Preliminary Result of Paper 2; Aggregation Method A

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 |

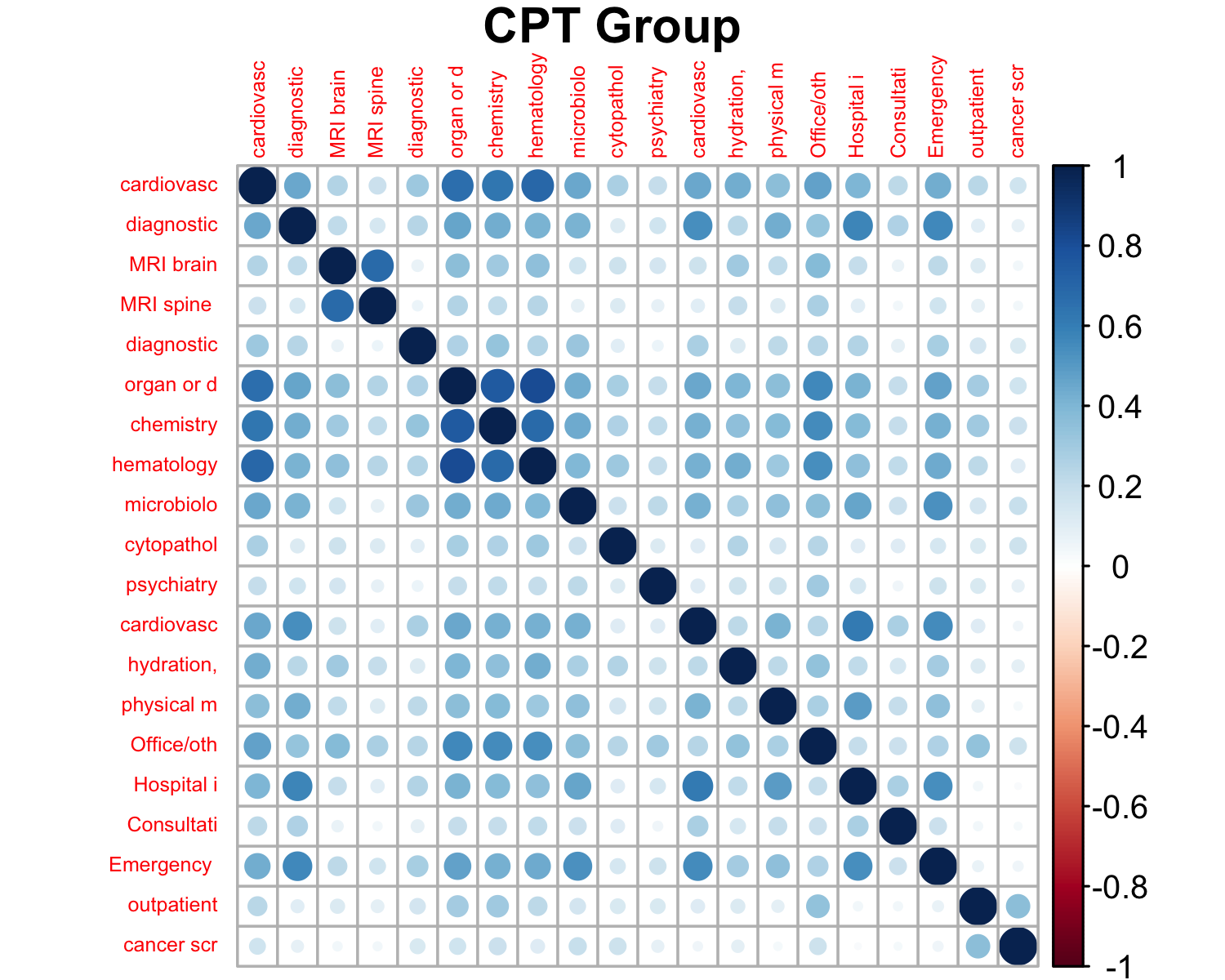
## Data Manipulation

For each encounter, it is classified as a case if the corresponding patient experienced relapse within 12-month window. The overall case rate is 0.254.

