

## **Supplementary Information**

### **Brain-cognition relationships in treatment-resistant late-life depression**

Peter Zhukovsky<sup>#</sup>, PhD, Meryl A. Butters, PhD, Helen Lavretsky, MD, Patrick Brown, PhD, Joshua S. Shimony, MD, PhD, Eric J Lenze, MD, Daniel M Blumberger, MD, Alastair J Flint, MD, Jordan Karp, MD, Steven Roose, MD, Erin W Dickie, PhD, Daniel Felsky, PhD, Ginger E Nichol, MD, Yar Al-Dabagh, Nicole Schoer, Feyifolu Obiri, Ashlyn Runk, Kayla Conaty, Benoit H Mulsant, MD, Aristotle N Voineskos<sup>#</sup>, MD, PhD

<sup>#</sup>Please address correspondence to Dr Aristotle N Voineskos or Dr Peter Zhukovsky, Center for Addiction and Mental Health, 250 College Street, Toronto, ON, Canada

#### **Overview**

1. MRI data acquisition
2. Participant inclusion flow
3. Timing of the neuropsychological and clinical assessments relative to the OPTIMUM clinical trial
4. Seed-based FC analysis probing precuneus connectivity
5. Sensitivity PLS regression analyses of FC-cognition associations
6. Sensitivity analyses of hold-out data by scanner
7. Markers predicting treatment outcomes
8. Clinical models predicting treatment outcomes
9. Stability of structural MRI and cognitive function between baseline and 6 month follow up
10. Brain-cognition associations in subsets of participants and after controlling for cognitive status

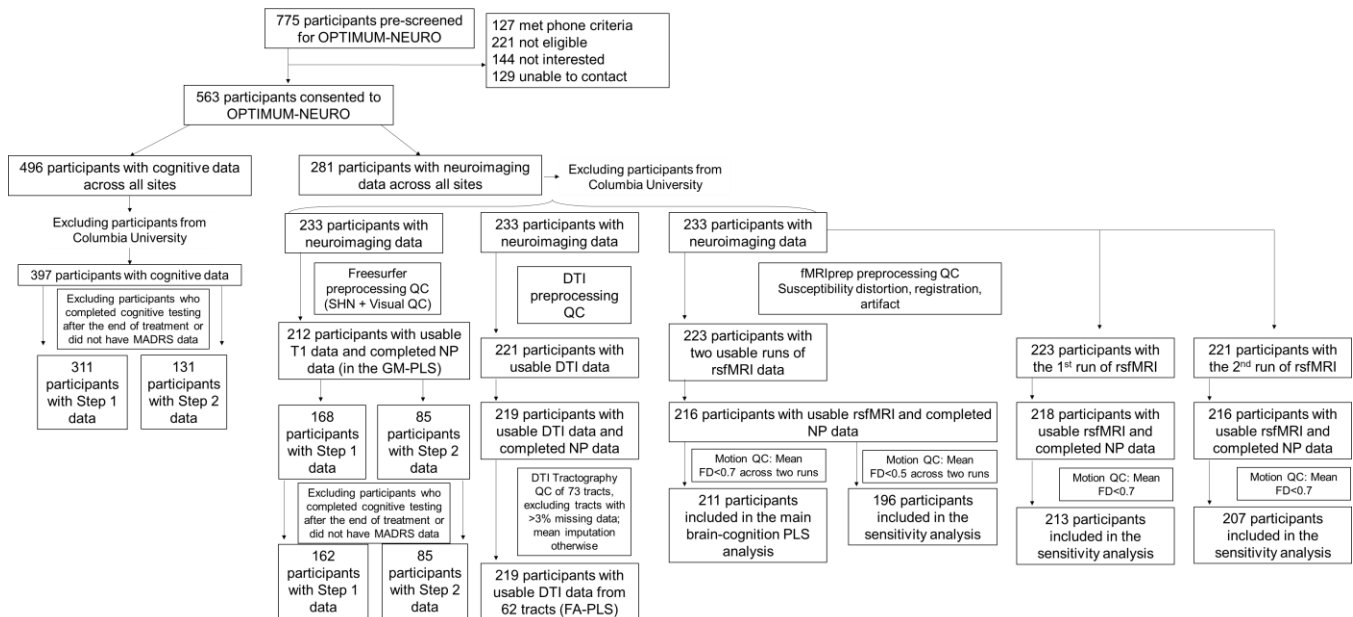
## 1. MRI data acquisition

MRI data were acquired using harmonized acquisitions for 3T Siemens Prisma (Pitt, UCLA, Wash U.) and 3T GE MR750 (CAMH/Toronto, Columbia) scanners. The acquisition parameters have been optimized for high quality T1-weighted, diffusion MRI (dMRI) and rs-fMRI data aiming to ensure that key parameters including voxel size, number of diffusion-weighted directions, and flip angle are consistent across sites. We also include multi-band acquisitions for echo planar imaging (BOLD and dMRI). These parameters have previously undergone reliability testing and rigorous quality assessments (QA) through the ABCD study. We implemented standard QA protocols at each site. In addition to sequences listed in the Supplementary Table 1, two-minute field maps with reverse phase encoding directions were collected for dMRI and fMRI scans to correct for anatomic distortions.

