

RESEARCH ARTICLE

Distributed functional connectivity predicts neuropsychological test performance among older adults

Seyul Kwak¹ | Hairin Kim¹ | Hoyoung Kim² | Yoosik Youm³ | Jeanyung Chey¹

¹Department of Psychology, Seoul National University, Seoul, Republic of Korea

²Department of Psychology, Chonbuk National University, Jeonju, Republic of Korea

³Department of Sociology, Yonsei University, Seoul, Republic of Korea

Correspondence

Jeanyung Chey, Department of Psychology, Seoul National University, Gwanak-ro 1, Gwanak-gu, Seoul, Republic of Korea.
Email: jychey@snu.ac.kr

Funding information

National Research Foundation of Korea, Grant/Award Number: NRF-2017S1A3A2067165; Ministry of Education, Science and Technology; National Institutes of Health (NIH), Grant/Award Numbers: R01EB009352, UL1TR000448, R01AG043434, P01AG003991, P01AG026276, P30NS09857781, P50AG00561

Abstract

Neuropsychological test is an essential tool in assessing cognitive and functional changes associated with late-life neurocognitive disorders. Despite the utility of the neuropsychological test, the brain-wide neural basis of the test performance remains unclear. Using the predictive modeling approach, we aimed to identify the optimal combination of functional connectivities that predicts neuropsychological test scores of novel individuals. Resting-state functional connectivity and neuropsychological tests included in the OASIS-3 dataset ($n = 428$) were used to train the predictive models, and the identified models were iteratively applied to the holdout internal test set ($n = 216$) and external test set (KSHAP, $n = 151$). We found that the connectivity-based predicted score tracked the actual behavioral test scores ($r = 0.08$ – 0.44). The predictive models utilizing most of the connectivity features showed better accuracy than those composed of focal connectivity features, suggesting that its neural basis is largely distributed across multiple brain systems. The discriminant and clinical validity of the predictive models were further assessed. Our results suggest that late-life neuropsychological test performance can be formally characterized with distributed connectome-based predictive models, and further translational evidence is needed when developing theoretically valid and clinically incremental predictive models.

KEYWORDS

brain connectomics, cognitive aging, dementia, machine learning, neuropsychological test, predictive modeling

1 | INTRODUCTION

The neuropsychological test has a distinct role in detecting and monitoring cognitive and functional changes associated with dementia-associated diseases (Fields, Ferman, Boeve, & Smith, 2011). Neuropsychological tests systematically describe the abilities to perform given cognitive tasks and evaluate how an individual will cope with daily-life functional activities (Donders, 2019; Fields et al., 2010). Thus, neuropsychological tests not only detect the presence of clinical impairment but also provide prognostically useful information in late-

life neurodegenerative disease (Belleville, Fouquet, Duchesne, Collins, & Hudon, 2014).

Despite strong evidence of the clinical utility of neuropsychological tests, however, it is unclear what neurobiological information the test scores provide. While the progression of the pathophysiological process of dementia strongly affects neuropsychological function (Jack et al., 2019; Mortamais et al., 2017), a large amount of variance is left unexplained even with multimodal biomarker information (Arenaza-Urquijo & Vemuri, 2018; Habeck et al., 2016; Vemuri et al., 2011). Moreover, it is hard to predict whether the presence of

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *Human Brain Mapping* published by Wiley Periodicals LLC.

pathophysiological markers directly leads to neural dysfunction within the behaviorally relevant regions (Hohman et al., 2016). These discordances suggest the need for narrowing down the explanatory gap between brain function and behavioral impairment that are relevant to neuropsychological tests.

To elucidate the underlying neural basis of neuropsychological tests, previous studies have examined the brain regional structures and functions highly associated with behavioral performance (Bayram, Caldwell, & Banks, 2018; Genon, Reid, Langner, Amunts, & Eickhoff, 2018; Mortamais et al., 2017). However, the previous studies typically narrowed down the region of interest within a focal brain structure (e.g., hippocampus) or averaged out overall brain measures features into fewer predictors (Charroud et al., 2016; Duchek et al., 2013; Shaw, Schultz, Sperling, & Hedden, 2015; Van Petten, 2004). While pinpointing or averaging the candidate neural correlate makes the theoretical account succinct, it is hard to discern whether the focally identified brain correlates are an optimal and exclusively relevant unit in explaining individual differences of cognitive function. Moreover, exploring multiple neural correlates with repetitive univariate tests leads to excessively conservative thresholding, and selective reports of focal correlates tend not to replicate in other datasets (Masouleh, Eickhoff, Hoffstaedter, & Genon, 2019). Such a limited scope of neural features may hinder the valid identification of replicable neural correlate.

An advancement of individualized neuroimaging and machine learning techniques has enabled characterizing the multivariate nature of the brain–behavior relationship (Bzdok & Ioannidis, 2019; Dubois & Adolphs, 2016; Jollans & Whelan, 2018). With the virtue of both approaches, it became possible to find the optimal balance between an overly simplified theory-driven model and a complex data-driven model of the brain–behavior relationship. More specifically, individuals' detailed profiles of functional connectivity (FC) patterns not only can classify clinical diagnosis but also can predict continuously measured behavioral traits. Accumulating evidence suggests that behavioral individual differences that are relevant in characterizing risks of neuropsychiatric disorders including attention task performance, personality, intelligence, and neuropsychological performance can be robustly predicted with functional connectivity patterns (Dubois, Galdi, Paul, & Adolphs, 2018; Lin et al., 2018; Nostro et al., 2018; Rosenberg, Hsu, Scheinost, Todd Constable, & Chun, 2018; Sui et al., 2018). Moreover, with the advantage of feasibility, resting-state FC has been highlighted as a promising approach in investigating the neural basis of brain aging and preclinical dementia mechanisms (Ferreira & Busatto, 2013; Sala-Llonch, Bartrés-Faz, & Junqué, 2015; Sheline & Raichle, 2013). Brain functional features of FC may better explain whether an individual will undergo significant behavioral impairment when the location and presence of neuropathology cannot fully account for such individual differences (Fox, 2018; Siegel et al., 2016).

In the current study, we aimed to identify the underlying brain connectivity basis of neuropsychological test performance in the older adult population. Connectome-based predictive modeling combined with a regularized regression technique was applied to examine the optimal combinations of neural features required for accurate

prediction. We hypothesized that a widespread set of FC features are predictive of neuropsychological performance in novel individuals. In addition, we further tested whether the identified predictive models generalize to an external dataset composed of heterogeneous demographic characteristics (Dwyer, Falkai, & Koutsouleris, 2018; Woo, Chang, Lindquist, & Wager, 2017). While several studies have identified the predictive pattern of neuropsychological test performance, previous attempts mostly lacked evidence of external validity. Most of the reported prediction accuracies are confined to internal cross-validation, which tests the generalizability of a model only within a homogeneous dataset. If the predictive model captures a large amount of variance unique to the given dataset (e.g., populational characteristics, MRI protocols), the theoretical reliability of the predictive model will be limited.

Finally, we further aimed to examine the discriminant and clinical validity of the connectome-based predictive models. Although neuropsychological test performances are highly correlated with each other, each test represents theoretically distinct domains of cognitive function (Park et al., 2012). The discriminant validity can be tested based on how the predictive models are discriminantly correlated with the specific construct of cognitive function (Habeck et al., 2015; Woo et al., 2017). The predictive model of cognitive function can either be confined to the targeted score or indistinctively generalize to other test scores (Avery et al., 2019; Jangraw et al., 2018; Rosenberg et al., 2015). The clinical validity, on the other hand, can be tested whether neurally predicted scores from the connectome provide incremental clinical information beyond the actual behavioral score (Jollans & Whelan, 2016; Moons et al., 2012). If the connectome-based prediction model only captures redundant information from the original behavioral score, the neurally expected score will not additively explain the severity of functional impairment relevant to dementia.

2 | METHODS AND MATERIALS

2.1 | Participants

Neuropsychological tests and neuroimaging datasets shared in Open Access Series of Imaging Studies (OASIS-3) were used to test the validity of the connectome-based predictive models. OASIS-3 dataset shared clinical, neuropsychological, neuroimaging, and biomarker data of 1,098 participants (age range: 42–95; www.oasis-brain.org; LaMontagne et al., 2018). We analyzed the initial baseline data of 644 participants who completed both neuropsychological tests and MRI scans. Participants with incomplete MRI scans (without T1 structural image and two sessions of resting fMRI) ($n = 366$), excessive head movement ($n = 37$), incomplete neuropsychological tests at the baseline ($n = 51$) were excluded from the analysis. Descriptive statistics are provided based on the CDR score (Clinical Dementia Rating; Table 1). The CDR is a semi-structured interview developed to provide a global rating of dementia severity, and it is useful for staging and tracking decline in AD (Fillenbaum, Peterson, & Morris, 1996; J. C. Morris et al., 1997; J. C. Morris, 1997). Each CDR scores represented levels

TABLE 1 Descriptive statistics of OASIS-3 (internal validation set, $n = 644$) and KSHAP (external validation set, $n = 151$)

	OASIS-3 internal validation dataset ($n = 644$)			KSHAP external validation dataset ($n = 151$)
	Not impaired ($n = 436$)	Very mild impairment ($n = 169$)	Mild-moderate impairment ($n = 39$)	
	Mean \pm SD (range)			
Age	70.9 \pm 6.52 (46–92)	72.2 \pm 6.87 (50–88)	76.1 \pm 8.79 (60–96)	71.7 \pm 6.56 (59–93)
Sex (female: Male)	239:197	72:97	16:23	96:55
Years of education	15.8 \pm 2.67 (8–29)	15.2 \pm 2.85 (7–23)	14.8 \pm 3.34 (8–20)	7.2 \pm 4.12 (0–23)
MMSE ^a /MMSE-DS	28.9 \pm 1.29 (21–30)	27.0 \pm 2.72 (18–30)	24.0 \pm 3.29 (18–30)	27.0 \pm 2.14 (21–30)
CDR	0	0.5	1.04 \pm 0.21	0.02 \pm 0.11
CDR-SOB ^b	0.02 \pm 0.12	1.73 \pm 1.02	5.35 \pm 1.54	

Note: Descriptive statistics of OASIS-3 dataset were presented based on CDR global score (0 = no impairment, 0.5 = questionable impairment, 1 = mild impairment, 2 = moderate impairment) which provide summarized estimate of dementia severity.

