

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Data from paper and pen neuropsychology tests (DKEFS and RBANS) and clinical interviews (eg MADRS) were collected and entered into RedCap. No custom code was used in data collection.
Data analysis	We share all code used in the manuscript for data analysis on GitHub: https://github.com/peterzhukovsky/brain_cognition_TRD/ We used fmripred v21 to preprocess data, and MATLAB R2022a with statistics toolbox to analyze data

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The data for the OPTIMUM-Neuro study can be obtained via the NIMH Data Archive (NDAR ID 2851, <https://nda.nih.gov/study.html?id=2851>). Data from the parent

OPTIMUM clinical trial are available upon request from the authors, and the data sharing statement was included as part of the publication of the original trial (Lenze EJ, et al. N Engl J Med. DOI: 10.1056/NEJMoa2204462).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	In the present data, participant Sex identification was based on participant self-report of sex assigned at birth (i.e., female or male) at point of enrollment. We report participants biological sex in Table 1. We focus on main effects given the limitations in sample size of participants with MRI. Future work focusing on larger subsets of OPTIMUM NEURO samples eg focusing on cognitive data only may explore sex differences via sex interaction testing.
Reporting on race, ethnicity, or other socially relevant groupings	We report participants race and ethnicity in Table 1, collected following NIH guidelines (ie Race: White/Asian/African American/Other; Ethnicity: Hispanic/Non-Hispanic).
Population characteristics	Table 1 summarizes participants' mean (SD) age 67.6 (5.4) and sex (n=168 72% female); response rates were 25.9% and 32.8% in step 2 and step 1, respectively.
Recruitment	Patients in the OPTIMUM clinical trial were recruited via referrals from primary care providers and psychiatrists, outreach from the trial team, automated alerts in electronic medical records and print, radio, social media, and office advertisements. Participants who had no contra-indications for MRI scanning were offered participation in the OPTIMUM-NEURO study, that was funded to evaluate the trajectories of cognitive function (focusing on memory, processing speed and executive function) and brain structural and functional decline (focusing on cortico-limbic and fronto-executive circuits).
Ethics oversight	Ethical approval was obtained from the institutional review boards of each of the five sites - Washington University in St. Louis; Columbia University; the University of California, Los Angeles; the University of Pittsburgh; and the University of Toronto.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was determined by the OPTIMUM-NEURO and OPTIMUM studies; this study is a secondary analysis of these data.
Data exclusions	Exclusions were guided by the availability of a) good quality imaging and cognitive data in brain-cognition analyses; b) in addition, availability of remission data from the clinical trial. More detailed breakdown is available in Supplementary Figure 1.
Replication	We have conducted cross-validation analyses, with a leave-one-site-out approach for brain-cognition PLS and a k-fold cross-validation approach to predicting remission in elastic net regularized regression. This is a unique dataset combining MRI, cognitive assessments and depression treatment outcomes from a clinical trial in treatment resistant late-life depression. Therefore, out of trial generalizability analyses were not feasible.
Randomization	Treatment allocation was randomized following the OPTIMUM clinical trial. In this study we do not compare treatment outcomes between different treatment arms as these data were published by Lenze et al NEJM 2023.
Blinding	OPTIMUM was a pragmatic clinical trial, and thus not blinded.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involvement in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	(OPTIMUM was funded by the Patient-Centered Outcomes Research Institute; OPT-NEURO ClinicalTrials.gov number, NCT02960763 and OPTIMUM Award TRD-1511-33321, while OPTIMUM-Neuro was funded by the NIMH via a collaborative R01 mechanism (Pittsburgh: MH114969; Washington University: MH114966, CAMH/Toronto: MH114970; UCLA: MH114981; Columbia: MH114980). Registration was first submitted on October 28 2016.
Study protocol	More detail on the study protocol can be found at the ClinicalTrials.gov registration NCT02960763.
Data collection	Data was collected between 2017 and 2024
Outcomes	1) Primary outcomes in brain-cognition analyses were cognitive domain tests (RBANS, DKEFS) 2) Primary outcomes in prediction of remission, MADRS≤10 at 10 weeks of treatment as well as change in MADRS scores between baseline and 10 weeks.

