Leveraging multimodal data to predict outcomes of antidepressant treatment

Preliminary results

Brice Ozenne, Vibeke Dam

December 9, 2021

1 Data

The data (n=98) we use has been provided by Emily Beaman. She processed relevant data from the CIMBI database. In addition we compute:

- a "MR" biomarker as the average of the thickness of the left lateral, right lateral, left medial, and right medial orbitol frontal cortex.
- a "PET" biomarker by combining the log binding potential from neocortex, hippocampus, caudate, and putamen via a latent variable model (adjusted for age, gender, and injected mass). This was done in a leave-one-out fashion i.e. the biomarker for individual i was obtained by fitting the model on all but the i-th individual and estimating the latent variable for the i-th individual based on its log-binding in the 4 regions.
- a "cognition" biomarker was obtained via a k-means algorithm on various cognitive outcomes (no leave-one-out here).
- the "outcomes" by computing the relative change in HAMD6 between week 8 (or 12) and baseline.

Individuals who had missing HAMD6 at both week 8 and 12 were excluded (n=13)

The corresponding R code is in the file `0-data-management.R` available on Github.

2 Data analysis

After discussion with neuroscientist, we have identified 10 candidate biomarkers[1] for predicting recovery after SSRI treatment:

- `MR_OFCthick`: thickness of the OFC brain region measured with MR.
- `HAMD17`: depression score at baseline.
- `low_hsCRP`: high sensitivity CRP (l levels of inflammation in the body).
- `lvpet`: summary of the brain log-PET binding.
- `cognitive_cluster`: summary of the cognition
- `EEG_vigilance`: EEG signal (vigilance slope B1 bl)
- `CATS_scoretotal`: ??
- `CAR_AUCi`: Difference between the cortisol value and the cortisol value at wake-up cumulated over an hour.
- `neuroticism`:

Missing values: to simplify the analysis, we will assume that missing data occured completely at random. In particular, that it is not related to the outcome (patient did not leave the study because they fully recover or they were so seriously depressed that they could not stay in the study).

Association between recovery and biomarkers[2]: to assess whether the biomarker were associated with recovery we fitted a logistic regression with gender and age vs. a logistic regression with gender and age plus all biomarkers (as additive effects). A likelihood ratio test was used to compare the two models, i.e. assess the assocation over all biomarkers. Wald tests were used to test the association for each biomarker. P-value were adjusted for multiple comparison (i.e. FWER control) using a max-test adjustment.
This procedure was performed 4 times: using week 8 or week 12 as the outcome, using complete case analysis (excluding CATS, CAR, and neuroticism as biomarkers) or using multiple imputation (MI) to handle missing values. When using MI, the original data was cloned 100 times. In each clone, missing values were imputed using Fully Conditional Specification (FCS) implemented by the MICE algorithm (Van Buuren and Groothuis-Oudshoorn, 2011). This algorithm alternates between learning the relationship between variables, using a linear regression for continuous variables, logistic regression for binary variables, and a a proportional odds model

[1] fMRI is missing in the list

[2] how does the recovery vary in average (i.e. at a population level) as a function of the biomarkers

for categorical variables with all variables (outcome, age, gender, biomarkers) as predictors (as suggested in Moons et al. (2006)), and impute by sampling from the resulting distributions (rougthly speaking, a noisy version of the best prediction).

Predictive value of the biomarkers[3]: to assess whether the biomarkers can be used to discriminate between patients who will recover and patients who won't, we tested whether the Area Under the Curve (AUC) of a logistic model with age, gender, and the biomarkers was greated than 0.5, and whether it was greater than a model with only age and gender. We also compared the performance with a random forest (default hyperparameters: mtry = 3, 500 trees, node size 1). The AUC was estimated via 10 fold cross-validation repeted 100 times. An AUC was computed over all folds of a given repetition and then averaged acrossed repetions. The p-value was computed via a permutation test (100 repetitions).
Calibration plot were also computed using the local polynomial regression fitting (loess).
Missing value was handled either using complete case analysis (excluding CATS, CAR, and neuroticism as biomarkers) or predicting the recovery probability for a given subject based on a logistic model containing all biomarkers for which the subject has available data.

The corresponding R code is in the file `1-prediction.R` available on Github.