We further take a case-control subsample as following. For each patient, his or her follow-up period (enrollment date or first encounter date, whichever comes first) are cut into 3-month time windows. Only one case/control will be sampled from cases/controls within the same 3-month time window. The overall case rate among this subsample is 0.236. No big difference.

## Correlations

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



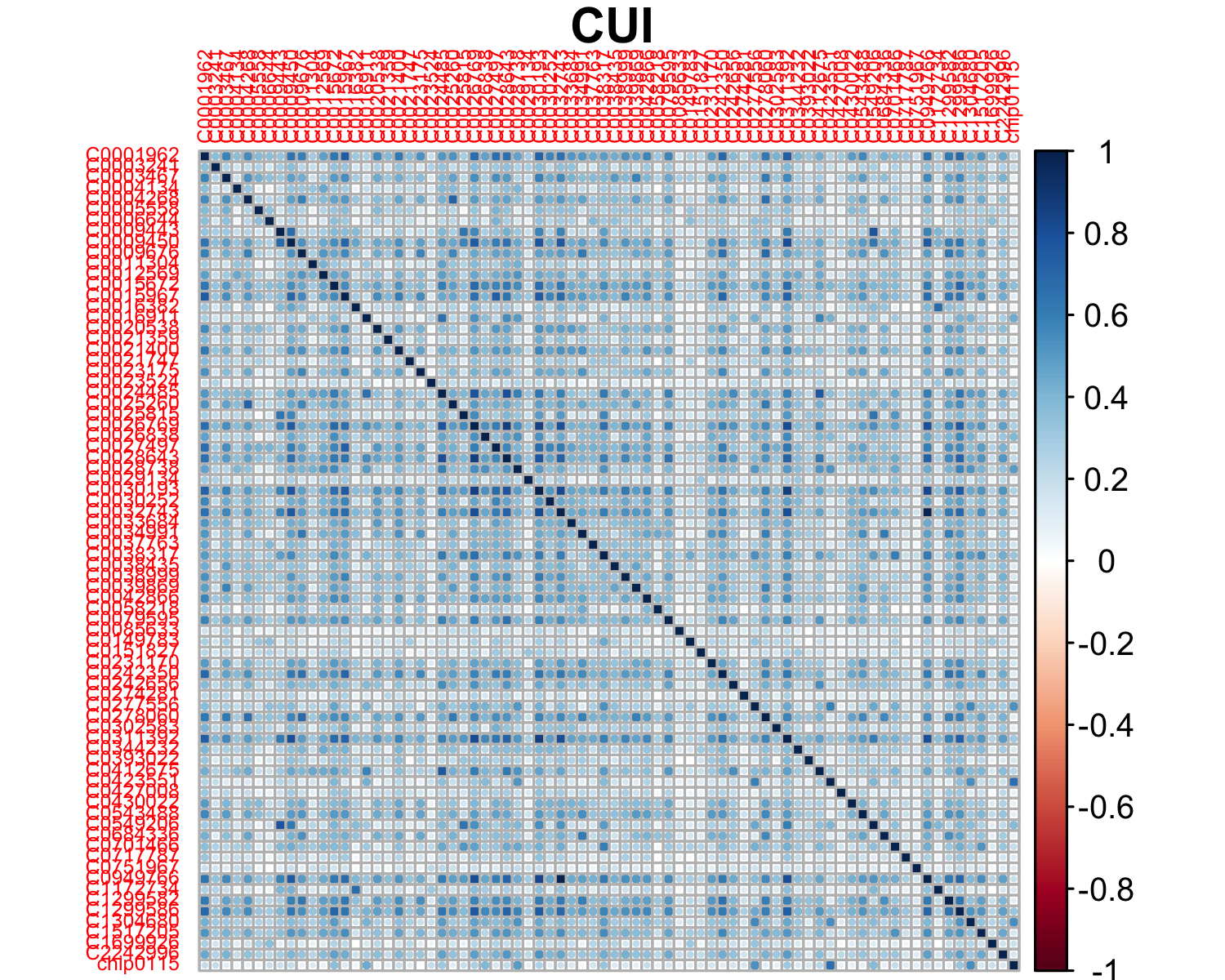
## Among CPT Groups:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.02204 0.14886 0.23926 0.27262 0.38118 0.81765

