Is Dementia predictable?

F. Di Filippo, E. Manfrin, E. Musiari, E. Palli

17 december 2021

Our team

Francesca Di Filippo

Erica Manfrin

Elena Musiari

Edoardo Palli

Dataset Dementia and Alzheimer longitudinal

Subject.ID	MRI.ID	Group	Visit	MR.Delay	M.F	Hand	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
OAS2_0001	OAS2_0001_MR1	Nondemented	1	0	M	R	87	14	2	27	0.0	1987	0.696	0.883
OAS2_0001	OAS2_0001_MR2	Nondemented	2	457	M	R	88	14	2	30	0.0	2004	0.681	0.876
OAS2_0002	OAS2_0002_MR1	Demented	1	0	M	R	75	12	NA	23	0.5	1678	0.736	1.046
OAS2_0002	OAS2_0002_MR2	Demented	2	560	M	R	76	12	NA	28	0.5	1738	0.713	1.010
OAS2_0002	OAS2_0002_MR3	Demented	3	1895	M	R	80	12	NA	22	0.5	1698	0.701	1.034
OAS2_0004	OAS2_0004_MR1	Nondemented	1	0	F	R	88	18	3	28	0.0	1215	0.710	1.444
OAS2_0004	OAS2_0004_MR2	Nondemented	2	538	F	R	90	18	3	27	0.0	1200	0.718	1.462
OAS2_0005	OAS2_0005_MR1	Nondemented	1	0	M	R	80	12	4	28	0.0	1689	0.712	1.039
OAS2_0005	OAS2_0005_MR2	Nondemented	2	1010	M	R	83	12	4	29	0.5	1701	0.711	1.032

where SES is Socioeconomic Status, MMSE is Mini Mental State Examination, CDR is Clinical Dementia Rating, eTIV is Estimated Total Intracranial Volume, nWBV is Normalize Whole Brain Volume and ASF is Atlas Scaling Factor.

Source: Kaggle

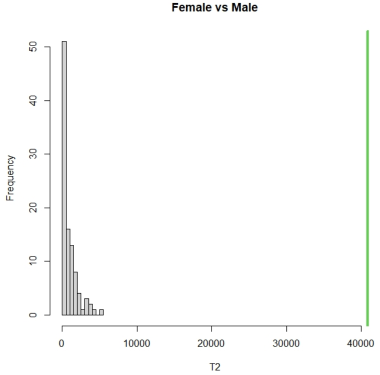
Analysis Male VS Female

Variable considered: $[Age, EDUC, MMSE, eTIV, nWBV, ASF]$

$$H_0 : \mathbf{Y}_{female} \overset{d}{=} \mathbf{Y}_{male} \text{ vs } H_1 : \mathbf{Y}_{female} \overset{d}{\neq} \mathbf{Y}_{male}$$

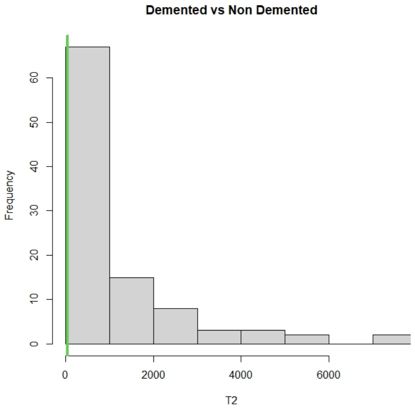
$$T_0 = |\bar{Y}_{female} - \bar{Y}_{male}|$$

$$pvalue = 0$$



Demented VS Nondemented

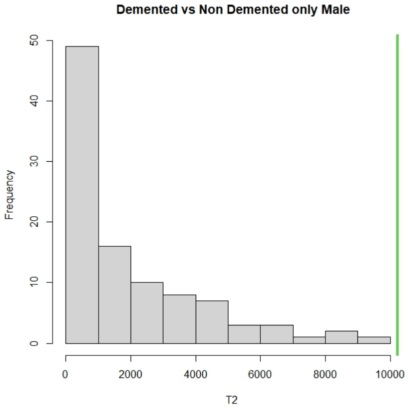
$$H_0 : \mathbf{Y}_{Demented} \overset{d}{=} \mathbf{Y}_{NonDemented} \;\; vs \;\; H_1 : \mathbf{Y}_{Demented} \overset{d}{\neq} \mathbf{Y}_{NonDemented}$$