**Supplementary Table 1.** Overview of the acquisition parameters.

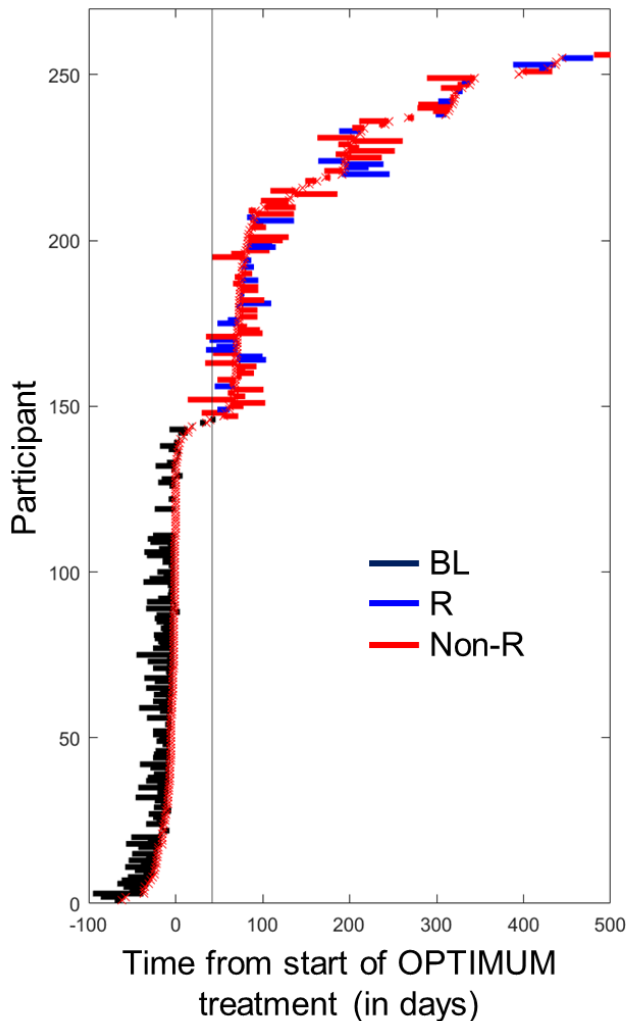
|           | Matrix    | Slices | FOV       | % FOV phase | Resolution (mm) | TR (ms) | TE (ms)    | TI (ms) | Flip Angle (deg) | Parallel Imaging | MultiBand Acceleration | Phase partial Fourier | Diffusion Directions | b-values mm <sup>2</sup> /s                                        | Acquisition Time (min) |
|-----------|-----------|--------|-----------|-------------|-----------------|---------|------------|---------|------------------|------------------|------------------------|-----------------------|----------------------|--------------------------------------------------------------------|------------------------|
|           |           |        |           |             |                 |         | GE Siemens |         |                  |                  |                        |                       |                      |                                                                    | GE Siemens             |
| T1        | 256 x 256 | 208    | 256 x 256 | 100%        | 1.0 x 1.0 x 1.0 | 2500    | 2 2.88     | 1080    | 8                | 2x               | Off                    | Off                   | N/A                  | N/A                                                                | 6:09 7:12              |
| T2        | 256 x 256 | 208    | 256 x 256 | 100%        | 1.0 x 1.0 x 1.0 | 3200    | 60 56.5    | N/A     | variable         | 2x               | Off                    | Off                   | N/A                  | N/A                                                                | 5:50 6:35              |
| Diffusion | 140 x 140 | 81     | 240 x 240 | 100%        | 1.7 x 1.7 x 1.7 | 4100    | 81.9 88    | N/A     | 77               | Off              | 3                      | 5.5/8                 | 102                  | 500 (6-dirs)<br>1000 (15-dirs)<br>2000 (15-dirs)<br>3000 (60-dirs) | 7:30 7:30              |
| rsfMRI    | 90 x 90   | 61     | 216 x 216 | 100%        | 2.4 x 2.4 x 2.4 | 800     | 30 30      | N/A     | 52               | Off              | 6                      | Off                   | N/A                  | N/A                                                                | 2 x 10:00 2 x 10:00    |

## 2. Participant inclusion flow



**Supplementary Figure 1. Overview of the quality control (QC) and exclusion process for the structural, diffusion, and rs-fMRI analyses.** Surface hole number was taken from the aseg file generated by the FreeSurfer pipeline. Visual quality checks were conducted on images with higher numbers of surface holes (>220) across the two hemispheres and images were included or excluded or included based whenever FreeSurfer reconstruction either missed a larger cortical region or if systematic misalignment was found between the observed and the reconstructed gray and white matter boundary.