Abbreviations: CDR, Clinical Dementia Rating global score; CDR-SOB, Clinical Dementia Rating Sum of Boxes; MMSE, Mini-Mental Status Examination; MMSE-DS, Mini-Mental Status Examination Dementia Screen.

^aMissing value omitted (Not impaired: 1).

^bMissing value omitted (Not impaired: 11, Very Mild Dementia: 7). MMSE- DS for external validation dataset.

of functional impairment (0 = no impairment, 0.5 = questionable impairment, 1 = mild impairment, 2 = moderate impairment) and summarized estimate of dementia severity (Marcus et al., 2007; J. Morris, 1993).

To examine whether the predictive model is robustly generalized to the heterogeneous demographical dataset, participants subsampled from Korean Social Life, Health, and Aging Project (KSHAP) were used as an external validation dataset (J. Lee et al., 2014; Youm et al., 2014). KSHAP participants who completed a neuropsychological assessment, psychosocial surveys, and neuroimaging scans were included in the study. The following exclusion criteria were applied: the presence of psychiatric or neurological disorders, vision or hearing problems, having metal in the body that cannot be removed, having a history of losing consciousness due to head trauma, showed excessive head movement during scans. Furthermore, older adults who were highly suspected of dementia were screened out based on age and education-stratified norm (< -1.5 SD) of Mini-Mental State Examination for Dementia Screening (MMSE-DS) (Han et al., 2010). The CDR score ranged from 0 to 0.5, including seven participants with very mild functional impairment (CDR = 0.5, $n = 7$). Among participants who completed both neuropsychological tests and neuroimaging scans, the final data were composed of 151 subjects who did not meet any of the exclusion criteria (Table 1). The study was approved by the Institutional Review Boards of Seoul National University and Yonsei University. All participants provided written informed consent to the research procedures.

2.2 | Neuropsychological test

2.2.1 | OASIS-3 neuropsychological test

OASIS-3 consists of Neuropsychological Battery of the Uniform Data Set (UDSNB) developed in Alzheimer's Disease Centers (ADC) to

establish unified and standardized data collection (Weintraub et al., 2009). The dataset of 10 neuropsychological tests measuring attention/working memory, executive function, processing speed, language, and episodic memory are currently shared (<https://central.xnat.org>).

Digit Span Forward (DIGI-F) and Digit Span Backward (DIGI-B) were used to assess attention and working memory function with the subtests included in the Wechsler Memory Scale (WMS-R) (Wechsler, 1987b). Total correct trials from 2 to 7 (backward) or 3 to 8 (forward) length of digits were counted after a verbal presentation of digit numbers. Category fluency of animal (FLU-ANI) and vegetable (FLU-VEG) were measured with a total number of words generated in 1 min (Moms et al., 1989). Trail Making Test Part A (TRAIL-A) consisted of consecutively numbered circles arranged randomly on a sheet of paper. Participants were asked to draw a line between the circles in ascending order as quickly as possible. In Trail Making Test Part B (TRAIL-B), participants were asked to consecutively connect lines between the number (ascending order) and alphabet in an alternating way (1-A-2-B-...; Reitan & Wolfson, 1993). If the subject cannot complete the sample item for each part or exceeds the time limits, a maximum time score is assigned (TRAIL-A: 150 s, TRAIL-B: 300 s). Total time (s) to complete the lines were log-transformed and inversed to adjust high skewness and indicate the same performance direction. Digit Symbol Coding from WAIS-R (Wechsler, 1987a) was administered in the standard way, with the total number of items completed correctly in 90 s as the total score. Episodic memory function was assessed with Logical Memory (LOGI MEM) from WMS-R (Wechsler, 1987b). Total items of Story A in the Immediate and Delayed Recall trial were used. The Boston Naming Test score consisted of the total number of items named correctly within the 20-s limit plus the number of items named correctly with a semantic cue (Kaplan et al., 1983). The tests represented cognitive domains of attention (Digit Span), memory (Logical Memory), language (Boston Naming, Fluency), executive, and speed (Trail Making, Digit Symbol) (Hayden et al., 2011; Park et al., 2012).

Among 732 participants who participated in both neuropsychological tests and MRI scans, we excluded participants who did not complete any of the tests [Digit Span Forward, Category fluency, Logical Memory score missing ($n = 2$); Digit Span Backward, Trail Making Test Part A score missing ($n = 3$); Trail Making Test Part B missing ($n = 45$)], leaving 644 final analysis set.

2.2.2 | KSHAP neuropsychological test

To test the generalizability of the predictive model, we identified eight homologous neuropsychological tests included in the KSHAP dataset. Similarity and correspondence of administration procedures and contents were scrutinized. For the attention and working memory, Digit Span Forward, and Digit Span Backward tests included in the Elderly Memory disorder Scale (Chey, 2007) were available. The score was the sum of all correct trials were counted from 2 to 8 lengths of digits. The category fluency test asked participants to generate words from the two semantic categories (i.e., animal and supermarket) each within a minute (Kang, Chin, Na, Lee, & Park, 2000; Kang, Jang, & Na, 2012). While interference condition of the original Trail Making Test in UDSNB asks to alternate between numbers and alphabets (Reitan & Wolfson, 1993), modified Trail Making Test (mTMT) asked to alternate between numbers and symbols (triangle and rectangle) to minimize the floor effect of illiterate elderly (M. Park & Chey, 2003; Seo et al., 2006). The mTMT Part A (Trail A) consisted of 15 consecutively numbered circles arranged randomly on a sheet of paper. Participants were asked to draw a line between the circles in ascending order as quickly as possible. In mTMT Part B, they were asked to draw between triangles and squares alternately. In mTMT Part C (Trail C), participants consecutively connected lines between 8 numbers (ascending order) and 7 shapes (triangle and square) in an alternating way (1 - \triangle - 2 - \square - 3 - \triangle ...). Total time consumed to complete the mTMT-A and mTMT-C were used in the analysis, and mTMT-B was not used in the analysis due to the variability in the possible trail options. A few participants exceeded the typical time limit of mTMT-A (150 s, $n = 1$) and mTMT-C (300 s, $n = 7$), but a maximum of 300 and 600 s, respectively, were allowed to finish the task due to the participants' unskilled usage of pencil. Total time (s) to complete the lines were log-transformed and inverted. The Story Recall Test included in the Elderly Memory disorder Scale was used to assess episodic memory function. The Story Recall Test consisted of one story modified from the Logical Memory subtest in WMS-III into a more culturally familiar content (An & Chey, 2004). The SRT required subjects to recall a paragraph containing 24 semantic units and the delayed recall subtests were administered 15–30 min after the immediate recall session.

2.3 | MRI acquisition and preprocessing

2.3.1 | OASIS-3 neuroimaging

Neuroimaging dataset in OASIS-3 was collected in a 16-channel head coil of different scanners (Siemens TIM Trio 3T, Siemens BioGraph

mMR PET-MR 3T, Siemens BioGraph mMR PET-MR 3T, Siemens Sonata 1.5T, Siemens Vision 1.5T). High resolution T1-weighted structural image (TR = 2.4 s, TE = 3.08 ms, FOV = 256×256 mm, FA = 8° , voxel size $1 \times 1 \times 1$ mm³) and resting-state functional image (EPI; TR = 2.2 s, TE = 27 ms, FOV = 240×240 mm, FA = 90° , voxel size $4 \times 4 \times 4$ mm, 36 slices) were used. Participants who completed both structural and two consecutive runs of functional scans (6 min and 164 volumes per run) were analyzed.

Image preprocessing and denoising was performed using the SPM12 software (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK) with the Conn toolbox 18.a (<http://www.nitrc.org/projects/conn>) default preprocessing pipeline. Functional images were corrected for motion and warped into MNI standard space. Images were smoothed with a Gaussian kernel of 8 mm full-width half-maximum. In addition, the Artifact Detection Tools (https://www.nitrc.org/projects/artifact_detect/) was used to identify motion and signal intensity outlier images. Images with global mean intensity Z-value >5 and movement >0.9 mm were identified as outlier images. Estimated motion parameters and outlier images were used as nuisance covariates in the time-series linear regression. T1-weighted images were segmented into gray matter, white matter, and cerebrospinal fluid and warped into MNI standard space. Signals within white matter and CSF mask were regressed out to exclude nongray matter BOLD signal. Band-pass temporal filtering (0.008–0.09) was applied to exclude physiological noise signals. Participants with excessive head motion were excluded (max motion >4.5 , mean motion >0.6 , $n = 51$).