[3]are the biomarkers useful to predict recovery for an individual

3 Results

3.1 Descriptive statistics

The dataset contained 85 patients, 84 with the outcome at week 8 and 81 with the outcome at week 12. Some summary statistics are displayed below:

```
    sex          age          MR_OFCthick        HAMD17         hsCRP
male  :25   Min.   :18.24   Min.   :2.318   Min.   :18.00   Length:85
female:60   1st Qu.:22.22   1st Qu.:2.510   1st Qu.:20.00   Class :character
            Median :23.93   Median :2.558   Median :22.00   Mode  :character
            Mean   :27.22   Mean   :2.573   Mean   :22.76
            3rd Qu.:28.70   3rd Qu.:2.636   3rd Qu.:25.00
            Max.   :57.31   Max.   :2.889   Max.   :31.00
```

```
    lvpet          cognitive_cluster EEG_vigilance       CATS_scoretotal
Min.   :-0.82582   Min.   :1.00      Min.   :-1.500000   Min.   : 0.00
1st Qu.:-0.49196   1st Qu.:1.00      1st Qu.: 0.000000   1st Qu.:16.25
Median :-0.42337   Median :2.00      Median : 0.000000   Median :23.50
Mean   :-0.43227   Mean   :1.88      Mean   : 0.006173   Mean   :30.42
3rd Qu.:-0.34832   3rd Qu.:3.00      3rd Qu.: 0.000000   3rd Qu.:41.50
Max.   :-0.09773   Max.   :3.00      Max.   : 1.500000   Max.   :81.00
NA's   :2          NA's   :2         NA's   :4           NA's   :11
   CAR_AUCi         neuroticism     Y_w8           Y_w12
Min.   :-1070.30   Min.   : 67   Mode :logical   Mode :logical
1st Qu.:   76.69   1st Qu.:109   FALSE:37        FALSE:22
Median :  208.01   Median :119   TRUE :47        TRUE :59
Mean   :  180.26   Mean   :120   NA's :1         NA's :4
3rd Qu.:  382.60   3rd Qu.:133
Max.   :  768.90   Max.   :155
NA's   :17         NA's   :24
```

The dataset contained many missing values. The pattern of the missing values is summarized on figure Figure 1. 48 patients had full data and the rest of the patients had between 1 and 4 missing data (number of red boxes per line). CATS, CAR, and neuroticm had a large number of missing data (11, 17, and 24) and this is why they were excluded from some analyses.