## CPT Group vs others:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.00279 0.15311 0.24407 0.26559 0.36137 0.79795

|  |  |  |
| --- | --- | --- |
| CUI | CUI | Corr |
| C0026769 | C0030193 | 0.853 |
| C0026769 | C0028643 | 0.861 |
| C0030193 | C0311392 | 0.867 |
| C0032743 | C0949766 | 0.984 |

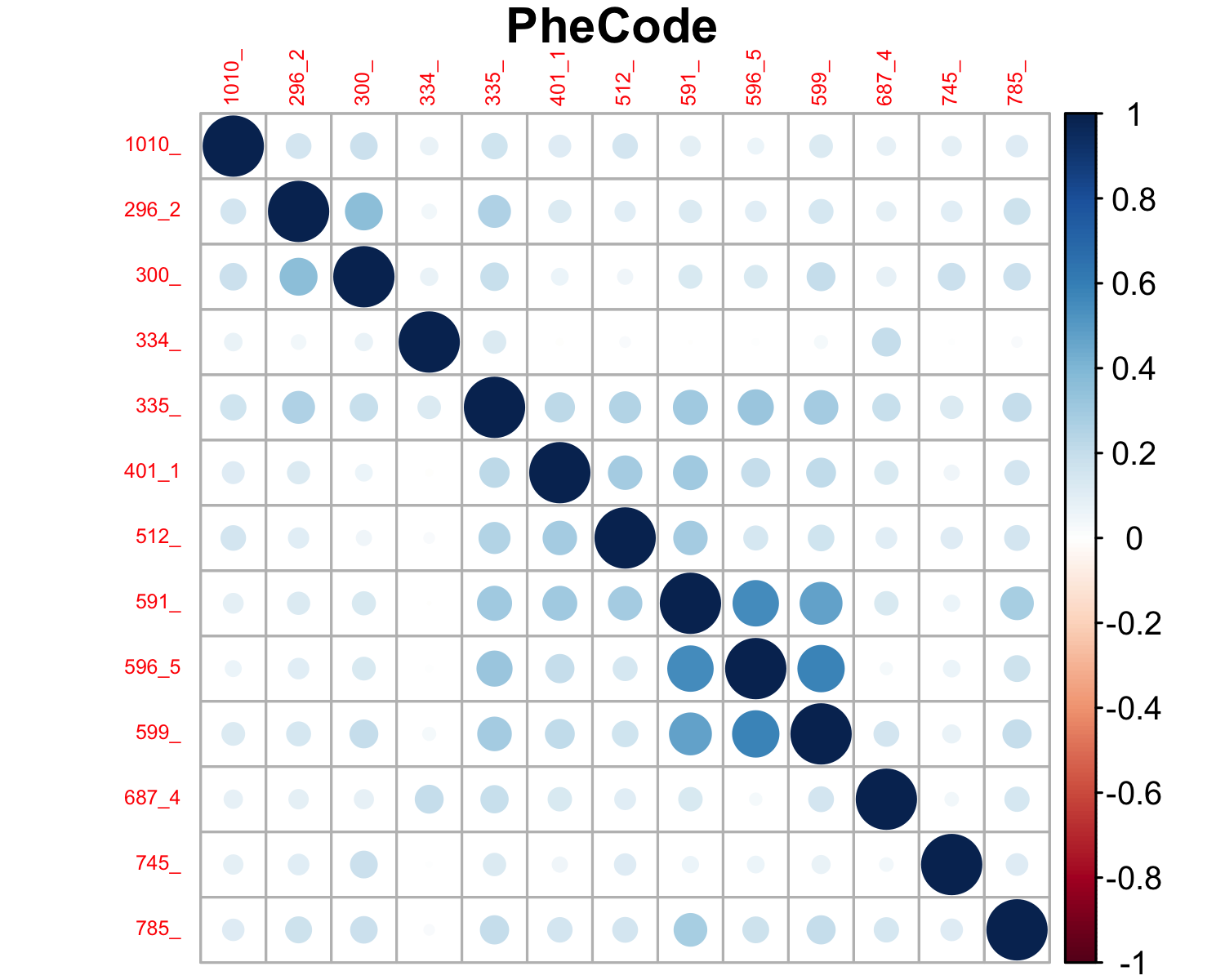