$pvalue = 0.9$

Demented VS Nondemented only Male

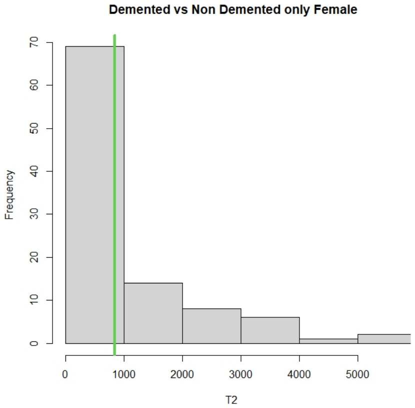
$$H_0 : \mathbf{M}_{Demented} \stackrel{d}{=} \mathbf{M}_{NonDemented} \; vs \; H_1 : \mathbf{M}_{Demented} \stackrel{d}{\neq} \mathbf{M}_{NonDemented}$$



$pvalue = 0$

Demented VS Nondemented only Female

$$H_0 : \mathbf{F}_{Demented} \overset{d}{=} \mathbf{F}_{NonDemented} \ vs \ H_1 : \mathbf{F}_{Demented} \overset{d}{\neq} \mathbf{F}_{NonDemented}$$



$pvalue = 0.37$

TWO WAYS ANOVA: EDUCATION (I)

$$EDUC = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon$$

$j = \{male, female\}$, $i = \{Demented, NonDemented\}$
$\alpha = sex$, $\beta = diagnostic$, $\gamma = interaction$

$$H_0 : \gamma_{ij} = 0 \ \ vs \ H_1 : \gamma_{ij} \neq 0$$

TEST STATISTIC: $T0 = F - STATISTICS$
$p - value = 0.082$ at level of confidence 95% there's no evidence to reject H_0, so we reduce the model

TWO WAYS ANOVA: EDUCATION (II)

$$EDUC = \mu + \alpha_i + \beta_j + \epsilon$$

$$H_0 : \beta_j = 0 \;\; vs \; H_1 : \beta_i \neq 0$$

$p - value = 0.069$ at level of confidence 95% there's no evidence to reject H_0

$$EDUC = \mu + \alpha_i$$

$$H_0 : \alpha_i = 0 \;\; vs \; H_1 : \alpha_i \neq 0$$

$p - value = 0.08$ we could say that's neither of the grouping is significant at 95%
while with parametric test at least the diagnostic division is significant

TWO WAYS ANOVA: MMSE (I)

$$MMSE = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon$$

$j = \{male, female\}$, $i = \{Demented, NonDemented\}$

$$H_0 : \gamma_{ij} = 0 \ \ vs \ H_1 : \gamma_{ij} \neq 0$$

TEST STATISTIC: $T0 = F - STATISTICS$
$p - value = 0.875$
there's no evidence to reject H_0, so we reduce the model

$$EDUC = \mu + \alpha_i + \beta_j + \epsilon$$

$$H_0 : \beta_j = 0 \;\; vs \; H_1 : \beta_j \neq 0$$

$p - value = 0.446$ there's no evidence to reject H_0

$$EDUC = \mu + \alpha_i$$

$$H_0 : \alpha_i = 0 \;\; vs \; H_1 : \alpha_i \neq 0$$

$p - value = 0$ there's evidence to reject H_0 so the most significative model is $MMSE \sim Diagnostic$ where Diagnostic is the division between Demented and Non Demented

Regression

Used a logistic model of classification Demented-Nondemented, with smoothing splines (degree 3) for EDUC, nWBV, Age, MMSE

$$\log \frac{p}{1-p} = EDUC + nWBV + Age + MMSE + CDR$$

with $p =$ probability that the observation is 'Demented'

```
        Shapiro-Wilk normality test

data:  model_gam$residuals
W = 0.93137, p-value = 2.796e-11
```