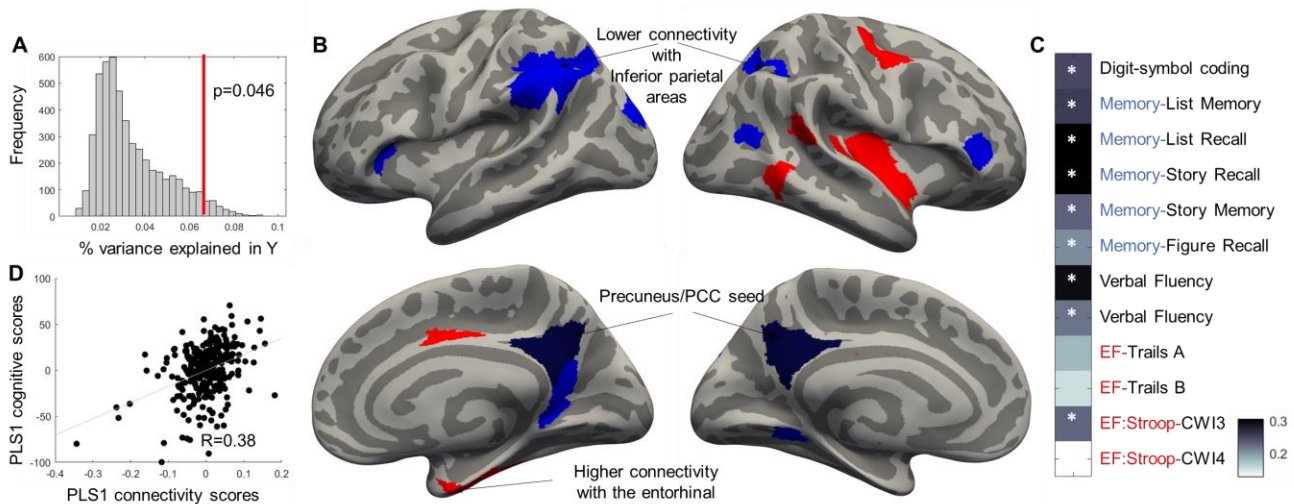
### 3. Timing of the neuropsychological and clinical assessments relative to the OPTIMUM clinical trial



**Supplementary Figure 2. Timing of the neuropsychology (NP) assessment relative to the start of treatment and the corresponding MADRS assessment.** For each participant, each row shows the timing of the MADRS score (indicated by a red cross) that occurred closest to the 'baseline' (BL) NP assessment and the corresponding NP assessment itself, connected by a line. Non-remitted participants (non-R) who were part of the OPTIMUM trial for over 6 weeks are shown in red; remitted participants (R) who were part of the OPTIMUM trial for over 6 weeks are shown in blue, while the participants scanned at the start of the OPTIMUM trial (BL) are shown in black. Previous data show that 4 weeks may not be sufficient to assess efficacy in non-treatment resistant depression<sup>1</sup>. In addition the Antidepressant Treatment Response Questionnaire requires an adequate monotherapy trial to be at least 6 weeks long<sup>2</sup>. Participant data was sorted according to the delay between the time of the MADRS assessment relative to the start of treatment.

### 4. Seed-based FC analysis probing precuneus connectivity

For seed-based analyses, we correlated the average timeseries of the bilateral precuneus region (consisting of the following ROIs from the HCP parcellation: L\_7m\_ROI, L\_v23ab\_ROI, L\_d23ab\_ROI, L\_31pv\_ROI, L\_31pd\_ROI, R\_7m\_ROI, R\_v23ab\_ROI, R\_d23ab\_ROI, R\_31pv\_ROI, R\_31pd\_ROI) with 360 ROIs in the HCP parcellation. To reduce the number of predictor variables, we collapsed across hemispheres, resulting in 180 bilateral connectivity values from the precuneus seed. A 215x180 predictor matrix was entered into a PLS regression, alongside the 215x12 outcome matrix with cognitive data. As in the main analyses, rs-fMRI timeseries were concatenated across two runs and participants with mean FD>0.7 were excluded. The PLS regression with two components explained 6.5% of variance in the cognitive data, significantly more than expected by chance ( $p=0.046$ ). Bootstrapping identified several regions whose connectivity with the precuneus predicted better cognitive performance, primarily involving memory, verbal fluency and processing speed tests (Supplementary Figure 3).