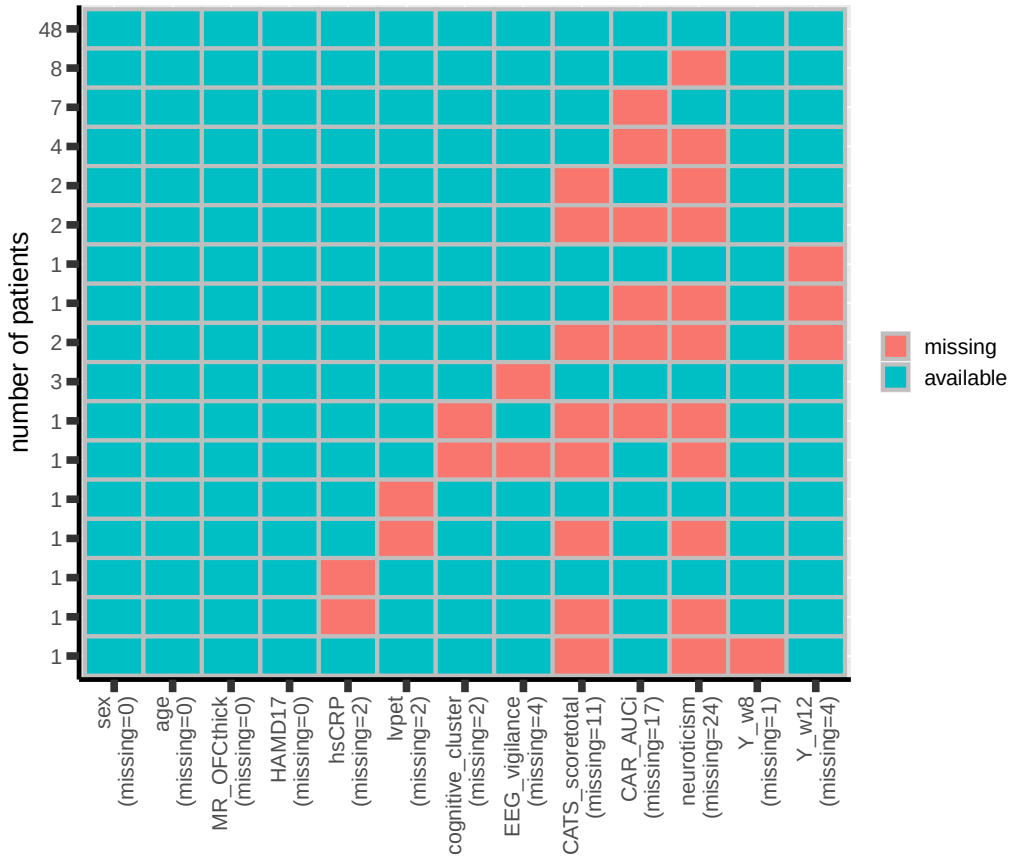

Figure 1: Missing data patterns

3.2 Association study (week 8)

Complete case: excluding CATS, CAR, and neuroticism, we fitted two logistic regressions (one with and one without the biomarkers) on the 75 patients with complete data. This likelihood ratio test showed evidence for an association between biomarkers and recovery:

```
Analysis of Deviance Table

Model 1: Y_w8 ~ female + age
Model 2: Y_w8 ~ female + age + MR_OFCthick + HAMD17 + low_hsCRP + lvpet +
    cognitive_cluster2 + cognitive_cluster3 + EEG_vigilance
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1        72    100.696
2        65     80.818  7   19.878  0.00584 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Looking at the biomarker specific effects, high vigilance appeared to be associated with poor recovery: odd ratio 0.179 (unit?) adjusted p-value of 0.047. There was also a similar trend for cognitive cluster 3: odd ratio 0.15, adjusted p-value of 0.07.

	estimate	std.error	odd ratio	p.value	adjusted p-value
(Intercept)	9.15328187	8.39872601	9.445388e+03	0.275782879	NA
female	-0.35932885	0.64097717	6.981447e-01	0.575073390	NA
age	0.05471804	0.04216889	1.056243e+00	0.194427297	NA
MR_OFCthick	-5.84569691	3.18471343	2.892318e-03	0.066424416	0.23752311
HAMD17	0.16179531	0.09514669	1.175620e+00	0.089040138	0.24208663
low_hsCRP	1.46954263	0.72733733	4.347246e+00	0.043337596	0.19466776
lvpet	-2.04888723	2.23330861	1.288782e-01	0.358921367	0.41535424
cognitive_cluster2	-0.81977323	0.69278301	4.405315e-01	0.236688505	0.41535424
cognitive_cluster3	-1.89634154	0.75932143	1.501168e-01	0.012510207	0.06993304
EEG_vigilance	-1.72062639	0.63876369	1.789540e-01	0.007066751	0.04675833

Multiple imputation: as a sensitivity analysis, we now used all patients and all biomarkers and use multiple imputations (100 datasets) to handle missing value. Results are rather similar to the complete case analysis, with a slight tendency for stronger effects.

```
summary(pool(e.glm_impw8))[,c(1,2,3,5:6)]
```

	term	estimate	std.error	df	p.value
1	(Intercept)	4.1516290163	8.6252902526	68.91234	0.631805406
2	sexfemale	-0.5825951700	0.6420666199	68.66383	0.367382059
3	age	0.0754554026	0.0456848728	68.79669	0.103163672
4	MR_OFCthick	-5.9114103612	3.0378756765	68.80810	0.055751906
5	HAMD17	0.1808561403	0.0925544076	68.24495	0.054794466
6	hsCRPlow	1.6868813295	0.7699177493	67.17518	0.031925680
7	lvpet	-2.6139562643	2.3770155497	68.85573	0.275301586
8	cognitive_cluster2	-1.1511850446	0.7226909239	68.23190	0.115801761
9	cognitive_cluster3	-2.7674590923	0.8726252184	68.19986	0.002273849
10	EEG_vigilance	-1.9950780957	0.6945417890	68.76199	0.005411043
11	CATS_scoretotal	-0.0001481392	0.0153985265	68.98892	0.992351959
12	CAR_AUCi	0.0007498367	0.0009340191	66.90661	0.424929161
13	neuroticism	0.0351547075	0.0211703226	68.99083	0.101339384

3.3 Association study (week 12)

Complete case: excluding CATS, CAR, and neuroticism, we fitted two logistic regressions (one with and one without the biomarkers) on the 72 patients with complete data. This likelihood ratio test showed no clear evidence for an association between biomarkers and recovery:

```
Analysis of Deviance Table

Model 1: Y_w12 ~ female + age
Model 2: Y_w12 ~ female + age + MR_OFCthick + HAMD17 + low_hsCRP + lvpet +
    cognitive_cluster2 + cognitive_cluster3 + EEG_vigilance
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1        69     76.256
2        62     64.376  7    11.88   0.1046
```

This was confirmed when looking at the biomarker specific effects. We can also see that the biomarkers for which with have most evidence against the null (`MR_OFCthick` and `cognitive_cluster2`) differ from week 8