For each subject, mean time-series were extracted by averaging all voxels composing each region for each time point from 227 regions from the Shen et al. (2013) brain atlas (37 cerebellar regions were excluded; Shen et al., 2013; Figure S1). Pearson correlation across time-series was calculated between each pair of regions and transformed to Fisher's Z scores. Therefore, 644 individuals' whole-brain connectivity matrices containing $(227 \times [227-1])/2 = 25,651$ pairwise functional connectivity (FC) values were constructed and vectorized in the following analyses.

2.3.2 | KSHAP neuroimaging

T1-weighted magnetic prepared rapid gradient echo (MP-RAGE) image and Resting-state fMRI data were acquired on a 3T Siemens Trio 32channel scanner (T1: Sagittal slices, slice thickness 1 mm, TR = 2,300 ms, TE = 2.36 ms, FOV = 256×256 mm, FA = 9° , voxel size $1 \times 1 \times 1$ mm³; Resting-state EPI: TR = 2000 ms, TE = 30 ms, FOV = 240×240 mm, FA = 79° , voxel size $3 \times 3 \times 3$ mm, gap = 1 mm, acquisition time = 5 min). During the scan, participants were instructed to rest quietly with their eyes open and not to fall asleep. We acquired two runs of 150 contiguous functional images. Thus, the total length and the number of images were similar to the OASIS-3 dataset (OASIS-3:328 volumes, 12 min; KSHAP: 300 volumes, 10 min). To acquire high spatial resolution, cerebellar regions were excluded from the acquisition. All of the preprocessing and functional network construction were conducted with the same procedure,

except for adding slice timing correction. Participants with excessive head motion were excluded (max motion >4.5, mean motion >0.6, $n = 26$). Constructing individuals' FC matrix of each participant was also the same as the procedures conducted with OASIS-3 data. Therefore, 151 individuals' 25,651 vectorized functional connectivities were used in the following analyses.

2.4 | Connectome-based predictive modeling

2.4.1 | Internal validation: Prediction from discovery to holdout sample

Due to the high dimensionality of whole-brain FC features, regularization techniques are widely used to select an optimal set of predictive features. Ridge regression develops a model that minimizes the sum of both the mean prediction error and L2-norm regularization term (sum of the squares of regression coefficients). Unlike the Ordinary Least Square (OLS) method, this technique not only minimizes the errors but also penalizes the total amount of explanatory weight in the model. A regularization parameter λ is used to control the trade-off between the prediction error in the training data and the size of total regression coefficients, that is, model complexity. A large λ gives a stronger penalty to the larger sum of beta coefficients and leads to a model with more shrunk regression coefficients, while a small λ allows large room for regression coefficients to explain the target variable. The optimal value of λ that minimizes prediction error is tuned by the cross-validation procedure. Previous studies have successfully predicted individual differences in various behavioral phenotypes using L2-norm regularization (ridge regression) (Cui & Gong, 2018; Dadi et al., 2019; Gao, Greene, Constable, & Scheinost, 2019; Siegel

et al., 2016). L2-norm (sum of squared betas) penalizes irrelevant features by reducing the size of weights.

Connectome-based predictive modeling was initially conducted by randomly splitting the train set and the test set (Figure 1). The training dataset was used to identify optimal predictive weights of the connectivity that contributes to accurate prediction in the behavioral scores, while the test dataset was used to evaluate how much the identified FC pattern is expressed in the novel individuals. The pattern expression score (i.e., FC-predicted score) was calculated with the dot product ($\beta \times X$) between predictive weights and the FC vector. The dot product of all 25,651 features estimated the predicted behavioral score based on the patterns of FC. The Pearson's correlation (r) between the actual (observed) neuropsychological test score and the FC-predicted score in the test dataset indicated the predictive accuracy.

Connectome-based predictive models of neuropsychological test scores were identified by utilizing a nested cross-validation procedure (Varoquaux, 2018). Initially, the internal cross-validation was initially conducted with the OASIS-3 dataset (Figure 2a). Dataset was randomly split into the dataset for model training (discovery sample; 2/3, $n = 428$) and the dataset for testing predictive accuracy (holdout sample; 1/3, $n = 216$). The target variable was scaled and mean-centered within the training dataset.

In the discovery sample, hyperparameter (λ) which regulates the complexity of the predictive model was tuned using three-fold cross-validation. The grid range of λ was generated based on the automatic algorithm implemented in the *glmnet* package. It generated 100 values of λ linear on the log-scale starting from the maximum λ value that converges all of the coefficients to zeros (Hastie et al., 2016). After finding a λ value that minimizes the mean absolute error in the nested cross-validation, the tuned λ is used to fit the prediction model and

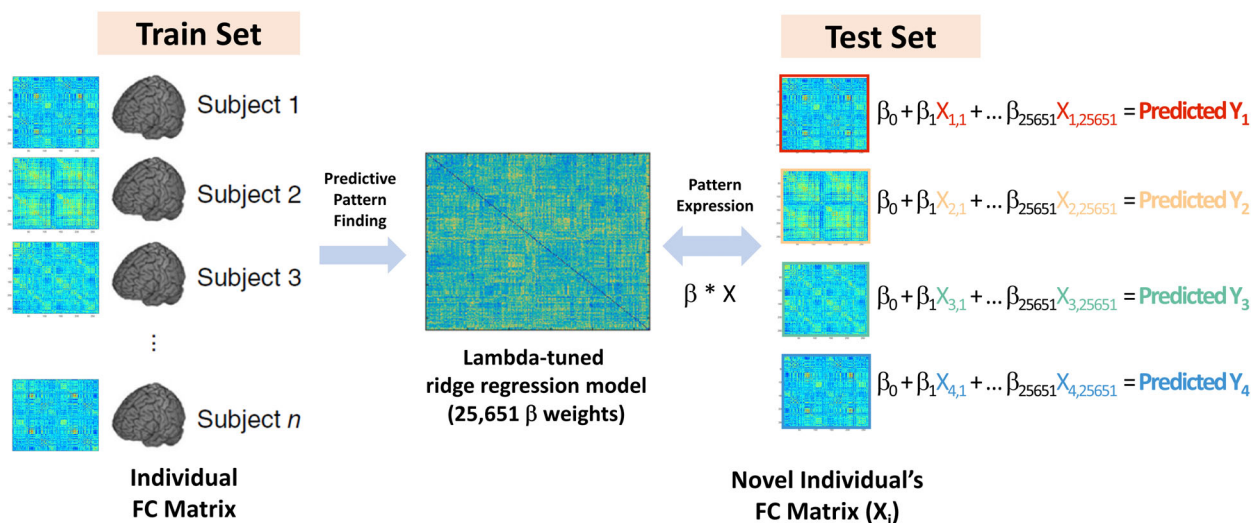


FIGURE 1 Schematic overview of the connectome-based predictive modeling. Train set of individuals' FC matrix (left) is used to identify predictive weights across all connectivity features. The dot product between beta weights and the FC matrix of novel individuals estimates the FC-predicted test scores

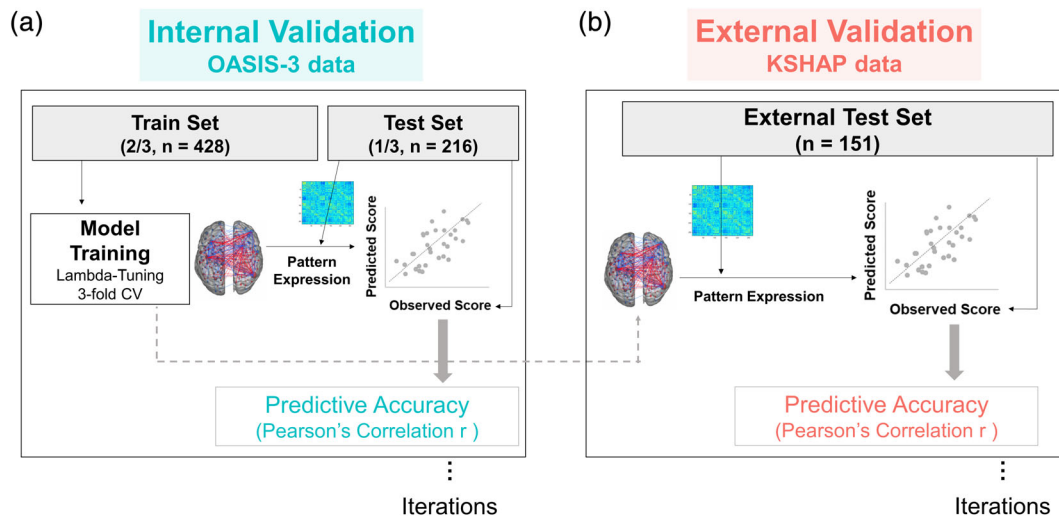


FIGURE 2 Schematic overview of nested cross-validation. Internal validation (left): Threefold cross-validation identifies lambda-tuned predictive weights in the train set. The trained predictive model is applied to test set FC. The FC-predicted score (pattern expression) was calculated by the dot product between predictive weights and an individual's FC. Correlation between FC-predicted score and the actual neuropsychological test score indicated predictive accuracy. External validation (right): Same iterative prediction was applied to KSHAP dataset

was applied to the holdout sample. To account for the variability of group splitting, the whole procedure was iterated 50 times. The iteration of the loop showed a range of predictive accuracy and deviations across iterations. When assessing the statistical significance of the prediction accuracy, the nested iteration of 1,000 times estimated the 95% confidence interval of the precision.