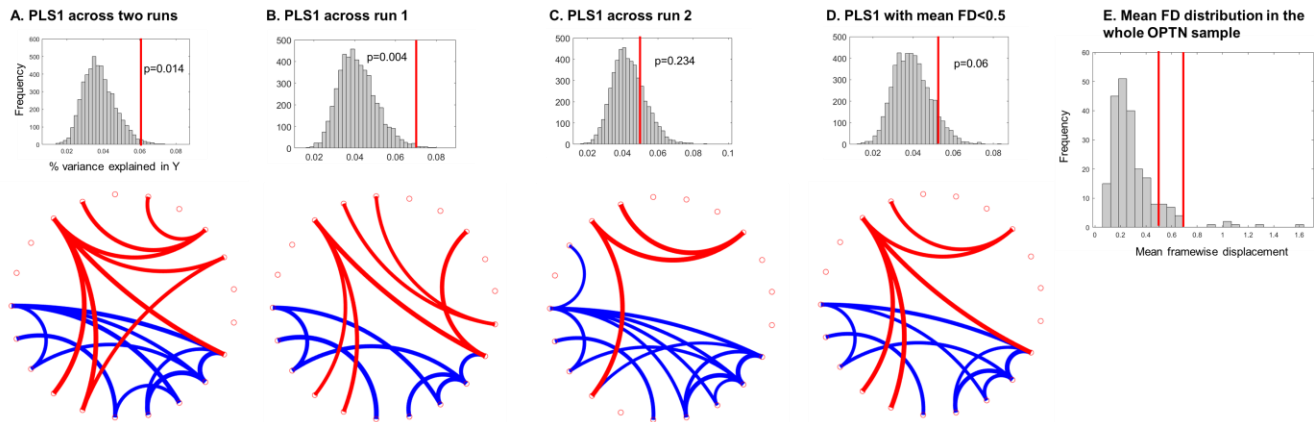


**Supplementary Figure 3. Seed-based precuneus connectivity analyses.** Seed-based connectivity analyses of the posterior medial DMN (pmDMN including precuneus and posterior cingulate cortex, PCC) showed that lower connectivity of the pmDMN with the inferior parietal cortex and some inferior frontal areas (part of frontoparietal network) and visual areas was associated with better cognition across the board. Higher connectivity of the pmDMN with the entorhinal and insular cortex was also associated with better cognition, especially in the memory domain. The overall PLS explained significantly more variance than expected by chance (A,D). PLS weights  $Z>3$  are shown in red and  $Z<-3$  are shown in blue (B), while Pearson's correlation between PLS1 X-scores and cognitive tests are shown in (C), with asterisks indicating significance after Bonferroni multiple comparison correction. Precuneus seed is shown in dark blue.

## 5. Sensitivity PLS regression analyses of FC-cognition associations

We conducted three sensitivity analyses. Given that two distinct runs of resting-state fMRI data were collected at the start and at the end of the scanning protocol, we repeated the FC-cognition PLS in each run separately. The overall PLS model focusing on run 1 explained significantly more variance in cognitive function than expected by chance ( $p=0.004$ ), and included many of the connectivities captured by the main model ( $|Z|>3$ , Supplementary Figure 4B). While the overall PLS model focusing on run 2 did not explain more variance than expected by chance ( $p=0.244$ ), bootstrapping revealed that many of the connectivities captured by the main model were also captured in this PLS model ( $|Z|>3$ , Supplementary Figure 4C).

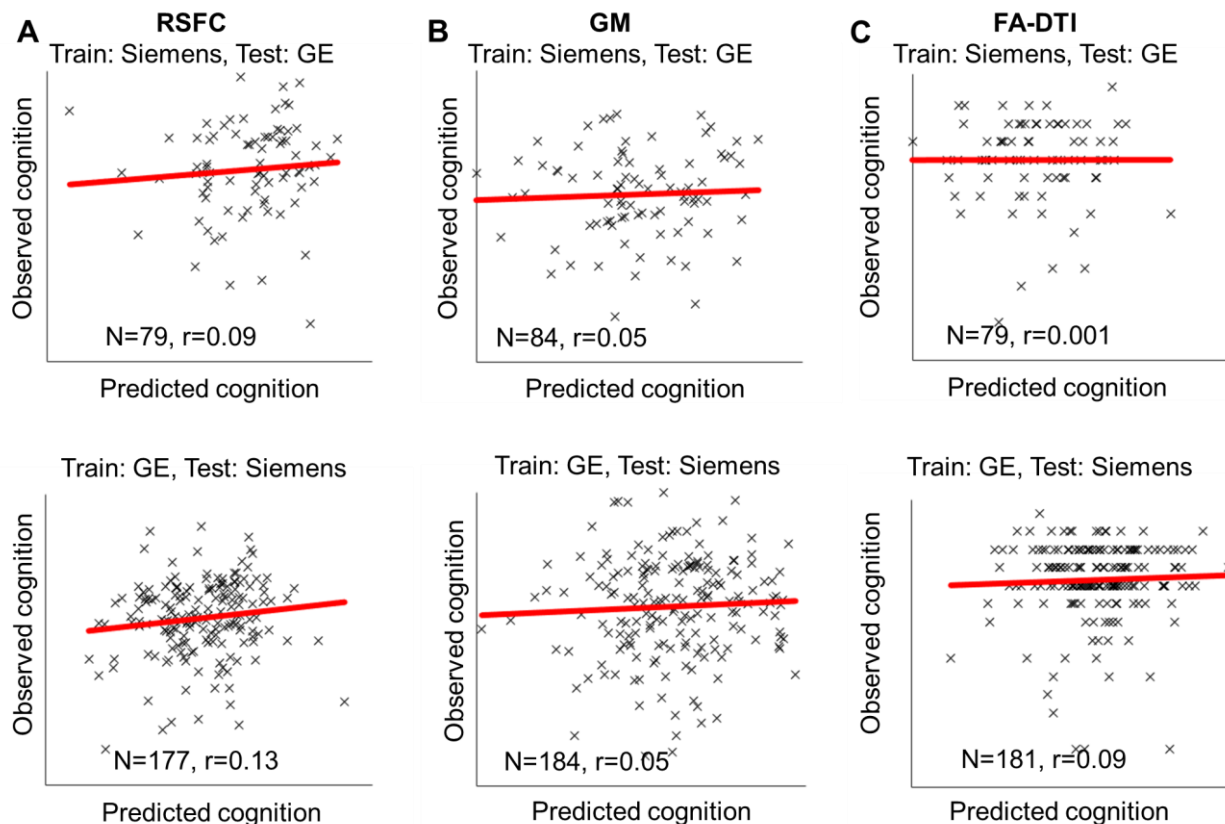
Further, we conducted another PLS analysis excluding participants with mean FD $>0.5$ . This PLS model did not explain more variance than expected by chance ( $p=0.06$ ); however, bootstrapping revealed that many of the connectivities captured by the main model were also captured in this PLS model ( $|Z|>3$ , Supplementary Figure 4D).