## Among CUIs:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.003663 0.197381 0.294905 0.314593 0.416435 0.983883

## CUI vs others:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.01389 0.14094 0.22549 0.24865 0.33570 0.87489



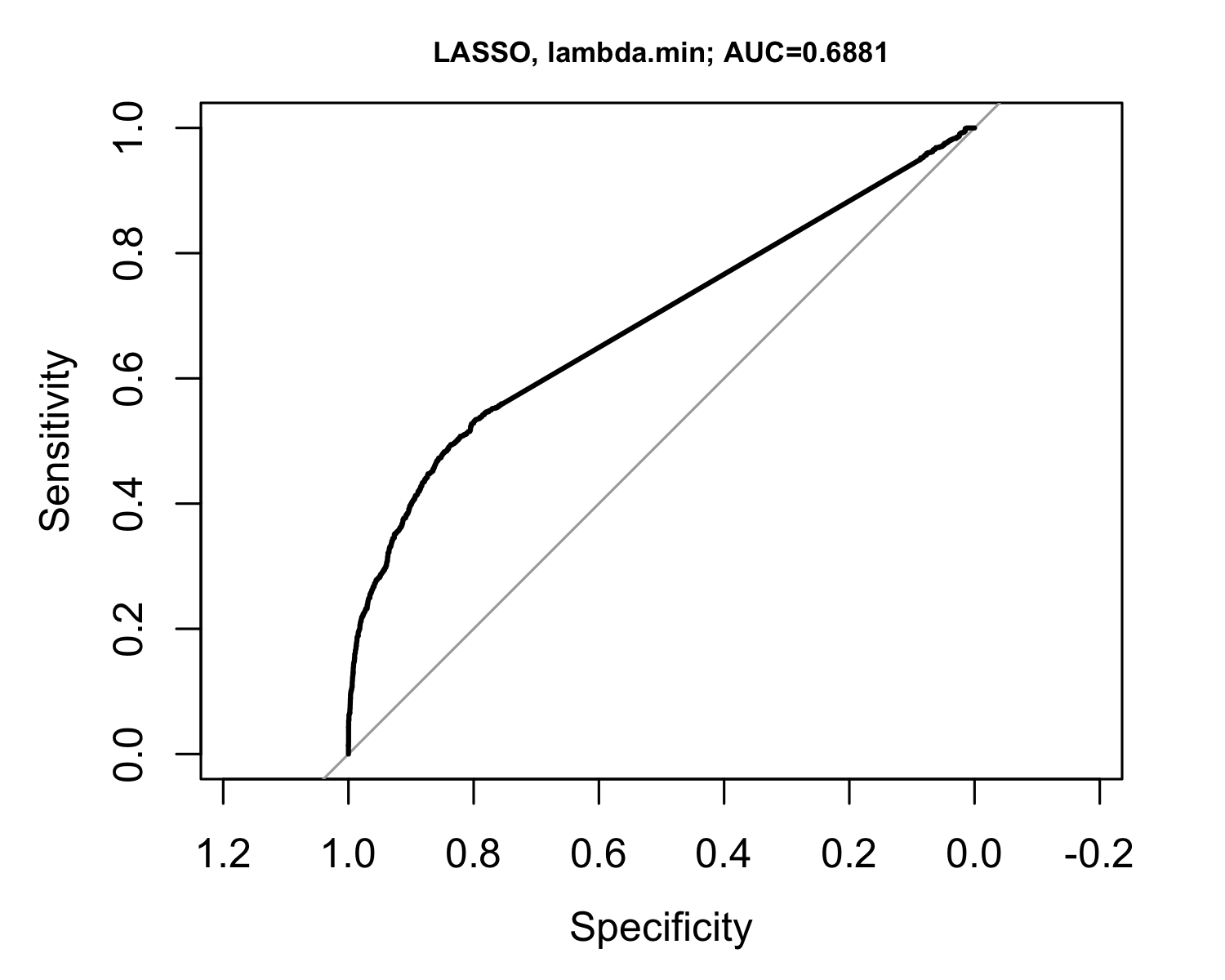
## Among PheCodes:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.008564 0.083921 0.136360 0.156280 0.198828 0.581401