```
                    estimate  std.error    odd ratio    p.value adjusted p-value
(Intercept)       15.1377108 9.66268612 3.751667e+06 0.11720467               NA
female            -0.1631019 0.71162765 8.495046e-01 0.81871691               NA
age                0.1239538 0.07958311 1.131964e+00 0.11934255               NA
MR_OFCthick       -8.2502682 3.58368566 2.611885e-04 0.02132536        0.1342289
HAMD17             0.1706117 0.10751351 1.186030e+00 0.11253840        0.4384615
low_hsCRP          1.0809026 0.83834457 2.947339e+00 0.19728347        0.5786588
lvpet             -0.5703845 2.60457709 5.653080e-01 0.82665539        0.9698902
cognitive_cluster2 -1.4448166 0.79614533 2.357893e-01 0.06956004        0.3344590
cognitive_cluster3 -0.8051151 0.88884433 4.470365e-01 0.36504179        0.7385351
EEG_vigilance     -0.0668419 0.60067967 9.353431e-01 0.91139660        0.9698902
```

Multiple imputation: as a sensitivity analysis, we now used all patients and all biomarkers and use multiple imputations (100 datasets) to handle missing value. Results are rather similar to the complete case analysis, but with a stronger evidence for an assocation between OFC thickness and recovery. Note that cognition and CATS are bordeline significant without adjustment for multiple comparisons.

```
summary(pool(e.glm_impw12))[,c(1,2,3,5:6)]
```

```
                term     estimate   std.error       df     p.value
1        (Intercept) 16.760516575 9.890355948 65.99623 0.094859430
2          sexfemale -0.898638556 0.731432568 65.79797 0.223597826
3                age  0.115967103 0.075427653 66.00500 0.128960677
4        MR_OFCthick -9.520173723 3.499858676 65.92820 0.008334242
5             HAMD17  0.141151055 0.103171990 65.81972 0.175928909
6          hsCRPlow  1.004339184 0.857512761 65.03615 0.245782198
7              lvpet -0.643880653 2.660297543 65.99493 0.809504882
8  cognitive_cluster2 -1.906920071 0.887507143 65.52557 0.035364314
9  cognitive_cluster3 -1.725228423 0.949783976 65.71480 0.073863526
10      EEG_vigilance -0.382219479 0.630217817 66.01354 0.546270922
11    CATS_scoretotal  0.037284123 0.020188426 66.02463 0.069257123
12           CAR_AUCi  0.001358667 0.001228576 65.78712 0.272803070
13       neuroticism  0.017883788 0.023653524 65.95409 0.452297597
```

3.4 Predictive value (week 8)

Complete case: excluding CATS, CAR, and neuroticism, we assessed the predictive performance of two logistic regressions (one with `glm_ccw8` and one without the biomarkers `glm0_ccw8`) as well as a random forest model (`rf_ccw8`) on the 75 patients with complete data:

```
     method metric     model   estimate
1  internal    auc glm0_ccw8 0.62769010
2  internal    auc  glm_ccw8 0.81133429
3  internal    auc   rf_ccw8 1.00000000
4  internal  brier glm0_ccw8 0.23870239
5  internal  brier  glm_ccw8 0.17485453
6  internal  brier   rf_ccw8 0.09276568
7        cv    auc glm0_ccw8 0.52305595
8        cv    auc  glm_ccw8 0.67530846
9        cv    auc   rf_ccw8 0.55341822
10       cv  brier glm0_ccw8 0.25887838
11       cv  brier  glm_ccw8 0.23932881
12       cv  brier   rf_ccw8 0.25372824
```

After cross-validation, we observe that both the AUC and brier score of the random forest (with biomarkers) are similar to the logistic regression without biomarkers. This indicates poor predictive ability of the random forest that will not be considered further. The logistic model with biomarker has a higher AUC (i.e., better discrimination) and lower brier score (i.e., smaller discrepancy between prediction and observed outcome) compared to the logistic model without biomarkers. The permutation test confirmed that the logistic model with biomarkers was informative (p=0.02 for the AUC and p=0.01 for the brier score ⚠ to re-run with more permutations) while there was no clear evidence with the logistic model without covariates (p=0.11 for the AUC and p=0.27 for the brier score). This difference between the predictions from the models (after cross validation) is illustrated in Figure 2, as well as the corresponding ROC Figure 3 and calibration curves Figure 4.

Note that the average AUC estimated by the permutation test was 0.5, supporting that the proposed cross-validation procedure is unbiased (under the null).

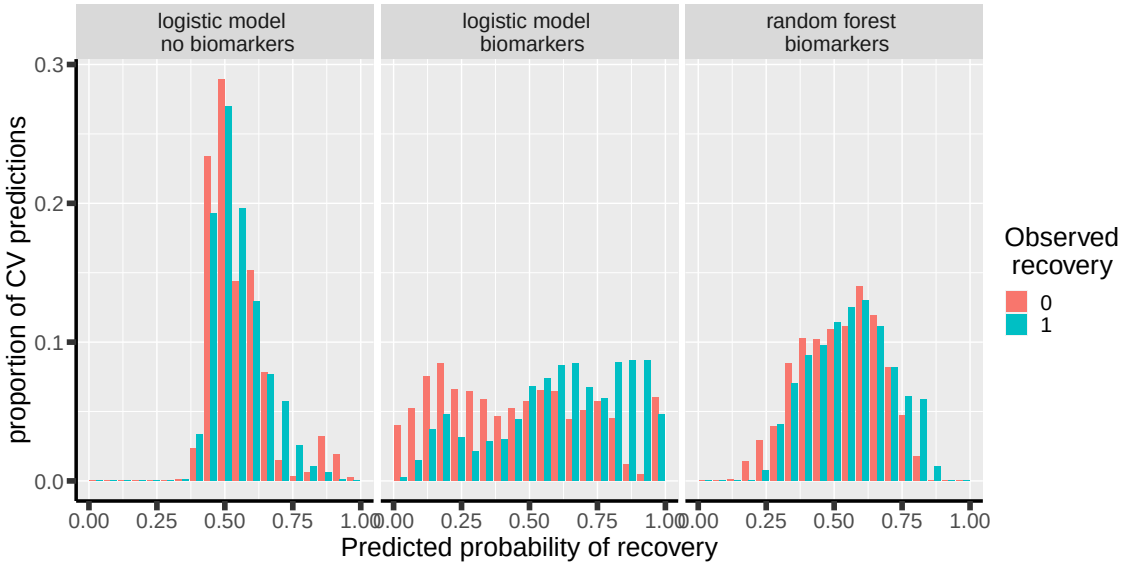

Figure 2: Distribution of the predicted probability of recovery according to the actual recovery for the various predictive models for week 8.

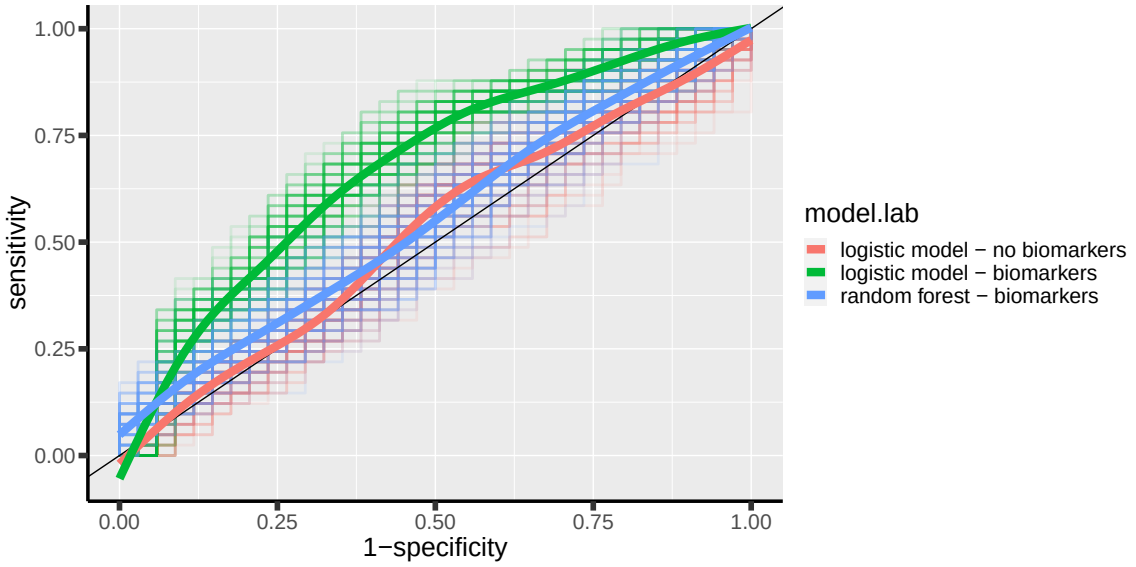

Figure 3: Roc curve associated to the cross-validated predictions for the various models (thick lines) for week 8. It is obtained by applying a smoother (lowess) on the 100 ROC curve obtain for each model and each of the 100 repetitions of the 10 fold cross validations (thin lines).

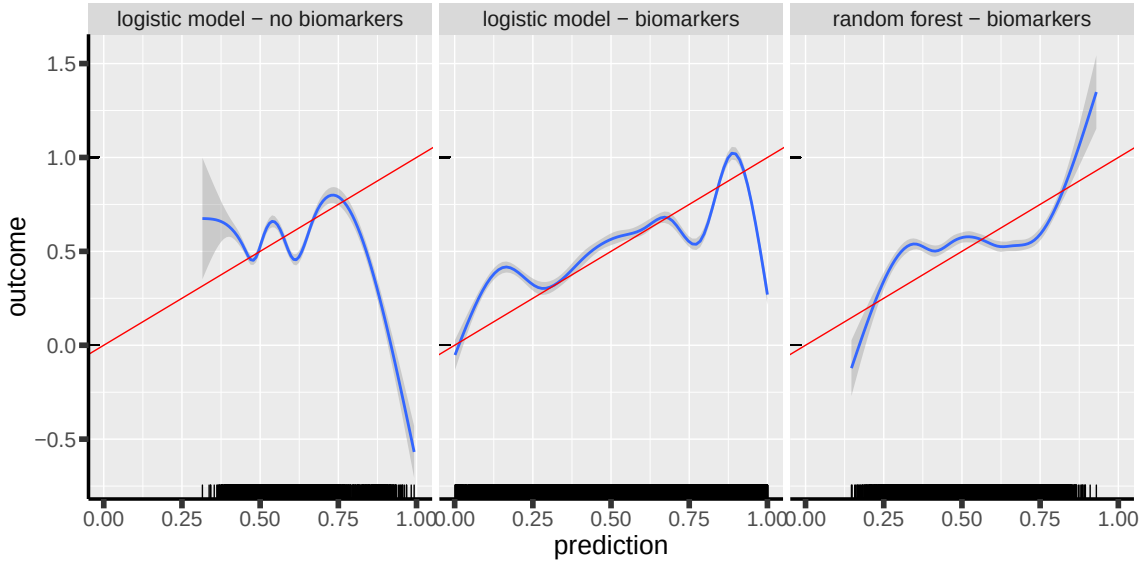

Figure 4: Calibration curve associated to the cross-validated predictions for the various models (thick lines) for week 8.

Full data: as a sensitivity analysis, we now used all patients and all biomarkers and modified the cross-validation procedure to handle missing data. We obtain slightly different results, but still in favor of the logistic model with biomarkers. The difference in AUC between the logistic models is similar to previously (about +0.1) but now the brier score is worse (+0.4 instead of 0.3) indicating poor calibration.