Prediction

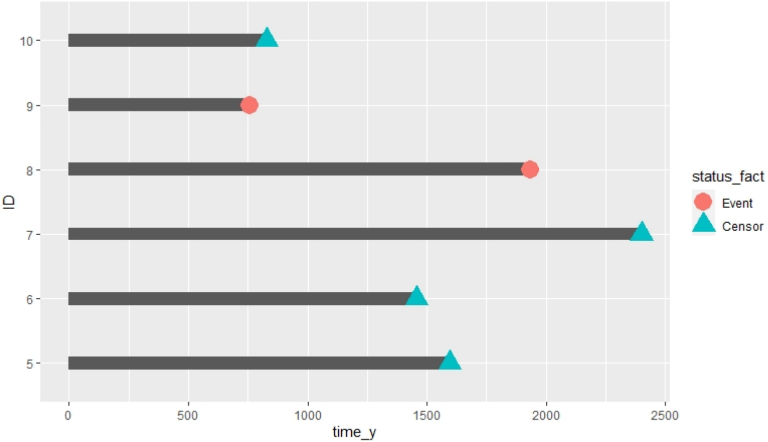
Prediction for Converted patients:

```
> options(scipen = 100, digits = 7)
> pred
            34         35         36         37         38         39         58         59         60
pred 0.05316856 0.09237775 0.674794 0.05863945 0.6125401 0.8149156 0.08133388 0.06052873 0.6179306
            82        83        84        115        116        195        196        219        220
pred 0.2627466 0.143016 0.6089418 0.07991747 0.8438092 0.1722836 0.7792069 0.08776355 0.5495934
           221        246        247        262        263        264        265        266        272
pred 0.5288126 0.04125604 0.8154823 0.0546213 0.5348506 0.5376504 0.5582587 0.5111782 0.6304302
           273        274        275        296        297              298        299        347        348
pred 0.4895987 0.1248788 0.6994842 0.05183226 0.5211393 -0.00002013915 0.5363979 0.04796561 0.04393508
           349
pred 0.4964958
```

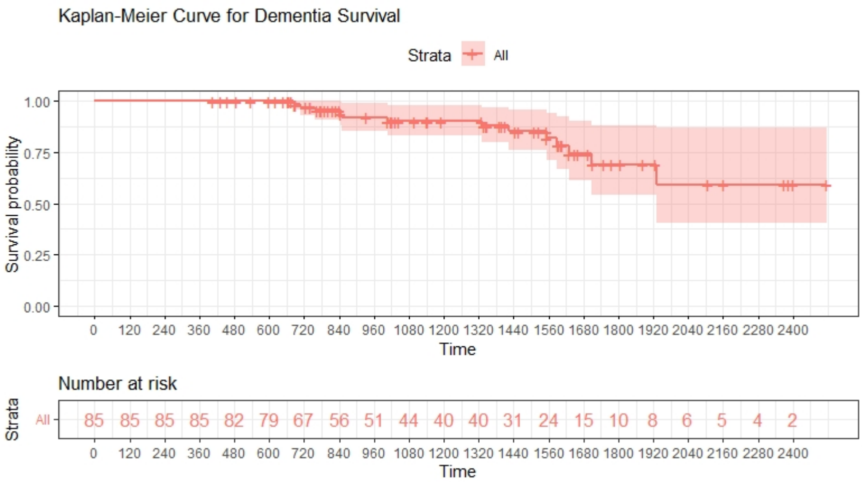
Prediction for Nondemented patients:

```
         1          2          6          7          8          9
0.33197223 0.08501153 0.16418897 0.21870711 0.23499001 0.64639607
        10         20         21         22         23         24
0.12727284 0.06378905 0.08046988 0.12099097 0.10163539 0.10233746
        25
0.18136779
```