We conducted the same analysis when controlling for age and education effect. Before conducting predictive modeling, we regressed out the effect of age or education (years) on the cognitive test score, respectively in the test set. We compared how predictive accuracy (correlation coefficient) decreases when residualizing the effect of age or education. An identical analysis was conducted adjusting for the effect of sex and head motion as a nuisance variable.

While the main predictive accuracy was evaluated with the iteration of the nested cross-validation procedure, the distributive relationship between the FC-predicted score and the observed behavioral score in the scatter plots are presented for visualization purpose. The predicted scores calculated based on nested three-fold cross-validation were plotted. The training and testing dataset folds are randomly split. Each predicted score of 1/3 of the data fold was allotted using the predictive model trained within the rest of 2/3 data folds. In the case of the whole dataset within the internal cross-validation procedure, the p values of the empirical correlation values, based on their corresponding null distribution, were computed by the following formula: $(1 + \text{the number of permuted } r \text{ values greater than or equal to the empirical } r) / 1,001$. The permutation test results revealed significant p values for the empirical r values.

2.4.2 | External validation: Prediction in external sample

After identifying eight homologous neuropsychological tests, the same prediction procedure was applied to test whether each test model

generalizes to external study with different populations and protocols (KSHAP dataset, $n = 151$; Figure 2b). Using the predictive model trained in the OASIS-3 discovery sample ($n = 428$), the dot product between predictive weights and FC vectors of KSHAP dataset estimated the FC-predicted score (pattern expression score). The training procedure identical to internal validation was applied and iterated 50 times using ridge regression. When assessing the statistical significance of the prediction accuracy, the nested iteration of 1,000 times estimated the 95% confidence interval of the precision.

Since clinical status and educational attainment widely differed between internal validation (OASIS-3) and external validation dataset (KSHAP), we additionally examined whether the predictive model trained with a more homogeneous population (clinically normal, lower-educated) shows better predictive accuracy.

2.5 | Spatial pattern of predictive model

The predictive accuracy can differ according to the range and extent to which features are considered. In this study, we examined how feature thresholding, feature sparsity, or pre-defined network shows systematically different levels of predictability and exhibit optimally distributed predictive models.

2.5.1 | Feature filtering

To pinpoint the necessary range of brain predictors and infer the optimal sparsity of the model, we examined the systematic effect of the feature selection threshold on the predictive accuracy. Since using every FC feature in training can introduce noisy information which does not contribute to prediction, the filtering method may provide optimal precision of the predictive model. The filtering method

excludes the predictors which are not strongly or weakly associated with the target variable (i.e., test score) by conducting univariate tests on each feature. This method can reduce dimensionality into a fewer set of predictors that are relevant to the outcome variable. Previous studies have shown that the extent of feature selection threshold may systematically influence the predictive performance (Gao et al., 2019; Greene, Gao, Scheinost, & Constable, 2018; Jangraw et al., 2018).

The spatial distributedness was examined in the OASIS-3 dataset which was used in the internal cross-validation. In the training set, FC features that are not correlated with the target test score above the specified threshold (Pearson's correlation $|r| < .01, .02, \dots, .18$) were excluded in the predictive modeling. The same features selected in the training set were also applied to the testing dataset. As the feature selection threshold increases, the features of the predictive model will be constructed with focally correlating features. If the individual difference of neuropsychological test score is based on the connectivities of broad brain regions, thresholding of features will lead to a significant decrease in the predictive performance. We identified the threshold point of maximum prediction accuracy and the inflection point was considered optimal sparsity of the prediction model.

2.5.2 | Feature regularization type

To evaluate whether neuropsychological test performance is driven by selective features of functional connections, we additionally examined the prediction result across hyperparameter (α) that affects the sparsity of the predictive model. The mixing hyperparameter α (proportion of L2-regularization term) defines the tendency of predictive weights to shrink into zero as shown in the following formula.

$$\sum (y_i - x_i^T \beta_i)^2 + \gamma \sum \left(\alpha |\beta_i| + \frac{1}{2} (1 - \alpha) \beta_i^2 \right)$$

The current study will examine the sparsity of the features across the α values and infer the effect of feature sparsity (Gao et al., 2019). We examined whether predictive models with weak and distributed features (low- α elastic net or ridge regression) more accurately predict neuropsychological performance than models with a selective number of features (high- α elastic net or LASSO). The same predictive modeling procedure was iterated across the 21 grid values of α ([0.00, 0.05, ..., 1.00]²). The systematic effect of the increased model sparsity on the predictive accuracy was examined.

2.5.3 | Brain functional network composition of predictive model

In the previous study, brain regions are parcellated and clustered into eight functional networks (frontoparietal, medial frontal, default mode, subcortical/salience, motor/auditory, Visual I, Visual II, and visual association; Shen et al., 2013; Figure S1). The functional networks indicate a coherent set of regions that activates together and forms modular communities (Laird et al., 2013; Power et al., 2011). In order

to interpret the relative importance of specific functional networks in predictive models, we assessed how much predictive accuracy decreases when excluding each functional network in conducting the same predictive analysis (Dubois et al., 2018; Whelan et al., 2014). If a particular test score requires unique neural information from a specific functional network, excluding the network connectivity information will deteriorate the original predictive performance. On the contrary, a minimal change will indicate that the neural basis of test performance is not confined to the excluded functional network.

2.6 | Discriminant validity of predictive model

We examined whether the connectome-based predictive model can represent a specific cognitive construct with discriminant validity. If the predictive model distinctively predicts the cognitive test that has trained the model, the predictive model may have captured a specific cognitive construct that was attempted to measure (Woo et al., 2017). On the contrary, if a specific FC-predicted score is correlated not only with the test score used to train the very model but also with the other test scores representing distinct cognitive construct, it is more likely that the predictive model captures general and common component of cognitive function (Habeck et al., 2015; Rosenberg et al., 2015). The predictive models were trained with the OASIS-3 discovery sample ($n = 428$), and the FC-predicted score was estimated using the FC expression of the internal holdout sample (OASIS-3, $n = 216$) and KSHAP sample ($n = 151$). In this way, we generated eight FC-predicted scores of each dataset. The 8-by-8 correlation matrix between the actual behavioral scores and FC-predicted scores of the eight cognitive tests was averaged across iterations of 1,000 times predictions. A higher correlation in the diagonal elements than off-diagonal elements indicated convergent and discriminant characteristics of the predictive model.

2.7 | Clinical validity of predictive model

To examine the clinical validity of the predictive model, we tested Spearman's rank correlation between neuropsychological test score (observed or predicted) and clinician-rated functional impairment score across each pair of the eight homologous tests. The Clinical Dementia Rating scale-sum of boxes (CDR-SOB) was used to evaluate the ecological relevance of the predictive models. The CDR-SOB provides levels of functional impairment across six domains of function (memory, orientation, judgment, community affairs, home and hobbies, personal care) (Lynch et al., 2005; J. C. Morris, 1997).

The FC-predicted score in the OASIS-3 dataset was estimated using the predictive models constructed with the homologous tests included in the KSHAP external dataset. Using three-fold cross-validation, optimal hyperparameter (λ) that minimizes mean absolute error (MAE) was identified in ridge regression ($\alpha = 0$). Then the fitted regression model was applied to FCs of OASIS-3 and estimated each of the eight FC-predicted scores. The partial rank correlation between the neuropsychological score (observed or predicted) and CDR-SOB controlling for the counterpart score (either the predicted or observed

score) was tested to examine the incremental value of the FC-predicted score. The remaining correlation even after adjusting the effect of the actual behavior score indicated the incremental information provided from the FC-predicted score. Partial rank correlation analysis was conducted using *ppcor* package (Kim, 2015).

3 | RESULTS

3.1 | Internal cross-validation

We first examined the correlation between observed behavioral test scores and FC-predicted scores. The three-fold cross-validation

yielded prediction results in OASIS-3 total dataset (Figure 3). The ridge regression was used to regularize prediction coefficients in each iteration. The permutation testing also confirmed a significant correlation ($p < 0.001$) between the predicted and observed scores except for the marginal significance in the Digit Span Forward test ($p = .049$).