```
     method metric   model  estimate
1: internal    auc glm0_w8 0.5991949
2: internal    auc  glm_w8 0.8878666
3: internal  brier glm0_w8 0.2409137
4: internal  brier  glm_w8 0.1432597
5:       cv    auc glm0_w8 0.5574226
6:       cv    auc  glm_w8 0.6542563
7:       cv  brier glm0_w8 0.2581057
8:       cv  brier  glm_w8 0.3015577
```

3.5 Predictive value (week 12)

Complete case: excluding CATS, CAR, and neuroticism, we assessed the predictive performance of two logistic regressions (one with `glm_ccw12` and one without the biomarkers `glm0_ccw12`) as well as a random forest model (`rf_ccw12`) on the 72 patients with complete data:

```
     method metric      model   estimate
1  internal    auc glm0_ccw12 0.66037736
2  internal    auc  glm_ccw12 0.79443893
3  internal    auc   rf_ccw12 1.00000000
4  internal  brier glm0_ccw12 0.18160078
5  internal  brier  glm_ccw12 0.14937617
6  internal  brier   rf_ccw12 0.05981177
7        cv    auc glm0_ccw12 0.57554121
8        cv    auc  glm_ccw12 0.60582920
9        cv    auc   rf_ccw12 0.76768620
10       cv  brier glm0_ccw12 0.19544784
11       cv  brier  glm_ccw12 0.21388319
12       cv  brier   rf_ccw12 0.16378866
```

After cross-validation, we observe that both the AUC and brier score of the logistic regression with biomarkers are similar to the logistic regression without biomarkers. This time it the random forest approach that shows a higher AUC (i.e., better discrimination) and lower brier score (i.e., smaller discrepancy between prediction and observed outcome) compared to the logistic