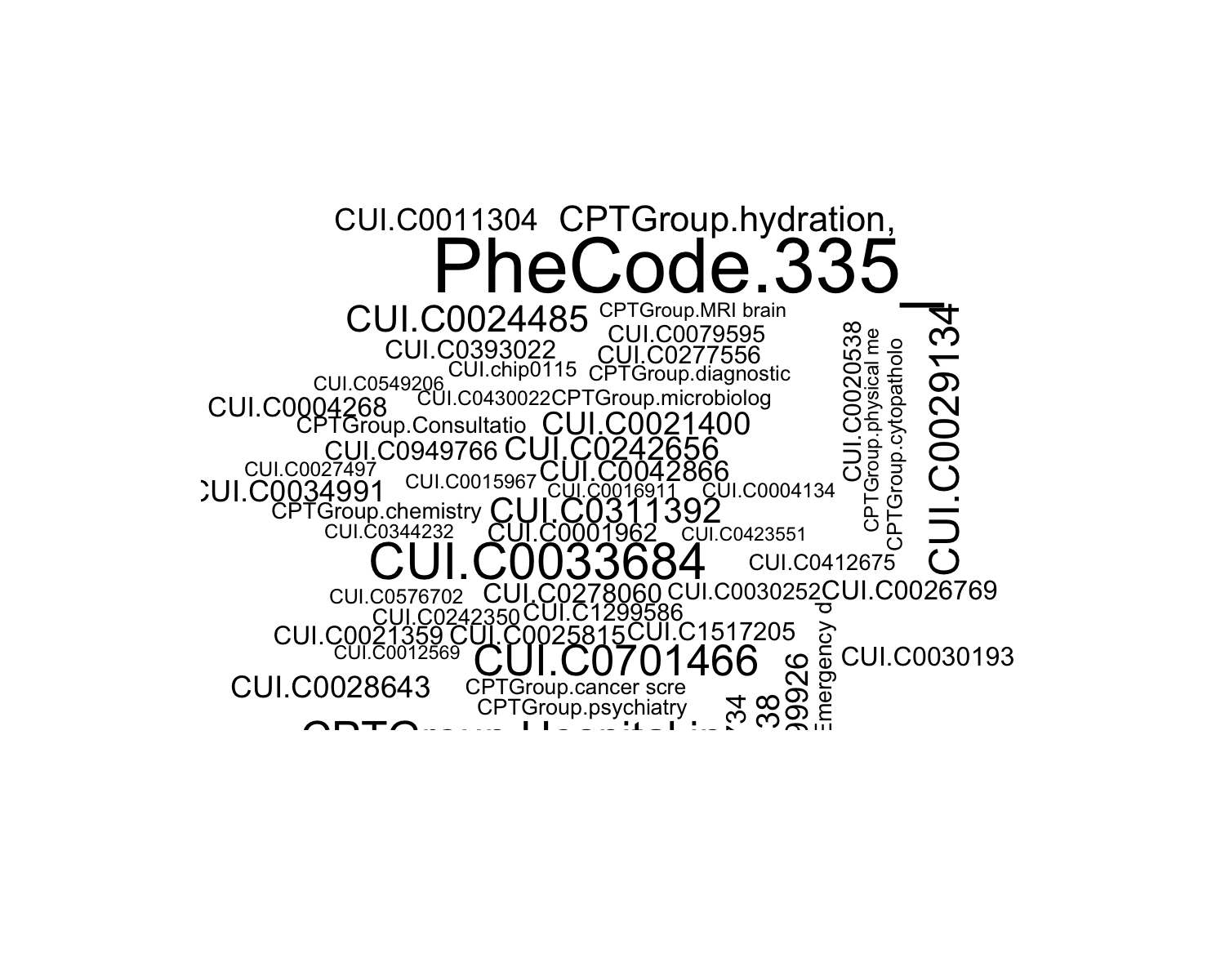
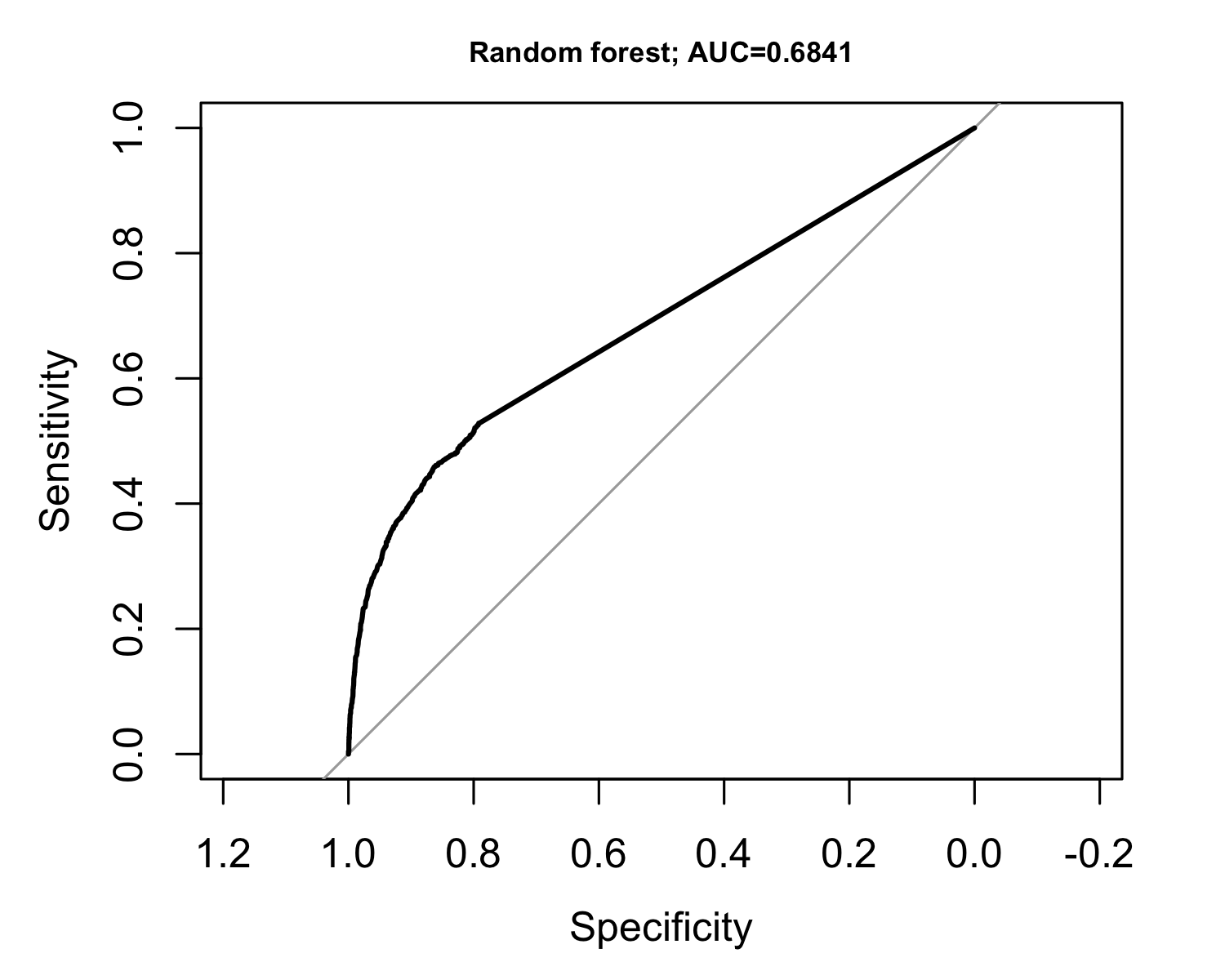
## LASSO model with optimal (in the sense of minimize deviance) lambda (tuning parameter)



## 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