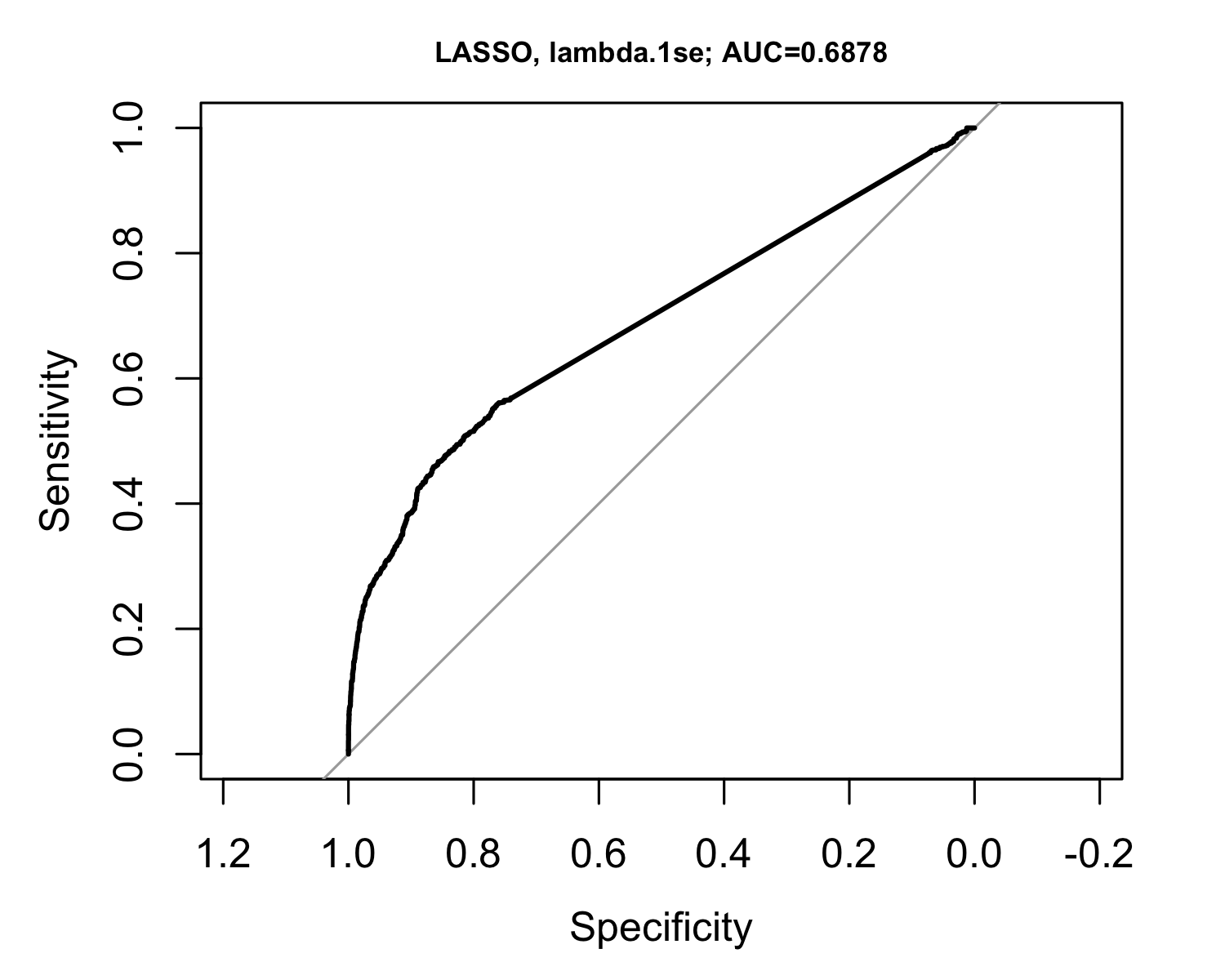
## PheCode vs others:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.01389 0.10601 0.18875 0.20808 0.28761 0.87489