model without biomarkers. The permutation test confirmed that the logistic model with biomarkers was informative (p=0.01 for the AUC and p=0.01 for the brier score ⚠ to re-run with more permutations) while there was no clear evidence for the other logistic models (p>0.1 for the AUC and brier score). This difference between the predictions from the models (after cross validation) is illustrated in Figure 5, as well as the corresponding ROC Figure 6 and calibration curves Figure 7.

Note that the average AUC estimated by the permutation test was 0.5, supporting that the proposed cross-validation procedure is unbiased (under the null).

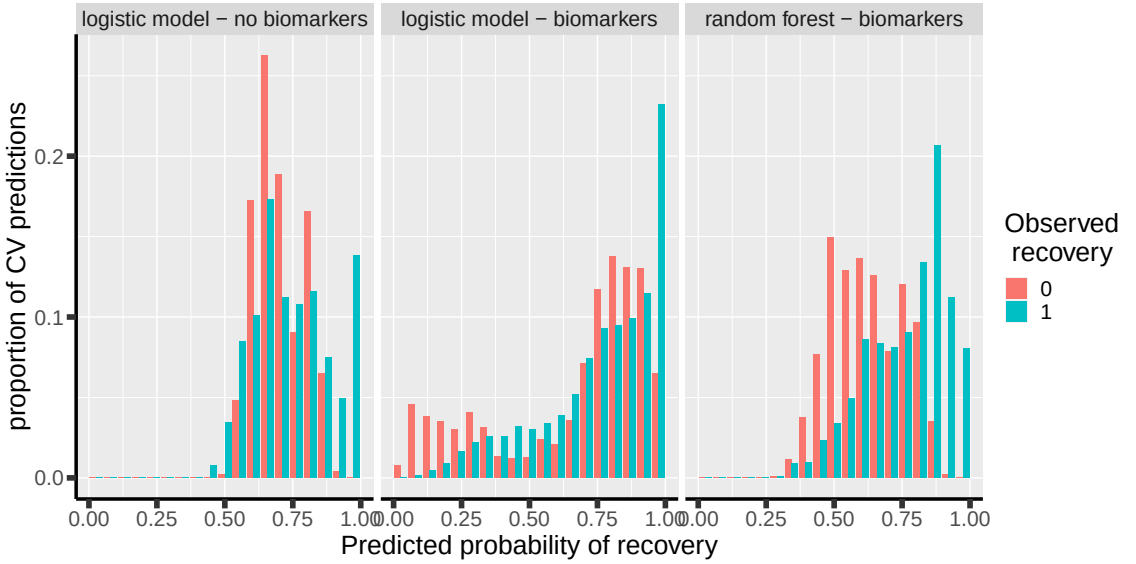

Figure 5: Distribution of the predicted probability of recovery according to the actual recovery for the various predictive models for week 12.

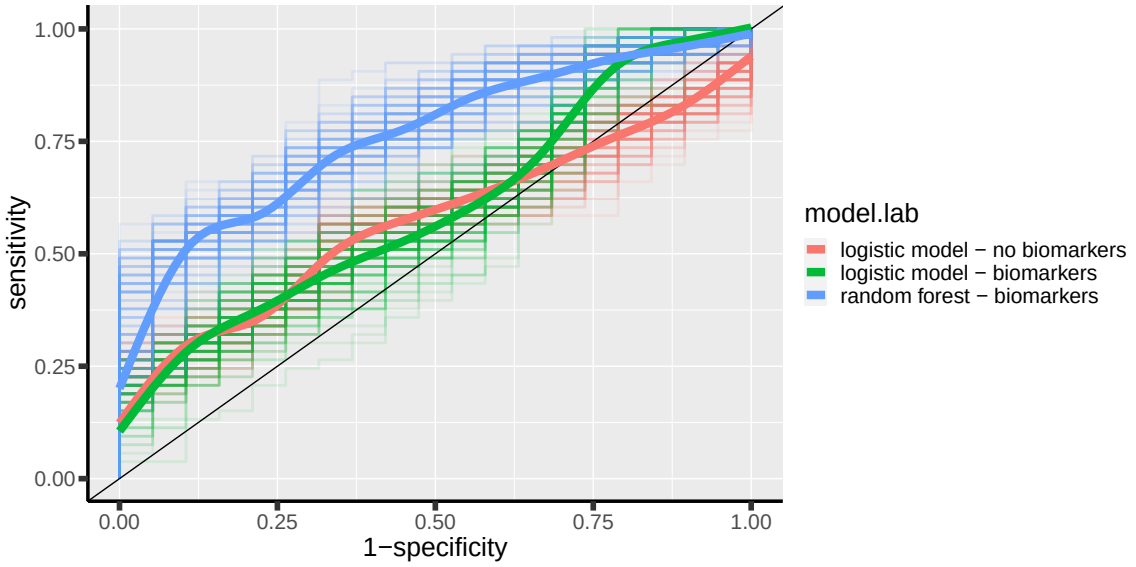

Figure 6: Roc curve associated to the cross-validated predictions for the various models (thick lines) for week 12. It is obtained by applying a smoother (lowess) on the 100 ROC curve obtain for each model and each of the 100 repetitions of the 10 fold cross validations (thin lines).

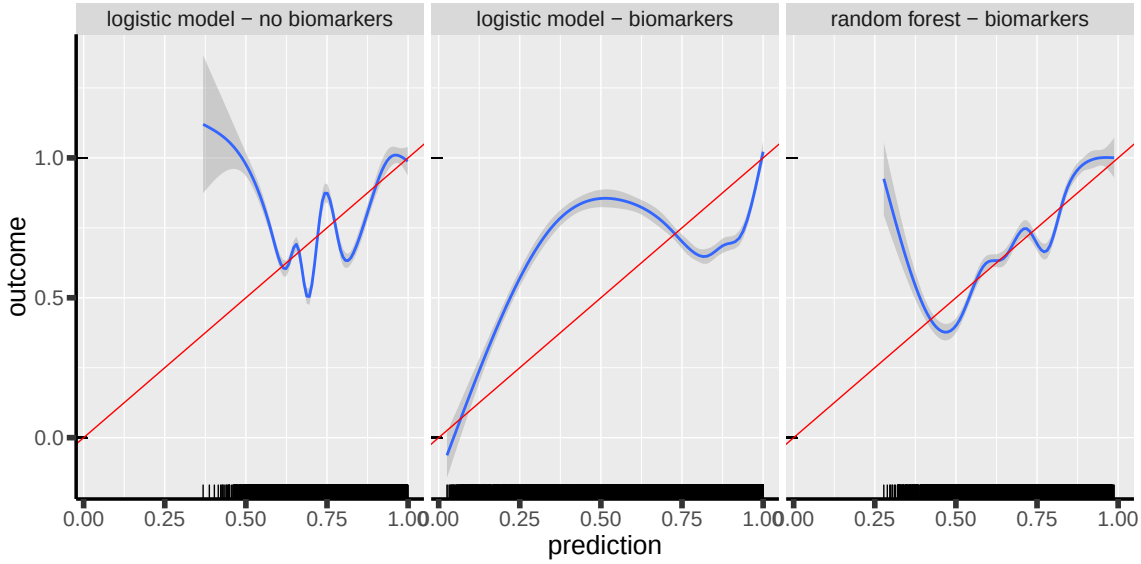

Figure 7: Calibration curve associated to the cross-validated predictions for the various models (thick lines) for week 12.

Full data: as a sensitivity analysis, we now used all patients and all biomarkers and modified the cross-validation procedure to handle missing data. The results are in line with the complete case, with worse performances when adding the biomarkers.