Plants

Seed stocks	NA
Novel plant genotypes	NA
Authentication	NA

Magnetic resonance imaging

Experimental design

Design type	resting state fMRI, T1 structural MRI, diffusion weighted MRI
Design specifications	1h of scan time per subject, two runs of resting state fMRI, one T1 and one DWI run
Behavioral performance measures	no task fMRI behavioral measures from OPTIMUM-NEURO included in these analyses

Acquisition

Imaging type(s)	resting state fMRI, T1 structural MRI, diffusion weighted MRI
Field strength	3T
Sequence & imaging parameters	Echo planar imaging for fMRI and DTI
Area of acquisition	whole brain
Diffusion MRI	☒ Used ☐ Not used

Parameters multi-echo, b values of 500 (6 directions); 1000 (15 directions); 2000 (15 directions); 3000 (60 directions);

Preprocessing

Preprocessing software

fMRI: Two runs of fMRI data (2 x 10:00 minutes) were pre-processed using fmripred; the resulting minimally preprocessed images (in the Nlin6 MNI space) were denoised by regressing out 24 noise components (Supplementary Information) and smoothed with a Gaussian kernel with full-width half measure of 3mm. The first three volumes were discarded to reach steady-state equilibrium. Demeaned and normalized timeseries from the two timeseries were concatenated.

dMRI: Diffusion image pre-processing followed previous studies and included (i) brain masking (using AFNI and MRtrix3 dwi2mask), (ii) motion and eddy current correction (FSL eddy), and (iii) susceptibility distortion correction (BrainSuite BDP). We used 3D slicer to fit DTI tensors and reconstruct white matter tracts via deterministic unscented Kalman filter tractography(<https://github.com/SlicerDMRI>). Next, we ascertained individual white matter tracts by clustering fibers and applying supervised groupwise registration to the ORG (O'Donnell Research Group) atlas. Finally, we analyze the fractional anisotropy (FA) as a measure of tract integrity.

T1: We used FreeSurfer (Version 6.0.0) to derive total intracranial volume, gray, white, and CSF volumes and cortical thickness and subcortical volumes.

Normalization

We registered/normalized rs-fMRI data to MNI152 standard space using fmripred tools (ANTs and FreeSurfer)

Normalization template

fMRI: MNI152 standard space (Nlin6 in fmripred)

DTI: tracts traced in subject space

T1: cortical thickness and subcortical volumes in FreeSurfer subject space (mapped onto Talairach standard space)

Noise and artifact removal

fMRI motion: we control for mean framewise displacement and exclude subjects with mean FD>0.7 and in sensitivity analyses mean FD>0.5.

Volume censoring

we did not perform volume censoring to ensure the timeseries autocorrelation features were conserved.

Statistical modeling & inference

Model type and settings

We analyze 210 connectivity pairs for FC, 74 cortical thickness and subcortical volumes and 62 track mean FA values. We used partial least squares regression for brain-cognition associations and elastic net logistic regression for predicting remission. Remission prediction models return area under the curve for a binary 1/0 (remitted/non remitted) classification. PLSR returns %of variance explained in the cognitive outcomes.

Effect(s) tested

We test for brain-cognition associations and develop logistic regression models predicting remission to depression treatment.

Specify type of analysis: ☒ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference

we use large scale parcellations in a multivariate regression, with bootstrapping for robustness and permutation testing for significance

(See [Eklund et al. 2016](#))

Correction

Bootstrapping multivariate PLS regressions generates bootstrap PLS weights, which we threshold at $|Z|>3$ following prior work. We apply Bonferroni $p<0.05$ to identify cognitive tests associated with PLS latent variable X-scores. We argue that for each modality (rsfMRI, DWI, T1) we had a separate hypothesis to address, thus applying multiple comparison correction within modality only. Follow-up analyses test whether PLS scores were associated with education, ATHF treatment resistance, and brain structure centile scores.