**Supplementary Figure 4. Sensitivity analyses of rsFC-cognition.** Sensitivity analyses showing FC-cognition associations for each of the two runs of rs-fMRI data (B,C) acquired and with a stricter FD threshold of 0.5 (D) alongside the main analysis (A). Motion distribution across all participants, including the 0.6 and 0.7 thresholds (E). FD: framewise displacement.

## 6. Sensitivity analyses of hold-out data by scanner

In order to test whether differences in scanner types used affect prediction accuracy of the brain-cognition PLS models, we have split our sample according to the scanner type used (GE vs Siemens Prisma) and trained a PLS model on data from one scanner and tested it on data from the other scanner. The resulting prediction performance was notably worse than prediction performance in the held-out data by site, especially for gray and white matter analyses. Correcting for batch effects resulting from the scanner rather than site slightly improved prediction accuracy in held out scanner data for the DTI PLS only.



**Supplementary Figure 5. Hold-out results by scanner.** Training the PLS model on data from one scanner type and testing its performance on data from the other scanner type (i.e. Siemens Prisma 3T and General Electric (GE) 3T scanners) has resulted in worse performance compared with testing the held-out data by each of the five sites (Figure 2, 3, 4) for resting state FC (A), gray matter (B) and white matter (C) analyses. GM: gray matter; RSFC: resting-state functional connectivity; FA-DTI: fractional anisotropy – diffusion tensor imaging.

## 7. Markers predicting treatment outcomes

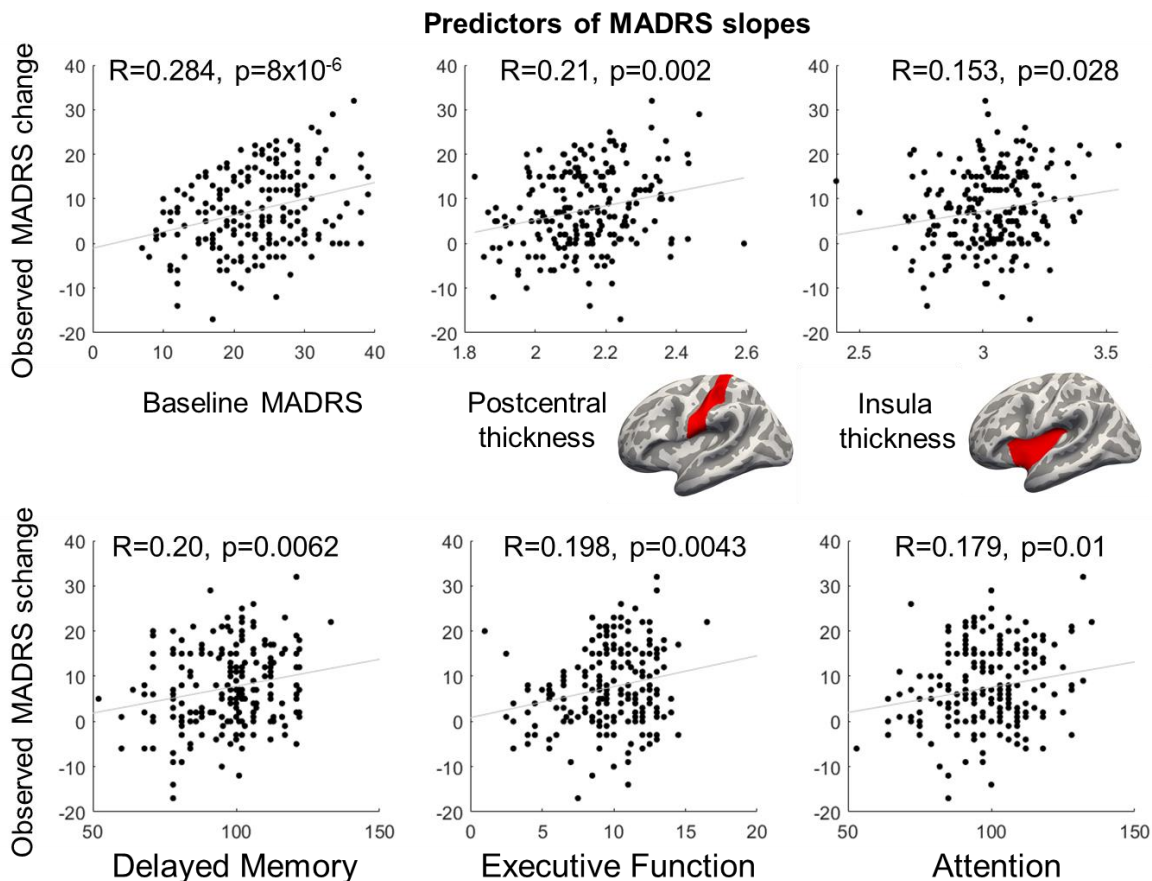
We created several sets of elastic net regression models aiming to predict treatment outcomes using demographic (age, sex), clinical (baseline MADRS), and neuroimaging data. The first set of models was ran in the subset of participants with gray matter data, i.e. including FreeSurfer derived cortical thickness variables in the 68 regions from the Desikan-Killiany parcellation<sup>3</sup> and lateralized volumes of the hippocampus, amygdala and striatum (accumbens, caudate and putamen) from the *aseg* parcellation, normalized for total intracranial volume. The second set of models was run in participants with resting state functional connectivity data, while the third set was ran in