For the main predictive analysis using nested cross-validation, prediction accuracy ranged from $r = 0.08$ – 0.44 (Figure 4). We additionally examined prediction results after adjusting the effect of age and education in both the training set and test set. Age-adjustment resulted in a large decrease in prediction accuracy whereas education-adjusted results showed a minimal decrease in the accuracy, indicating the age-related attributes of the predictive model (no covariate mean correlation = 0.253, education-adjustment mean correlation = 0.254,

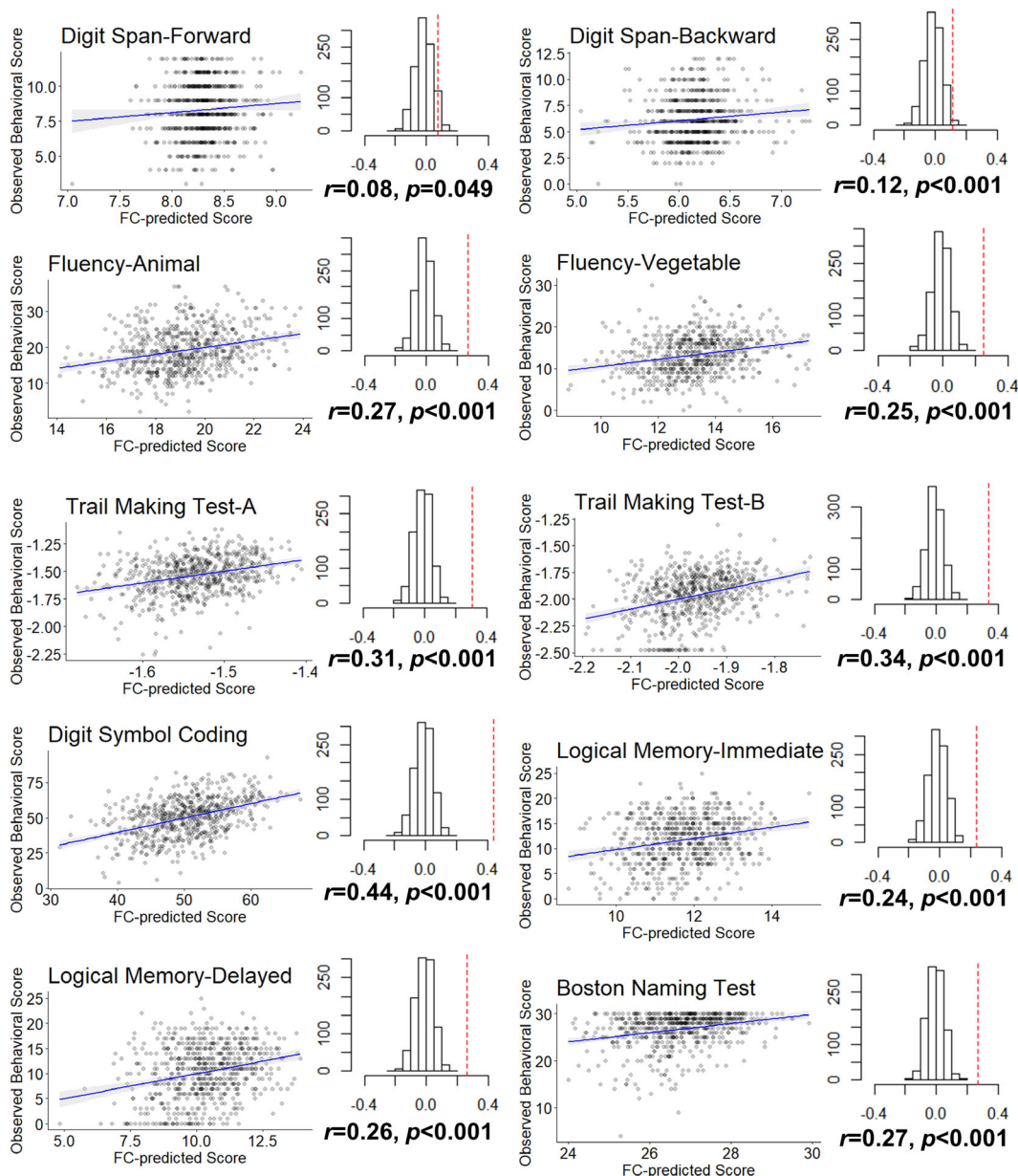


FIGURE 3 Internal cross-validated prediction result. Correlation between FC-predicted score and observed behavioral score in OASIS-3 ($n = 644$). The predicted scores were calculated based on three-fold cross-validation. Each of the folded datapoints (1/3) was calculated with the FC prediction model tuned and constructed within the other two folds of data (2/3). The permutation results are shown as the null distribution and the true predicted coefficient (dashed red line)

age-adjusted mean correlation = 0.179). The results also remained largely unchanged when adjusting for the effect of sex (mean correlation = 0.252) and head motion (mean correlation = 0.247). The confidence interval of the iterated prediction showed statistically significant predictability except for Digit Span tests (Figure S2).

3.2 | External cross-validation

We identified eight homologous neuropsychological tests that can be tested as an external dataset (Table 2). When conducting nested

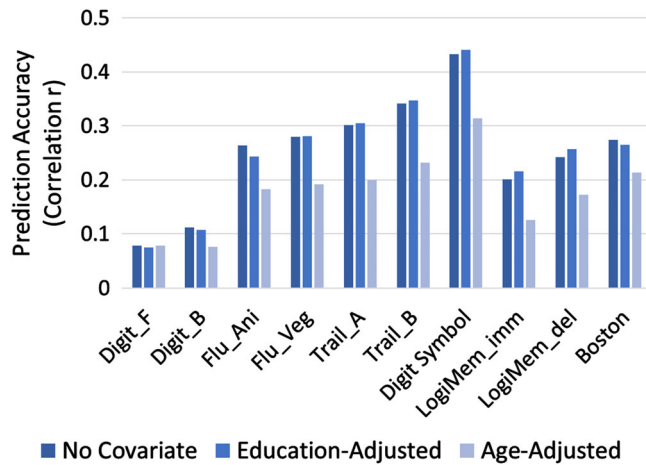


FIGURE 4 Prediction accuracy (correlation between FC-predicted score and observed NP score) in the test set ($n = 216$) when age and education effects are adjusted. The adjustments are conducted by residualizing age and education effect on the scores of the testing dataset. Mean correlation coefficients of 50 iterated prediction procedures are plotted. Digit_F, digit span forward; Digit_B, digit span backward; Flu_Ani/Veg, category fluency (animal/vegetable); Trail_A, Trail Making Test Part A; Trail_B, Trail Making Test Part B; LogiMem_imm, WMS-R Logical Memory I immediate recall; LogiMem_del, WMS-R Logical Memory II delayed recall; Boston, Boston Naming Test

cross-validation (Figure 2) across both internal (OASIS-3) and external (KSHAP) dataset showed a range of iterated prediction results (Figure 5). The result showed that internal cross-validation produced higher accuracy (mean $r = 0.226$) than external cross-validation (mean $r = 0.159$). Most of the correlation decreased whereas fluency (vegetable/supermarket) and Trail Making Test (A) showed preserved internal cross-validation's accuracy.

3.3 | Feature filtering

To examine the extent to which focal brain regions are attributable to the prediction result, we tested the effect of the feature filtering threshold on the prediction accuracy. By applying univariate tests to each connectivity feature, we filtered the features that correlated positively or negatively with the test scores higher than the threshold. We confirmed that increasing the filtering threshold resulted in a decrease in the features utilized in the training set (Figure S3). When examining the optimal filter threshold, we found that excluding features with a light threshold ($|r| < 0.06$) or applying no threshold ($|r| = 0.00$) showed the highest predictive accuracy in the test set (Figure 6). Therefore, maintaining a large proportion of the features (10,831–25,651 edges, 42–100%) in the ridge regression models produced the most accurate prediction. In contrast, however, several tests preserved their original accuracy even when the relatively strong threshold was applied. Excluding features with a high threshold ($|r| < 0.14$) minimally changed the predictive accuracy with the smaller number of features. We also confirmed a consistent result with the other accuracy metric (mean absolute error; Figure S4).

3.4 | Feature regularization

Similar to the feature filtering, we additionally examined the effect of the mixing parameter (α , proportion of L1-norm) that affects the

TABLE 2 Descriptive statistics of neuropsychological test performance in internal (OASIS-3) and external (KSHAP) validation dataset

	Internal validation set OASIS-3 ($n = 644$)			External validation set KSHAP ($n = 151$)			
	Mean	SD	Range	Mean	SD	Range	
Digit span forward	8.31	2.01	3–12	Digit span forward	7.64	2.16	4–14
Digit span backward	6.16	2.15	0–12	Digit span backward	4.39	2.12	1–14
Fluency animal	19.14	6.03	2–37	Fluency animal	13.72	4.21	3–30
Fluency vegetable	13.25	4.55	0–30	Fluency supermarket	15.76	6.41	3–34
Trail making test A	37.51	19.05	13–180	Modified Trail making test A	34.21	19.13	7–157
Trail making test B	106.43	63.95	20–300	Modified Trail making test C	132.64	79.37	25–522
Digit symbol coding	50.22	13.48	4–93				
Logical memory immediate	11.72	4.68	0–25	Story recall test immediate	13.44	6.76	0–28
Logical memory delayed	10.29	5.34	0–25	Story recall test delayed	11.48	6.89	0–28
Boston naming test	26.70	3.61	4–30				