Models & analysis

n/a | Involved in the study

☐ ☒ Functional and/or effective connectivity

☒ ☐ Graph analysis

☐ ☒ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

full Pearson's correlations

Multivariate modeling and predictive analysis

Statistical analyses. We used three partial least squares regressions (plsregress.m, SIMPLS algorithm) to identify latent variables capturing multivariate relationships between functional connectivity and cognitive function (FC-PLS), between fractional anisotropy and cognitive function (WM-PLS) and between gray matter structure and cognition (GM-PLS). Model significance was tested using permutation testing following previous studies ($n=5,000$). The PLS permutation procedure tests whether the model with one, two or three latent variables jointly explains more variance than expected by chance. Permutation involves random row-wise reshuffling of the Y data on each permutation, and then running a PLS regression, while keeping the correlation structure of the original X and permuted Y data.

We z-scored the predictor matrix X and the outcome matrix Y ($X=211 \times 210$ and $Y=211 \times 12$ 6 in the FC-PLS; $X=219 \times 62$ and $Y=219 \times 12$ 10 in the WM-PLS; $X=212 \times 74$ and $Y=212 \times 12$ 6 in the GM-PLS). PLS returns a set

of latent variables that attempt to maximize the covariance between the PLS scores summarizing X and Y. PLS scores are a linear combination of the predictor variables (X) and component loadings. We used bootstrapping ($n = 5,000$) to identify predictors that showed robust contributions to the each PLS latent variable. A threshold of $|Z| > 3$ was chosen to identify the most robust connectivities significantly associated with cognitive performance (see above). We correlated the latent brain scores (XS) with the cognitive tests and applied Bonferroni correction ($P < 0.05$) to identify cognitive tests significantly associated with each latent variable.

Robustness analysis in held-out data. To evaluate the robustness of PLS performance in each of the sites, we split our participants into four subsamples, one for each included site. We used three of these subsamples as training data and the remaining subsample as test data. We applied the PLS beta regression coefficients obtained in the training sample to the test sample and correlated the observed cognitive data with the predicted cognitive data to assess PLS performance in predicting cognitive function in held-out data⁸¹. Instead of keeping all cognitive tests, we created a composite cognitive variable using a principal component analysis of variables significantly associated with FC, tract integrity, and brain structure in the FC-PLS, WM-PLS, and GM-PLS respectively.

Prediction of treatment outcomes. We used regularized, cross-validated elastic net logistic regressions to predict remission ($\text{MADRS} \leq 10$) to approximately 10 weeks of acute antidepressant treatment in the parent OPTIMUM trial. Step 1 and step 2 were considered as separate studies. On each of 100 iterations, we split the data into training and test datasets, featuring 20 randomly selected participants in the test dataset for step 1 and featuring 20 randomly selected participants in the test dataset for step 2. This approximately corresponds to 8-fold and 3.5-fold cross-validation in the outer fold. On each iteration, we used 10-fold cross-validation to train the elastic net models in the inner fold. Area-under-the-curve (AUC) measures were then used to assess model performance in the held-out test data. Three sets of prediction models were run, each including demographic, clinical, cognitive data and one of the three imaging modalities. The clinical model included age, sex, and baseline MADRS scores. These models were run in the larger sample ($n=397$) including cognitive data, clinical data, and demographic data. Then the model was re-run in the sample ($n=234$) that also included neuroimaging to determine the potential improvement in the AUC when adding neuroimaging. Stringent cross-validation approach ensured that model training was completely separate from the test data. Following this iterative process, we selected the most parsimonious model that included predictors surviving regularization in over 95% of models ran. We then tested the performance of this most parsimonious model in an even 8-fold and 5-fold split of the step 1 and step 2 data, respectively, presenting confusion matrices of this model showcasing the true positives and negatives as well as false positives and negatives. Feature selection in this most parsimonious model was affected by train and test data, with cross-validation applying to the regression weights but not to the variable selection as variable selection process was done across the whole sample. In addition, we have used a PLS model predicting MADRS change (absolute difference in MADRS score between baseline and the end of last step completed) from clinical (MADRS), demographic (age, sex), cognitive and brain structure data. More information on predictive modeling can be found in the Supplementary Information.