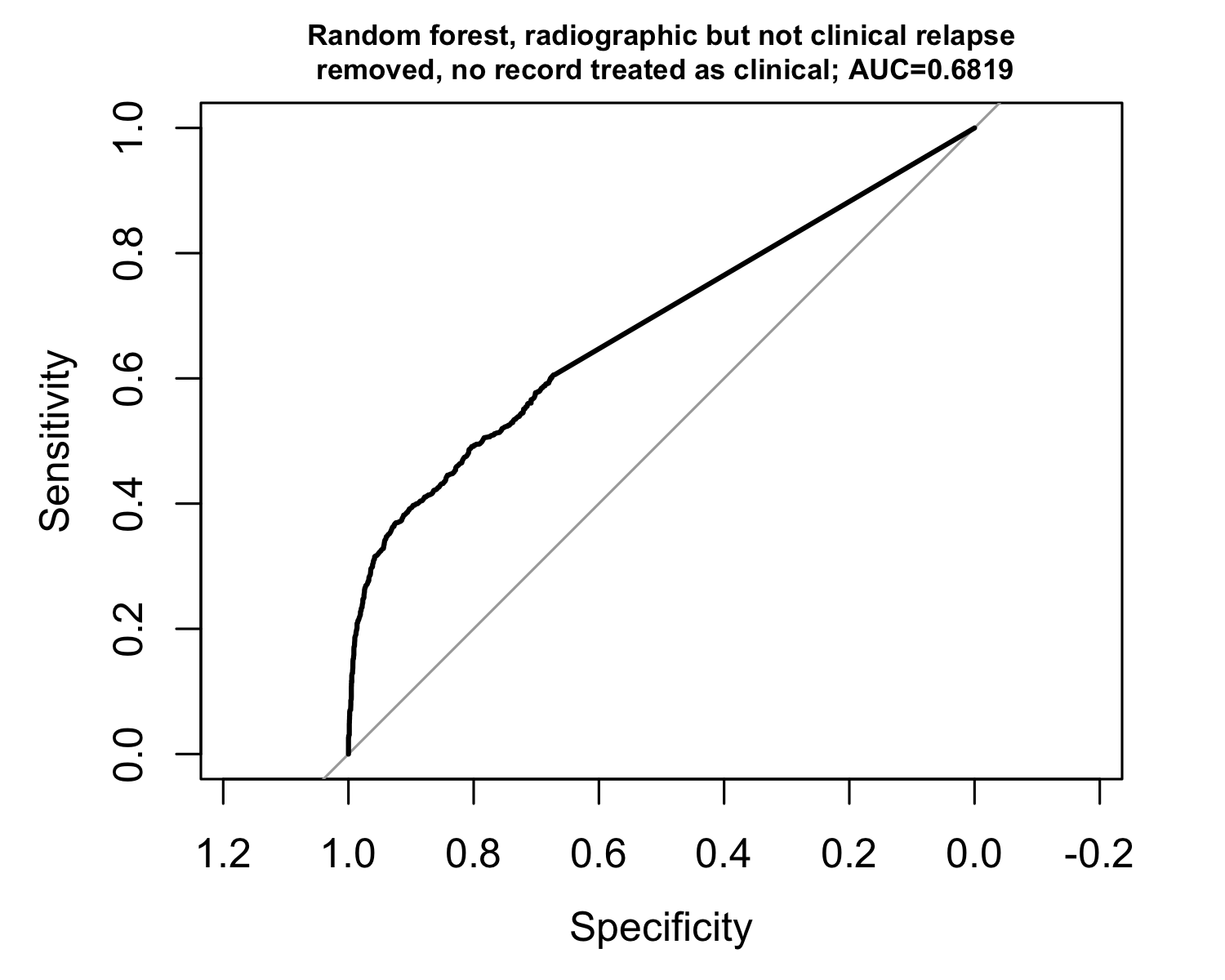
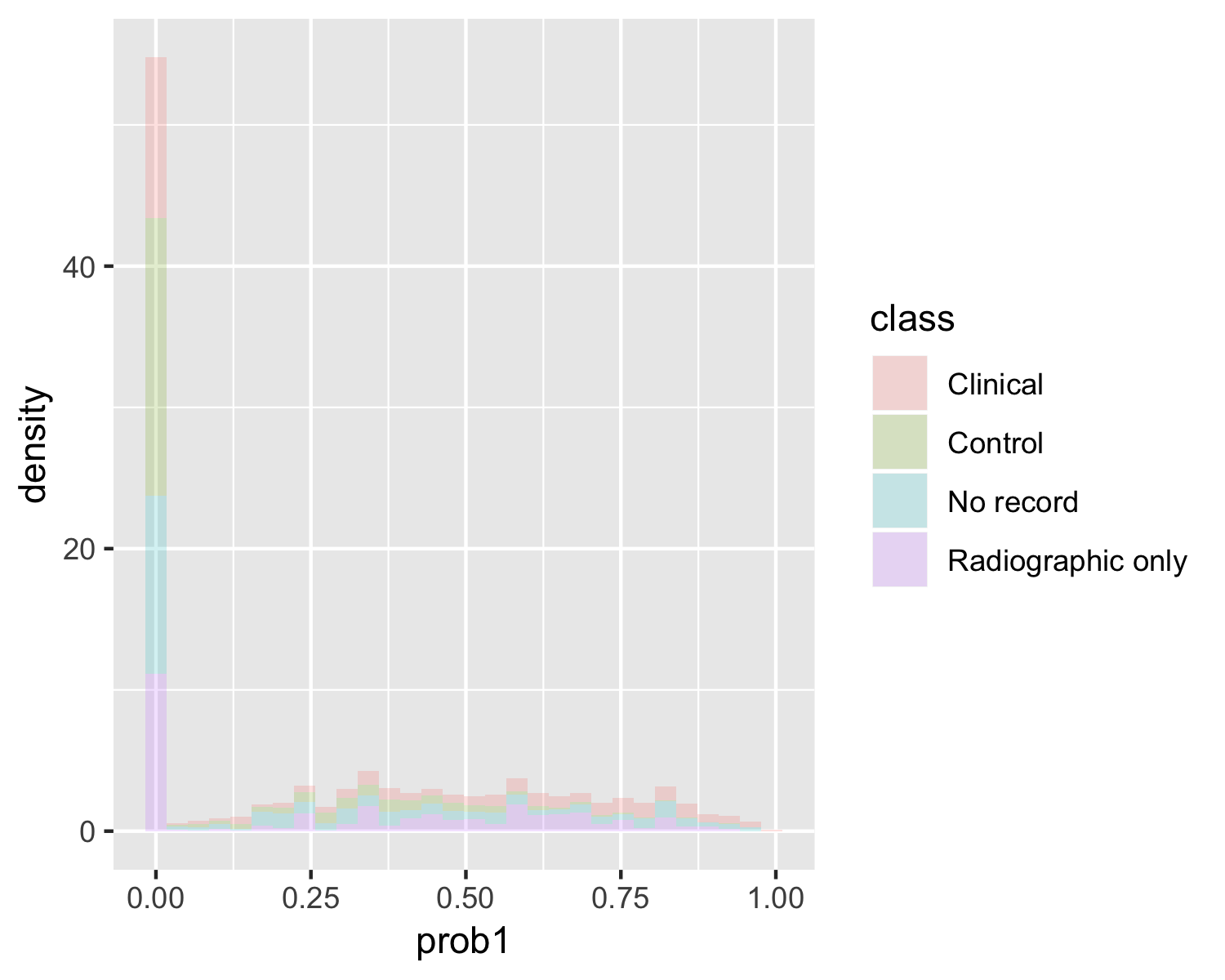