participants with fractional anisotropy data. We predicted response to step 1 and step 2 separately given the different classes of drugs used in each step. In order to assess prediction performance, we ran 100 elastic models for each model set. On each of these 100 iterations, we split the sample into training and testing subsets, randomly selecting 20 participants in the testing subset for step 1 and 20 participants in the testing subset in step 2, while using the rest of the sample for training the model. Elastic net regressions represent a compromise between lasso and ridge regressions. Lasso regressions take a conservative regularization approach that sets the regression coefficient for any predictors that don't contribute enough to the model to zero. Ridge regression takes a more liberal approach to regularization and shrinks the regression coefficient to very small values. A balance between lasso and ridge regularization was achieved by setting the alpha parameter to 0.1 (i.e. biased towards ridge regularization) in the *lasso glm* (MATLAB R2022) logistic regression model. On each of the 100 iterations, the regression model used 10-fold cross-validation to optimize the regression weights, setting a number of those predictors to zero. For each of the 100 models in step 1, we obtained an AUC value, visualizing the mean AUC ( $\pm 95\%$ CI) in Figure 5C. Similarly, for each of the 100 models in step 2, we obtained an AUC value, visualizing the mean AUC ( $\pm 95\%$ CI) in Figure 5G. We repeated this process while first, excluding the MRI data for each set of models, and second, excluding both the MRI data and the cognitive data, and used the resulting AUCs to assess the contributions from MRI, cognitive, and clinical/demographic data. Finally, we selected the most parsimonious model by only including predictors that survived regularization (i.e. those with non-zero values) in  $\geq 95$  of the 100 models. We assess and present the performance of this most parsimonious model in an 8-fold split of the step 1 data and a 5-fold split of the step 2 data. For each of the folds, we set the regularization parameter alpha to 0.001 aiming to include all variables and only tune their relative importance to the model. We then assess the AUCs and present a confusion matrix for the step 1 and step 2 data in Figure 5D and 5H, respectively.