```
     method metric    model  estimate
1: internal    auc glm0_w12 0.6502311
2: internal    auc  glm_w12 0.8936826
3: internal  brier glm0_w12 0.1855812
4: internal  brier  glm_w12 0.1172801
5:       cv    auc glm0_w12 0.6017917
6:       cv    auc  glm_w12 0.5762885
7:       cv  brier glm0_w12 0.2181458
8:       cv  brier  glm_w12 0.2679212
```

4 Conclusion

There is evidence that some biomarkers (EEG, to a lesser extend cognition) are predictive of recovery at week 8. The corresponding gain in AUC was about +0.1 with a small improvement in brier score. These results were not seen at week 12 and the overall predictive performance was not great though (AUC of about 0.6 and brier score >0.2).

There was no evidence for non-linear effect or interaction between biomarkers (as assessed via a random forest model) probably due to the limited sample size.

Generally the results where robust to how missing data were handled (complete case or multiple imputation). Effects had a slight tendency to be stronger when not using complete case. The only exception is for the brier score at w8 which became worse compared to the complete case.

5 References

Moons, K. G., Donders, R. A., Stijnen, T., and Harrell Jr, F. E. (2006). Using the outcome for imputation of missing predictor values was preferred. *Journal of clinical epidemiology*, 59(10):1092–1101.

Van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of statistical software*, 45:1–67.