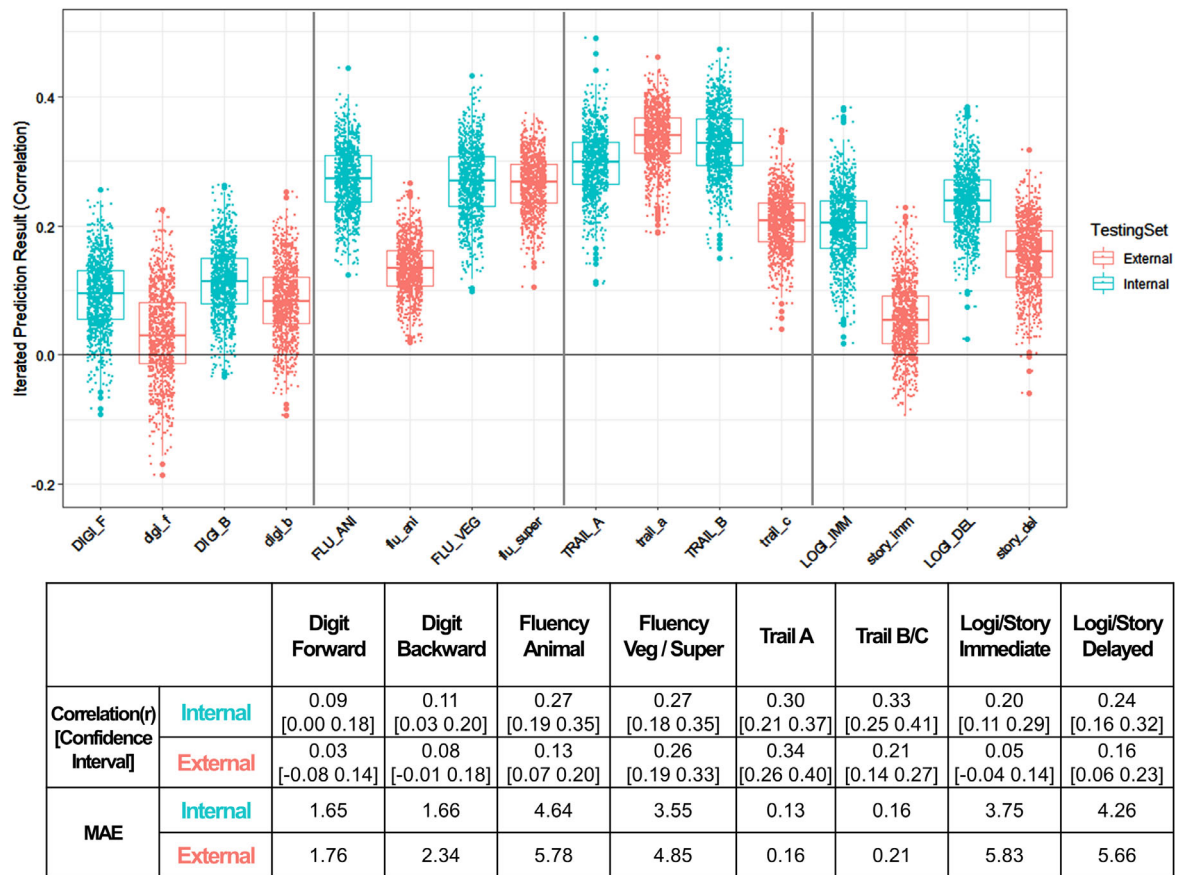


FIGURE 5 Prediction accuracy (correlation r between FC-predicted score and observed NP score) of corresponding neuropsychological test across internal (OASIS-3) and external (KSHAP) testing set. Model training: OASIS-3 ($n = 428$); Internal cross-validation: Holdout test dataset in OASIS-3 ($n = 216$). External cross-validation: KSHAP dataset ($n = 151$). Each dot indicates 50 times iterated results. DIGI_F/digi_f, digit span forward; DIGI_B/digi_b, digit span backward; FLU_ANI/flu_ani, category fluency (animal); FLU_VEG/flu_sto, category fluency (vegetable/supermarket); TRAIL_A/trail_a, Trail Making Test A; TRAIL_B/trail_c, Trail Making Test B/C; LOGIMEM_IMM/story_imm, Logical Memory/Story Recall Test immediate recall; LOGIMEM_DEL/story_del, Logical Memory/Story Recall Test delayed recall

model sparsity. If the α parameter increases, the penalized model produces highly sparse features with fewer predictors, whereas the lowest α parameter (ridge regression) leaves highly distributed weights of features. We initially confirmed that setting higher α left fewer sets of predictors (Figure S5). Consistent with feature filtering results, penalizing regression coefficients with increased α resulted in lower predictive accuracy.

3.5 | Leave-network-out prediction

To identify the brain functional networks that are critical to the prediction, we conducted a leave-network-out analysis. Similar to virtual lesion analysis, the predictive analysis was conducted after the connectivities of the designated functional network (features of within and between network) were excluded. The result showed that excluding specific functional networks (frontoparietal, default mode, subcortical/saliency) showed a larger decrease than other networks, but excluding a single network minimally affected the prediction accuracy in general (Figure 7).

3.6 | Single network prediction

When a single functional network was separately used as predictors, several functional networks achieved prediction accuracy similar to that of the full model (Figure S6). Digit Span Forward and Logical Memory-immediate tests were highly predictable when using frontal modules (medial frontal, default mode, or frontoparietal networks). On the other hand, Fluency-Vegetable, Digit Symbol Coding, Logical Memory-delayed, and Boston Naming tests were especially predictable using a relatively central part of the brain (Subcortical/Saliency and Sensorimotor). The posterior part (Visual, Visual Association) played a role especially in tests requiring motor and visuospatial speed (Trail Making Test A/B, Digit Symbol Coding), whereas Digit Span Forward, Logical Memory tests were minimally predictable with the posterior networks.

3.7 | Discriminant validity of predictive model

We also further examined whether the connectome-based predictive models specifically represent predictive information of the trained

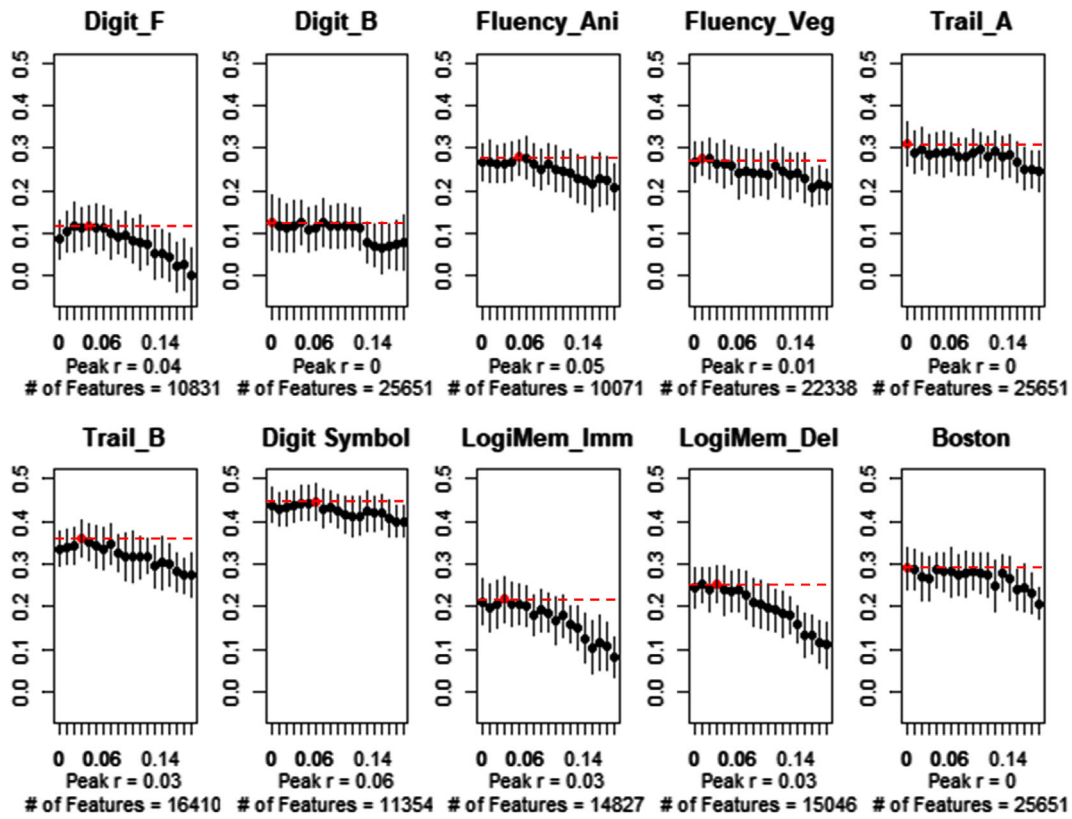


FIGURE 6 Predictive accuracy across feature filtering threshold ($|r| > .00, .01, \dots, .17, 0.18$). Error bar indicates SD of 50 times iteration results across thresholds. The highest peak accuracy points (red dots, dashed lines) and the number of connectivities that survived after filtering (averaged across iterations) are noted. Increasing the filter threshold left fewer features and produced lower predictive accuracy