Excluding participants with their neuropsychological testing completed over six weeks after start of treatment did not substantially affect the prediction performance of the elastic net models.

In addition, we operationalize treatment outcomes as change in MADRS scores throughout the OPTIMUM trial (regardless of step) and conduct a PLS regression to predict the change in MADRS scores using demographic, clinical, and gray matter data. The results of the PLS model also implicate better cognitive function (including Trails B) and higher cortical thickness of the precentral and rostral ACC as predictors of improvement in MADRS scores due to treatment. We show the associations between change in MADRS scores and these baseline variables in Supplementary Figure 6.

## 8. Clinical and cognitive models predicting treatment outcomes

We compared the predictive model results in the sample of participants with MRI and clinical and cognitive data to findings in the larger set of participants with clinical and cognitive data, regardless of whether they also had MRI data. We found consistent levels of performance in the full clinical/cognitive sample when including baseline MADRS, and cognitive data. For step 1 (n=290), across folds AUC mean=0.678, SD=0.14. For step 2 (n=152), across all folds AUC mean=0.568, SD=0.13. These values are consistent with the model performance observed in the subsample with clinical only data and in the subsample with both neuropsychological and MRI data (Figure 5).

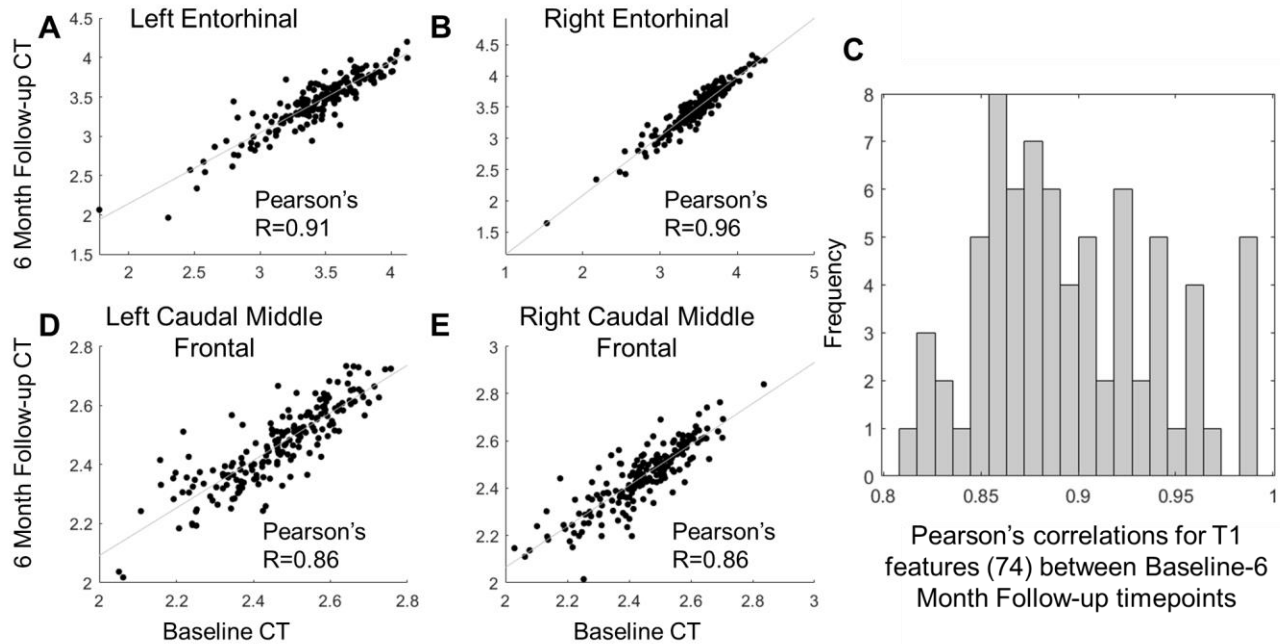


**Supplementary Figure 6. Predictors of MADRS change.** Scatterplots showing the association between baseline MADRS, cognitive function, and cortical thickness and change in MADRS scores pre vs post treatment. We show select variables that were significant in the PLS regression linking these data types with change in MADRS scores. MADRS: Montgomery-Asberg Depression Rating Scale. R values refer to Pearson's R correlations.

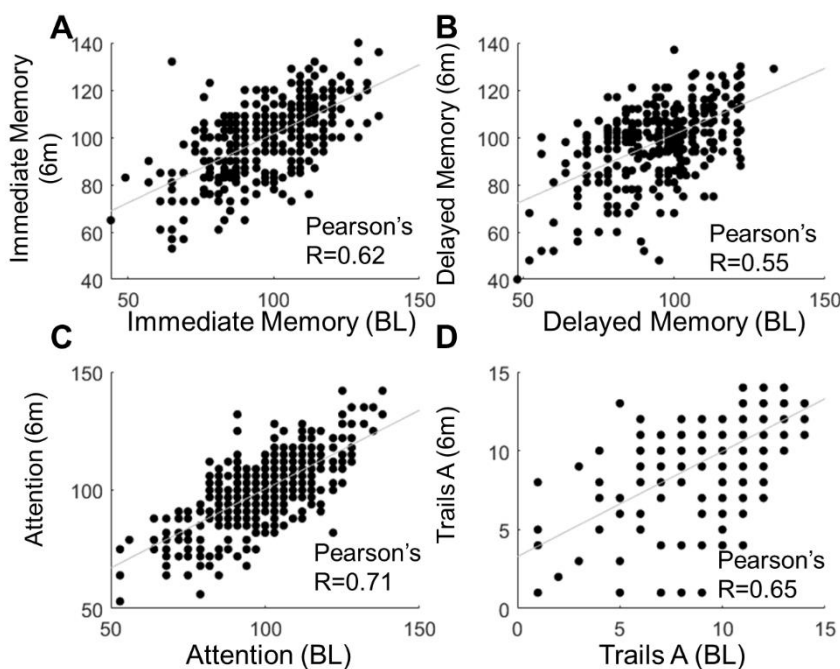
## 9. Stability of structural MRI and cognitive function between baseline and 6 month follow up

To provide a justification for using structural MRI features that were potentially collected several weeks after the start of the clinical trial (Supplementary Figure 2) in predicting prediction clinical trial outcomes, we leverage the 6-month follow-up data from OPTIMUM-NEURO. Specifically, we correlate baseline cortical thickness and subcortical volume values with those collected at 6 months, after longitudinal FreeSurfer processing on top of fMRIPrep. We found remarkable consistency of baseline-6 month

correlations across all structural features investigated (Supplementary Figure 7), with low within-subject variability across the 6-month follow-up period. This data supports the use of structural features ascertained even several months after the start of the clinical trial. Further, we found good consistency in cognitive performance between baseline and 6-month timepoints (Supplementary Figure 8), supporting the use of cognitive tests in predicting remission.



**Supplementary Figure 7. Baseline-to-6-month follow-up correlations between structural features.** Left and right entorhinal (A, B), and left and right caudal middle frontal gyri (D,E) were chosen as representative examples from medial temporal lobe and the prefrontal cortex. Distribution of all baseline-6-month correlations is shown in (C), with values ranging between 0.8 and 0.99. CT: cortical thickness. All  $p < 0.0001$ . CT: cortical thickness.