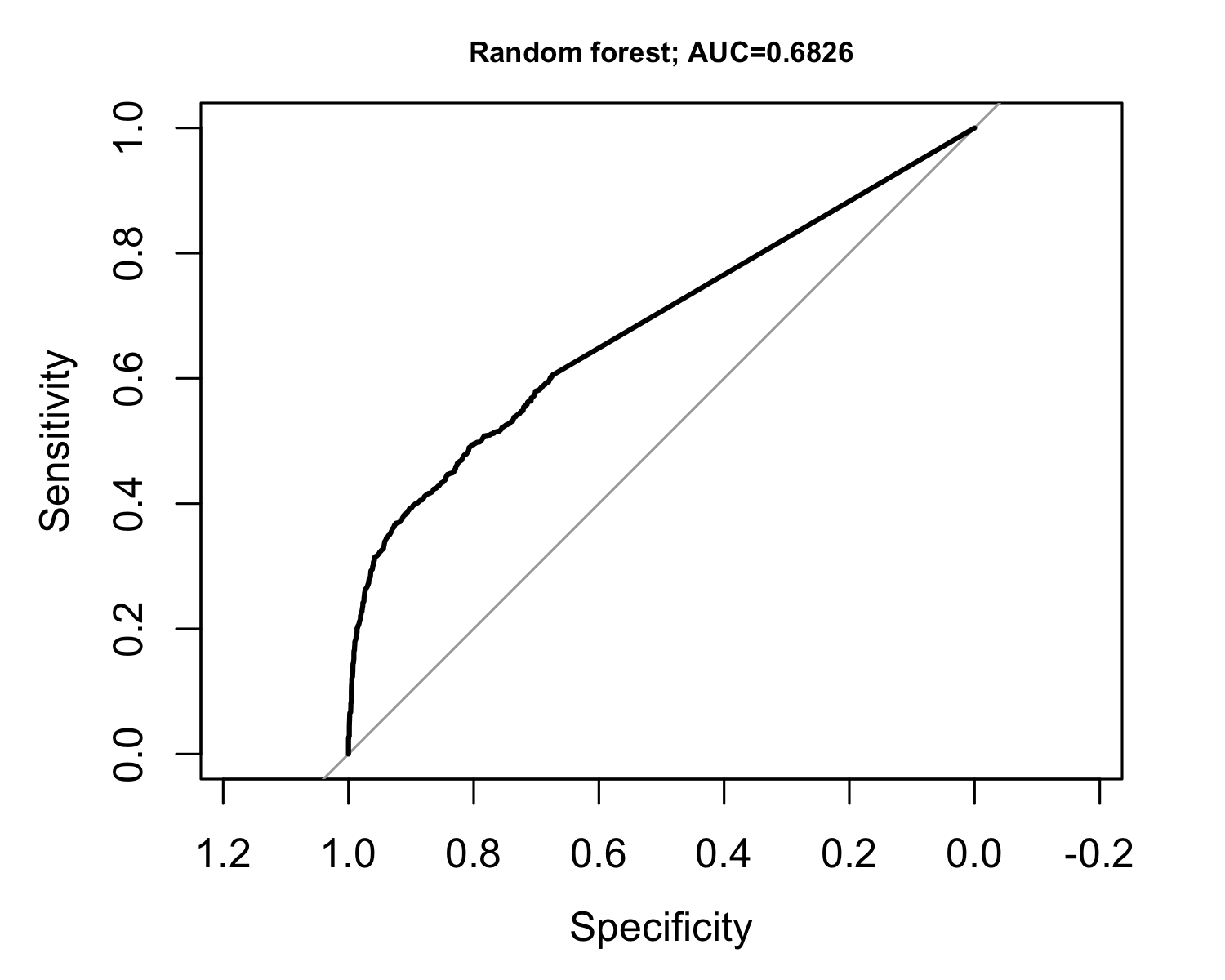
## LASSO model with optimal (in the sense of minimize deviance) lambda (tuning parameter)



## 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