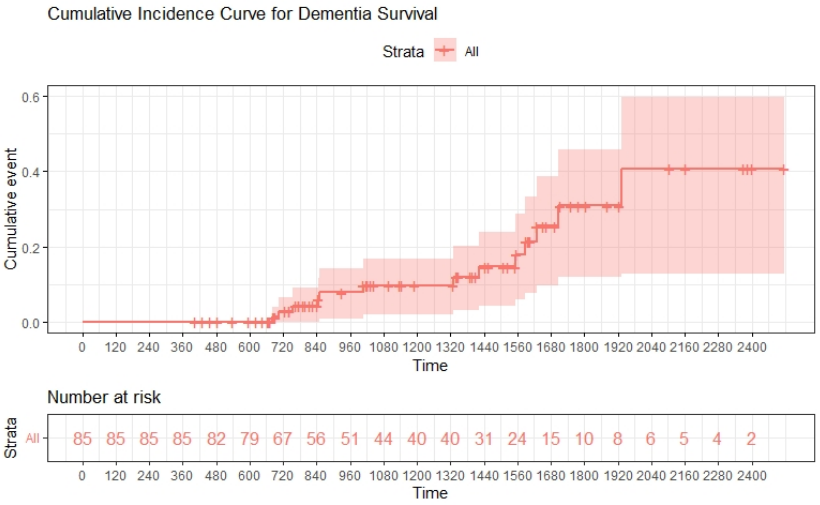
Event: disease occurred



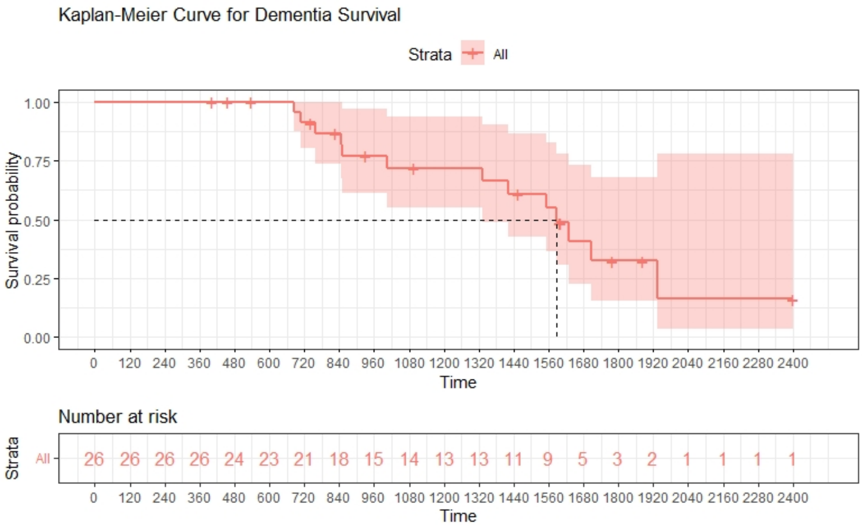
Survival Analysis (II)



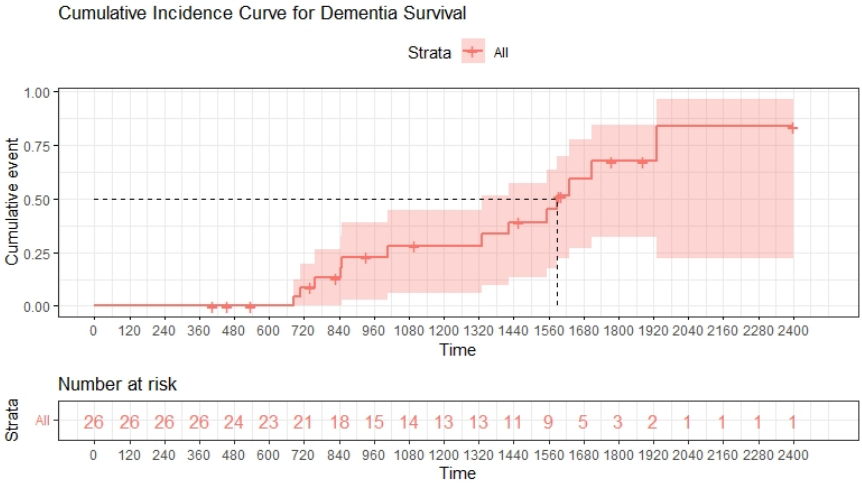
Survival Analysis(III)



Survival Analysis - Balanced (I)



Survival Analysis - Balanced (II)



Future goals

- Complete our survival analysis
- Solve the residual gaussianity problem
- Perform a prediction on the other dataset of patient (not labelled)