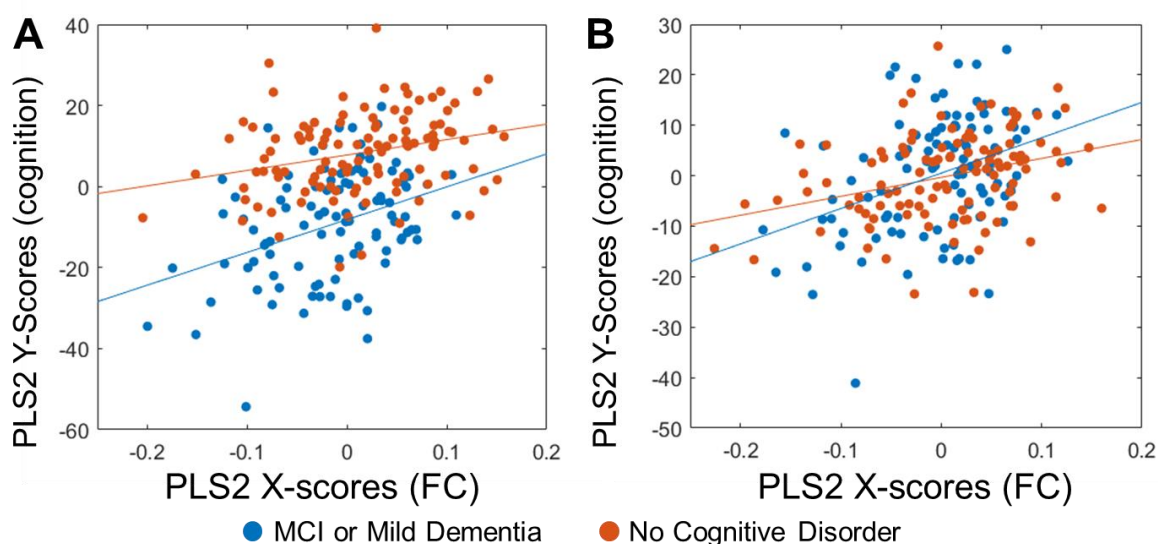


**Supplementary Figure 8. Within-subject consistency of cognitive scores over a 6-month follow-up.** Baseline-to-6-month follow-up correlations between key cognitive tests used in the analyses. Immediate and delayed memory (A, B), attention (C) and Trailmaking A (panel D) showed a high degree of consistency between baseline and 6 month. All  $p < 0.0001$ . BL: baseline; 6m – 6 month follow-up.

## 10. Brain-cognition associations in subsets of participants and after controlling for cognitive status

We next tested whether our findings in the overall OPTIMUM sample hold up in subsets of participants with MCI or mild dementia, and in those with no cognitive disorder. For the resting-state FC models, we found that in each subgroup individually, the overall PLSR was not significant, i.e. did not explain more variance in cognitive function compared to chance permutation models (cognitively normal subjects:  $n=112$ ,  $p_{\text{PERMUTATION}}=0.74$  for a 3-latent variable model; MCI and dementia subgroup:  $n=97$ ,  $p_{\text{PERMUTATION}}=0.36$  for a 3-latent variable model). For the diffusion-weighted fractional anisotropy models, we found that in each subgroup individually, the overall PLSR was not significant (cognitively normal subjects:  $n=109$ ,  $p_{\text{PERMUTATION}}=0.30$  for a 1-latent variable model; MCI and dementia subgroup:  $n=103$ ,  $p_{\text{PERMUTATION}}=0.12$  for a 1-latent variable model). For the T1 cortical thickness and subcortical volume models, we found that in each subgroup individually, the overall PLSR was not significant (cognitively normal subjects:  $n=113$ ,  $p_{\text{PERMUTATION}}=0.28$  for a 3-latent variable model; MCI and dementia subgroup:  $n=103$ ,  $p_{\text{PERMUTATION}}=0.54$  for a 3-latent variable model). This suggests that our brain-cognition findings could be driven by the variability in cognitive function across the two patient subgroups and reduced power to detect effects in the smaller cognitive subgroups. Greater variability in brain data and cognitive/behavioral performance drives larger effect sizes<sup>4</sup> and leads to significant findings in the whole sample.

Finally, we repeated the brain-cognition PLS regressions after controlling for the “no cognitive disorder” vs MCI/Mild dementia group in the brain connectivity data using ComBat and after regressing out cognitive status from the cognitive outcome variables. The resting-state FC PLS remained significant (permutation  $p=0.0286$ ), similar to the DTI PLS (permutation  $p=0.0386$ ) and the CT PLS (permutation  $p=0.0138$ ) after controlling for cognitive status in the brain data. However, regressing out group from the cognitive outcomes Y did affect the models as shown in Supplementary Figure 9. Specifically, we found the variability in cognitive outcomes Y to be reduced after regressing out cognitive status, leading to lower overall PLSR correlation of 0.33 vs 0.37.



**Supplementary Figure 9. Group-specific analyses of brain-cognition associations.** Latent variable correlation in the NCD (no cognitive disorder) and the MCI or Mild Dementia groups for the FC PLS ran across all participants before (A) and after (B) regressing out group from the cognitive outcome data.

We found a significant correlation between latent brain and cognitive scores from the resting-state FC PLS regression when not regressing out group (**A**, Pearson's  $R=0.2748$ ,  $p=0.0034$  in NCD, Pearson's  $R=0.3675$ ,  $p=1.8e-04$  in MCI/Dementia, overall  $R=0.3704$ ,  $p=0.03$ ). After regressing out group, we still observe significant correlations within cognitive subgroups, but the overall PLS model is no longer significant (Pearson's  $R=0.3215$ ,  $p=5.5e-04$  in NCD, and Pearson's  $R=0.3689$ ,  $p=1.7e-04$  in MCI/Dementia, overall  $R=0.3334$ , permutation  $P=0.2038$  for the overall PLS regression). FC: functional connectivity.

## References

1. Quitkin, F. M., Rabkin, J. G., Ross, D. & Mcgrath, P. J. Duration of Antidepressant Drug Treatment: What Is an Adequate Trial? *Arch. Gen. Psychiatry* **41**, 238–245 (1984).
2. Chandler, G. M., Iosifescu, D. V., Pollack, M. H., Targum, S. D. & Fava, M. RESEARCH: Validation of the Massachusetts General Hospital Antidepressant Treatment History Questionnaire (ATRQ). *CNS Neurosci. Ther.* **16**, 322 (2010).
3. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
4. Gratton, C., Nelson, S. M. & Gordon, E. M. Brain-behavior correlations: Two paths toward reliability. *Neuron* **110**, 1446–1449 (2022).