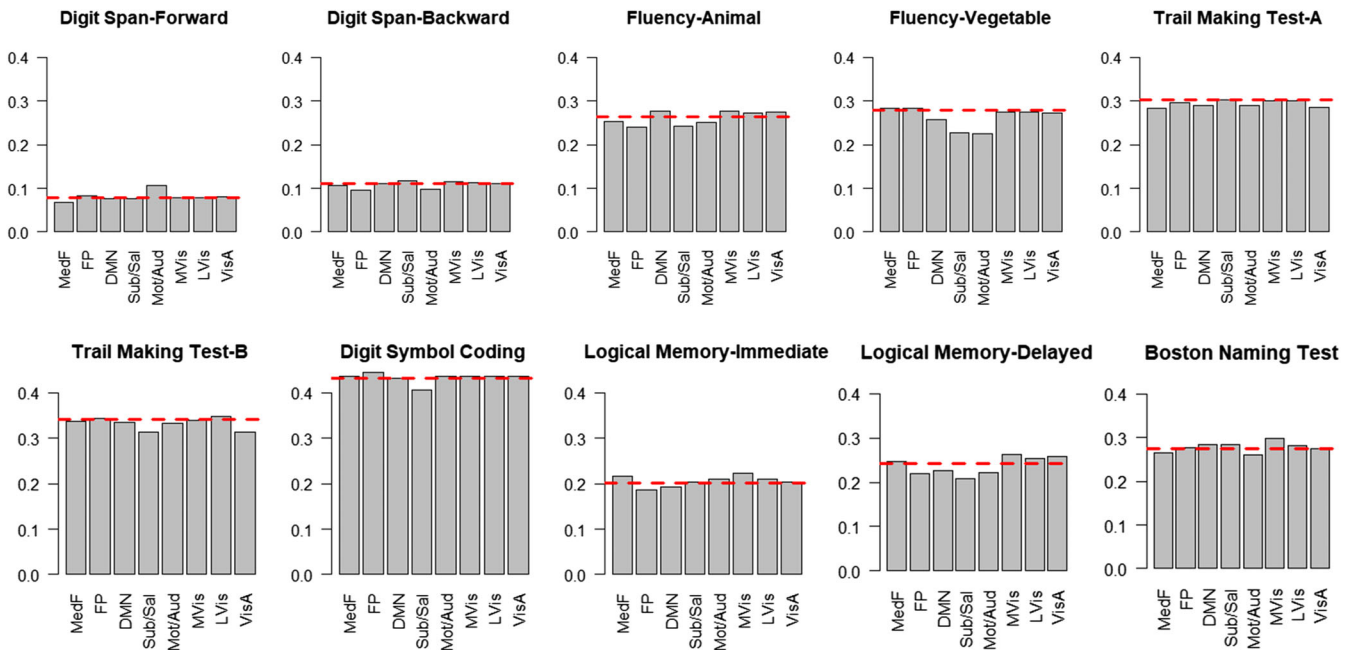


FIGURE 7 Prediction performance (averaged 50 times iteration) after excluding each functional network. Within (network-to-network) and between (network-to-all others) connectivities were excluded one at a time, and the same predictive modeling procedure was conducted. Red dashed line indicates full model performance using all of the functional networks (25,651 features). DMN, default mode network (3,915 edges); FP, frontoparietal (5,950 edges); LVis, lateral visual 2 (1,341 edges); MedF, medial frontal (6,148 edges); Mot/Aud, sensorimotor/auditory (9,898 edges); MVis, medial visual 1 (3,915 edges); Sub/Sal, subcortical/salience (12,121 edges); VisA, dorsal attention; visual association (3,706 edges)

test. Based on the predictive models constructed with the OASIS-3 training dataset ($n = 428$), FC-predicted scores were calculated in the OASIS-3 testing set and KSHAP. The correlations between eight of the predicted scores and eight corresponding behavioral scores were presented with the OASIS-3 dataset ($n = 216$; Figure 8, left, identical test) and KSHAP dataset ($n = 151$; Figure 8, right, homologous test). The result showed that some FC models (trained with Fluency – vegetable/supermarket and Trail Making Test A) were the most predictable to the corresponding test, whereas other test models were not specifically predictable to the corresponding tests. Rather, predictive models with generally high accuracy also tended to predict scores of other cognitive domains.

3.8 | Clinical validity of predictive model

Finally, we examined whether the brain predictive modeling approach captures clinically relevant information. We trained the predictive models in the KSHAP dataset ($n = 151$), and the models were applied to the OASIS-3 total dataset ($n = 644$). The prediction from the KSHAP model to the OASIS-3 dataset achieved a predictability pattern similar to the aforementioned analysis. Using these FC-predicted scores, we conducted a rank correlation analysis between dementia functional impairment score (CDR-SOB) and the neuropsychological performance (actual behavior score or FC-predicted score). Since the original measures of the cognitive test are highly predictive of the clinical impairment, a unique association of the predicted- or observed scores are assessed with partial correlation. The result showed that the FC-predicted score of Fluency (Supermarket) and Story/Logical Memory test (delayed) provided additive information when explaining the severity of clinical impairment (Table 3). Although FC-predicted

scores based on Fluency (Animal) and Trail Making Test (A) models were associated with the levels of clinical impairment, the other prediction scores did not remain statistically significant after adjusting for the effect of the actual behavioral performance score.

3.9 | Clinical and demographic stratification

We additionally examined whether inconsistent clinical status across training and testing samples was associated with prediction accuracy. When the same prediction analysis was confined within the clinically unimpaired group (CDR score = 0; training data: OASIS-3, $n = 288$), the prediction accuracy decreased both in the internal holdout sample (OASIS-3, $n = 148$) and external KSHAP sample ($n = 151$; Figure S7). This result shows that training a predictive model only with the clinically normal and homogeneous population did not enhance the predictive accuracy. Rather the predictability overall decreased, indicating relevance to the clinically impaired population in the predictive model.

We also examined how heterogeneity of educational attainment across internal (OASIS-3, mean education = 15.56) and external dataset (KSHAP, mean education = 7.17) systematically affects the prediction accuracy. We conducted an identical analysis procedure after stratifying the education group in the OASIS-3 sample. By training the predictive models with different education groups, we tested whether the model trained with similar educational attainment more accurately predicts neuropsychological performance in the external KSHAP dataset. Our result showed, however, that while the model trained with lower-educated older adults (7–12 years) showed relatively higher accuracy in Trail Making Test (A) and Logical Memory tests, the other tests did not show such tendencies (Figure S8).

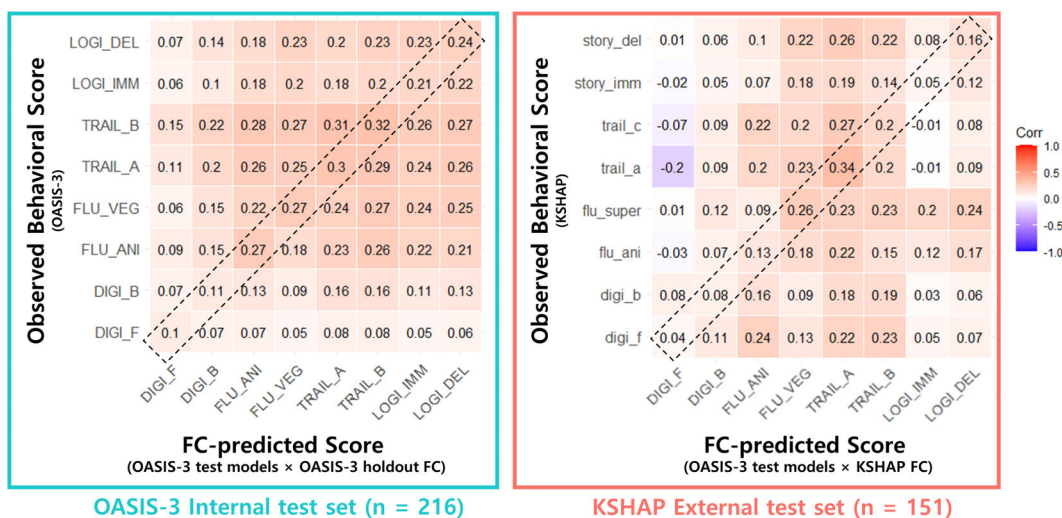


FIGURE 8 Discriminant validity of connectome-based prediction (averaged 50 times iteration). Correlation between FC-predicted and observed scores across eight subtests. X-axis: Specific test used to train predictive model and its predicted score in the test set. Y-axis: Observed neuropsychological test score in the test set. Diagonal elements within the dashed-box indicate prediction with the same (internal) or homologous (external) test models (e.g., First column: FC model trained with Digit Span Forward [DIGI_F] is used to test the correlation between FC-model predicted score and 8 behavioral scores)

TABLE 3 Incremental value as a clinical correlate

	Partial correlation between [observed score] and [CDR-SOB] (FC-predicted score controlled)		Partial correlation between [FC-predicted score] and [CDR-SOB] (observed score controlled)	
	Rho	p value	Rho	p value
Digit span forward	-0.156	7×10^{-5}	-0.012	.76
Digit span backward	-0.217	3×10^{-8}	-0.025	.53
Fluency (animal)	-0.368	5×10^{-22}	-0.077	.05
Fluency (vegetable/supermarket)	-0.418	1×10^{-28}	-0.128	1×10^{-3}
Trail making test A	-0.237	1×10^{-9}	-0.044	.26
Trail making test B/C	-0.379	2×10^{-23}	0.032	.42
Logical memory immediate	-0.438	2×10^{-31}	-0.048	.22
Logical memory delayed	-0.501	4×10^{-42}	-0.027	.50

