Preliminary Result of Paper 2; Aggregation Method A

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

For each encounter, it is classified as a case if the corresponding patient experienced relapse within 24-month window. The overall case rate is 0.375.

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 0.355. 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.   
## -0.007153 0.106989 0.181207 0.218875 0.284327 0.845964

## CPT Group vs others:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.01928 0.10897 0.19465 0.22483 0.31236 0.79673

|  |  |  |
| --- | --- | --- |
| CUI | CUI | Corr |
| C0030193 | C0032743 | 0.850 |
| C0030193 | C0949766 | 0.859 |
| C0026769 | C0030193 | 0.875 |
| C0026769 | C0028643 | 0.883 |
| C0030193 | C0311392 | 0.889 |
| C0032743 | C0949766 | 0.986 |

## Among CUIs:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.01243 0.18746 0.27874 0.30573 0.40895 0.98583

## CUI vs others:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.01953 0.11144 0.19369 0.21741 0.29743 0.89042

## Among PheCodes:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.01479 0.05795 0.13414 0.15142 0.21254 0.61460

## PheCode vs others:

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.01953 0.09229 0.17000 0.18875 0.26304 0.89042

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

A close up of a map

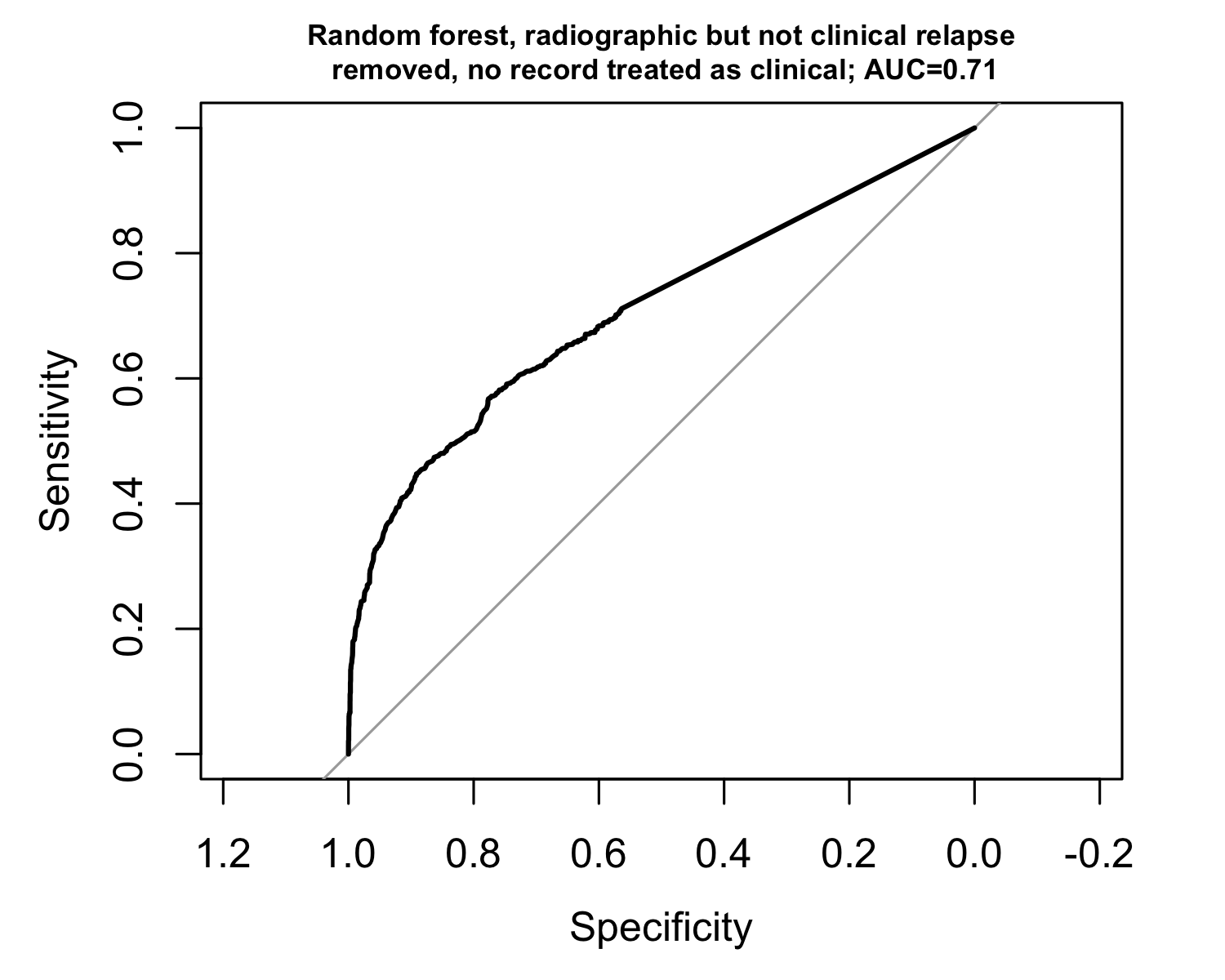
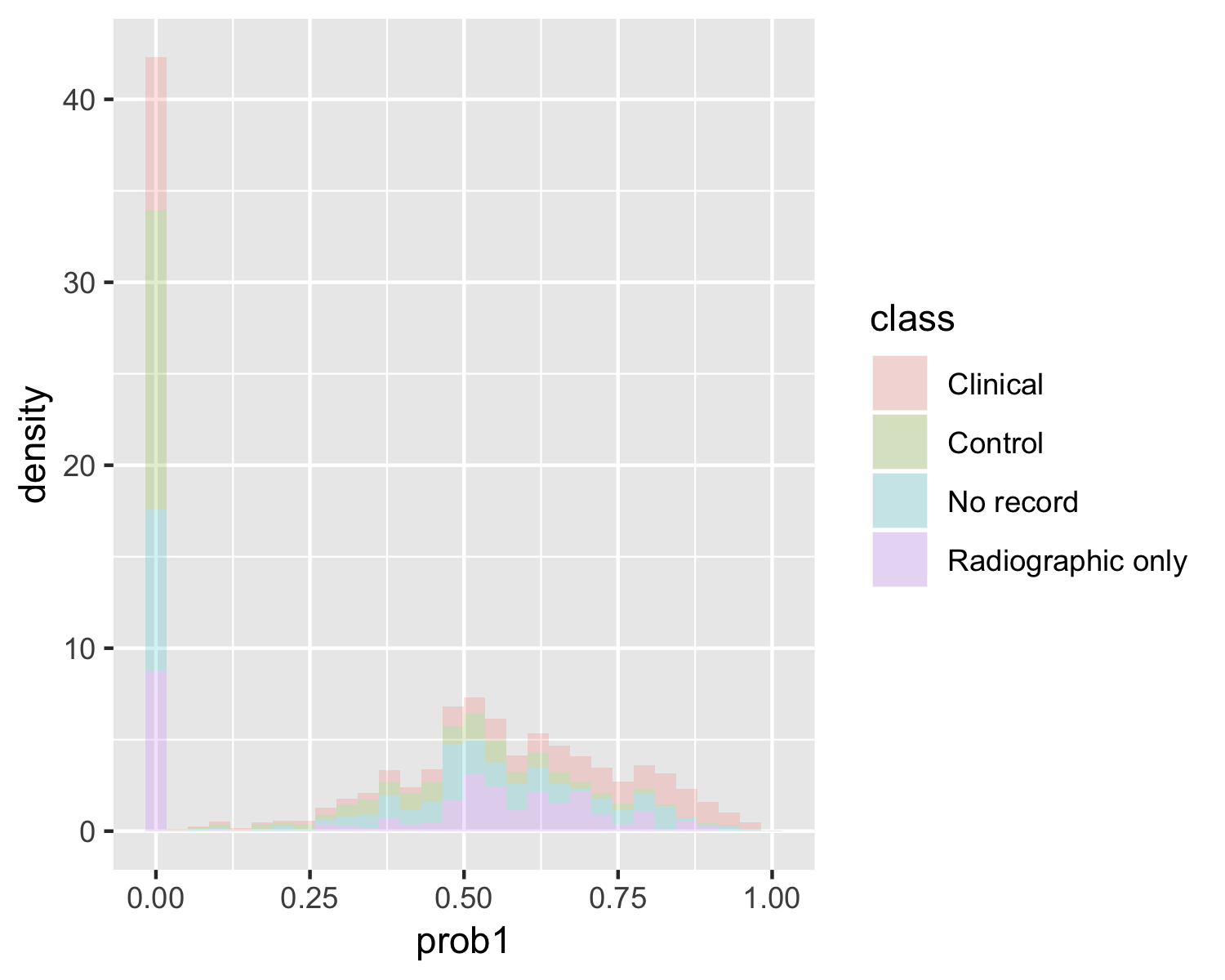
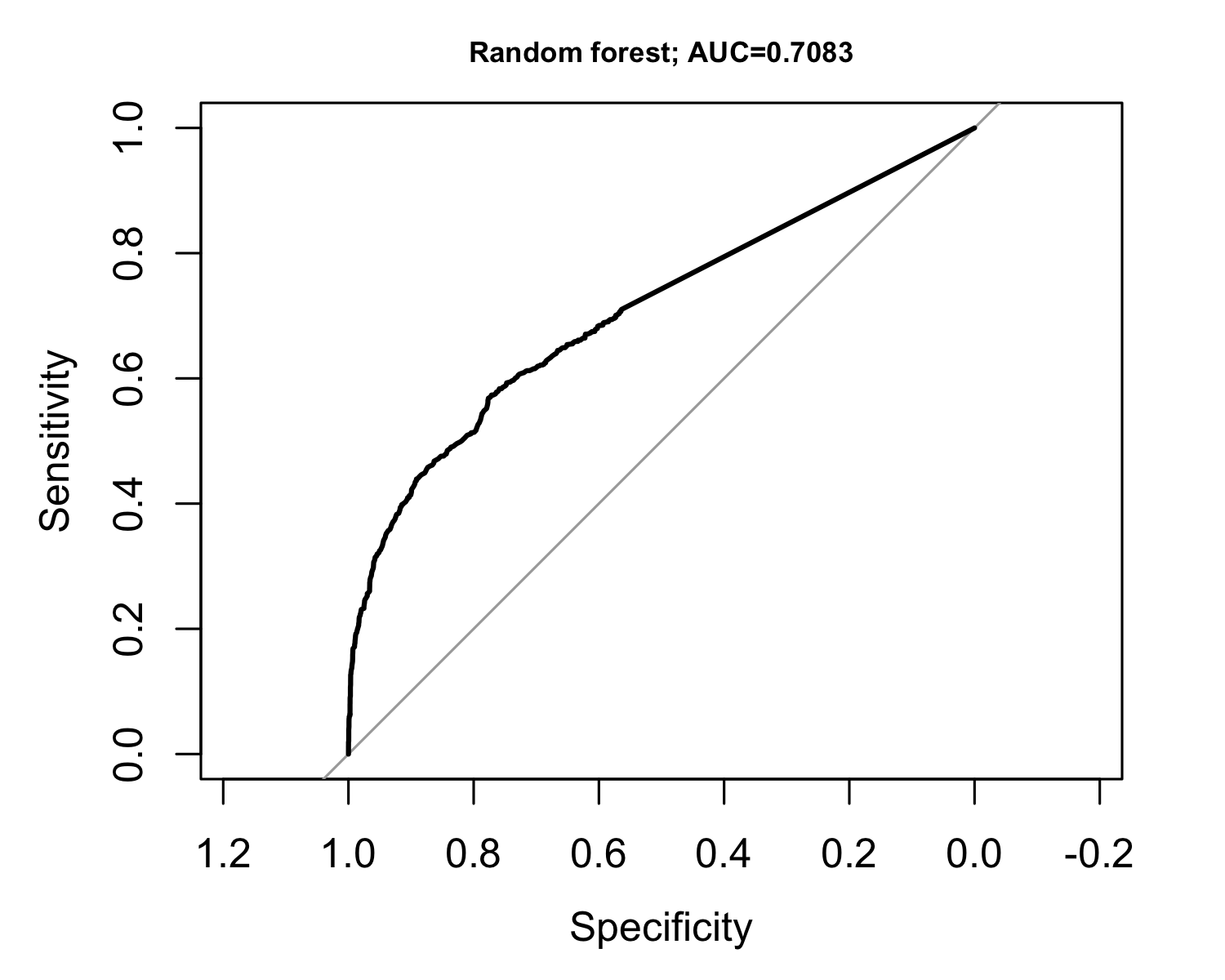
Description automatically generated

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

A close up of a map

Description automatically generated

## Random forest



## Ensemble logistic regression (LASSO) with random forest