Note: Spearman's rank partial correlation (rho) between the Observed score (real neuropsychological test performance), FC-Predicted score (scores estimated based on the connectivity-trained model), and clinical severity of dementia (CDR-SOB) in OASIS-3 total dataset ($n = 643$, missing = 1). Either the FC-predicted or observed score was controlled when examining the counterpart score's association. The FC-predicted score was estimated with the models trained with KSHAP homologous tests. A higher observed score indicates better behavioral performance. Higher CDR-SOB indicates a poorer clinical function. Bold value has significance threshold ($p < 5 \times 10^{-2}$)

4 | DISCUSSION

In this study, we identified patterns of brain connectivities that are predictive of neuropsychological test performances in older adults. The identified predictive model across functional networks and the best prediction model required most of the connectivity features (39–100%). In addition, the predictive models trained with cognitive tests were generalized to the external dataset with heterogeneous demographic features. The evidence of discriminant validity was unclear, suggesting that the predictive models were not distinctively predictive of the corresponding tests. Also, scores estimated from connectome-based models provided marginally additive information in explaining clinically relevant functional impairment. Our results demonstrate that late-life neuropsychological test performance can be formally characterized by multivariate connectome-based predictive models, and further translational evidence is needed when developing theoretically valid and clinically incremental predictive models.

In the current connectome-based predictive modeling, tests mainly representing executive function and processing speed (Fluency, Trail Making Test, and Digit Symbol Coding) showed relatively high predictive accuracy. Notably, these neuropsychological tests have consistently shown prognostic value in predicting the progression of dementia alongside episodic memory function (Amieva et al., 2014; Ewers et al., 2012; Mortamais et al., 2017; Younes et al., 2019). The tests that assess executive and processing speed function sensitively respond to subcortical and vascular lesions and subtle cognitive decline in these domains may reflect increased dementia risk through a somewhat independent source of typical Alzheimer's pathophysiology (Chouiter et al., 2016; Hedden et al., 2012; Jiang et al., 2018; Parks et al., 2011; Rabin et al., 2019). Accordingly, the decrease in the predictive accuracy was most prominent when excluding the subcortical network, especially in predicting vegetable fluency, Trail Making Test (B), and Digit Symbol Coding tests. These results indicate that some portion

of test performance requires necessary neural information from the subcortical network connectivities.

The tests representing attention and working memory function were less predictable with the connectome-based models. Possible reasons for low predictability can be discussed. First, contrary to the previous studies that measured attentional performance with extensive trials, the Digit Span test included in the battery may have lacked reliable information from only 12 trials. Second, Digit Span tests are relatively less sensitive to vascular or Alzheimer's diseases' pathophysiological processes (Belleville et al., 2014; Petersen et al., 2010; Ramirez-Gomez et al., 2017). It is possible that the reliability of predictive neural correlates may be largely driven by the test's relevance to the pathological process of late-life neurocognitive disorders.

Another possible explanation for the low predictive accuracy especially in Digit Span tests may be due to the nonordinal brain-behavior relationship. Unlike other neuropsychological tests that require homogeneous and repetitive trials, Digit Span test trials are incrementally ordered in terms of task loads. Although working memory tasks induce a parametric increase in neural response as a function of task load (i.e., the length of digits to maintain or manipulate), older adults undergoing age-related brain changes fail to show a typical parametric increase, indicating early arrival to their maximum neural capacity (Nagel et al., 2009; Schneider-Garces et al., 2010; Seghier & Price, 2018). As older adults utilize idiographic brain regions to compensate for the decreased neural inefficiency, the cognitive process during Digit Span tests may require neural features that are not uniform and ordinal (Carp, Gmeindl, & Reuter-Lorenz, 2010). In other words, the corresponding neural correlates that each trial reflects may qualitatively differ across span lengths (e.g., 4–5 vs 7–8). These complications have also been referred to as Simpson's paradox, suggesting that the brain-behavior correlation can even be reversed across the subgroups or the individuals composing the population (Cabeza et al., 2018; Kievit, Frankenhuis, Waldorp, & Borsboom, 2013). For

example, increased activity of the frontal cortex is one of the distinct features when comparing young adults and old adults, and it appears to be a beneficial sign for those who already underwent age-related changes (Cabeza, Anderson, Locantore, & McIntosh, 2002; Davis, Kragel, Madden, & Cabeza, 2012). Interestingly, our study showed that the predictive model of the Digit Span Forward test predicted the worse performance of Trail Making Test A in the external dataset ($r = -0.21$). This result suggests that the behavioral relevance of neural correlates can be easily reversed, and more fine-grained decomposition of individual differences is required in the future (Borsboom et al., 2016; Michell, 2012).

In the practice of connectome-based predictive modeling, identifying which brain regions constitute the models has been a challenging issue. By conducting a leave-network-out analysis, logically similar to hierarchical regression, we examined whether the necessary predictive information is embedded in the specific units of brain predictor (Yarkoni & Westfall, 2017). Consistent with the previous attempt to specify regional specificity of predictive models, excluding a large portion of functional networks (5–47% of edges) minimally affected the predictive performance, indicating that the individual difference of neuropsychological functions is not solely based on a specific functional network (Nielsen, Barch, Petersen, Schlaggar, & Greene, 2019; Rosenberg, Finn, Scheinost, Constable, & Chun, 2017). One notable previous study also had shown that cognitive impairment of associative domains (attention and memory function) could be better predicted with diverse contributions of functional networks. In contrast, less-associative domains (visual, somatomotor) can be better predicted with the locational information of stroke lesions (Siegel et al., 2016). In other words, if the neuropsychological test requires a collaboration of multiple cognitive modules, it becomes more unlikely to find a unitary neural predictor. In general, appears that neural disruption in multiple units of information processing modules (i.e., functional networks) combinatorially leads to behavioral impairment in the neuropsychological tests.

Moreover, when examining the optimal sparsity of the model, our results showed that utilizing a large proportion of the connectivity predictors in the model showed optimal predictive accuracy. Most of the test models required more than half of the total connectivities ($r > .01$ – $.06$; 10,071–25,651 features) in order to achieve the best accuracy. While some tests (Digit Span backward, Digit Symbol Coding) largely maintained the original predictability even when applying a stronger filter threshold, the decreasing tendency was consistent across all tests. This result indicates that previous studies capturing focal neural correlate with conservative multiple comparison corrections (i.e., familywise error rate) may have omitted behaviorally relevant neural features (Masouleh et al., 2019). If a researcher identifies neural correlates by applying a relatively strong univariate test threshold, such focal models may not fully represent the highly distributed nature of the neuropsychological functions. Also, the predictive model that selects sparse features (higher α regularization) showed poorer predictive accuracy. Consistent with previous studies, accurate individualized prediction of test performance requires a fine-grained

combination of neural features rather than highly selective neural features (Cui & Gong, 2018; Gao et al., 2019; Hatoum et al., 2019).

This viewpoint may seem to run counter with the extensive discussions that have been made on the highly specified functional roles of the brain networks (Buckner, Krienen, & Yeo, 2013; Fox & Raichle, 2007; Vaidya, Pujara, Petrides, Murray, & Fellows, 2019). If functional neuroimaging reflects available regional neural resources, it is plausible that the inter-individual difference mainly stems from the relevant brain areas (Lebreton, Bavard, Daunizeau, & Palminteri, 2019; Navon & Gopher, 1979; Poldrack, 2015). However, our brain-wide predictive models suggest that inter-individual difference in cognitive function is not only reflective of the focal brain characteristics but also of the distributed network systems (Fox, 2018; Sutterer & Tranel, 2017). This may be due to the high interdependency of functional connectivities. Even with small and focal brain lesions, the effect may result in the neural changes extending to adjacent connectivities (Fornito, Zalesky, & Breakspear, 2015; Gratton, Nomura, Pérez, & D'Esposito, 2012). This emergent property of the network may result in highly dispersed neural correlate of cognitive functions (Alstott, Breakspear, Hagmann, Cammoun, & Sporns, 2009; Stam, 2014). A growing body of research using other neuroimaging modalities also suggest that previously identified regional specificity of neurocognitive function (i.e., frontal lobe task), in fact, reflects distributed effects of inter-dependent neural systems (Burgess & Stuss, 2017; Cole, Yarkoni, Repovš, Anticevic, & Braver, 2012). Future work is needed to bridge the explanatory gap between functional connectivity and dynamic reorganization of neural states in performing cognitive tasks (Barbey, 2018; Gu et al., 2015; Shine et al., 2015). Our result suggests that only a single sensory modality, especially visual modal, is less predictive of the cognitive tests compared to the higher-order association system.

It should be noted, however, that focally selected connectivities also showed comparable predictive accuracies. For example, thresholding connectivities into a set of highly selective features showed preserved prediction accuracy, to some degree, in several tests (i.e., Digit Span Backward, Fluency, Trail Making A). This tendency was also described when a single functional network was used in the prediction. These results indicate that most of the individual differences can be represented with a relatively small proportion of important connectivities, and taking account for every other neural feature minimally adds to the total predictability. It is possible that neurodegenerative pathology is typically characterized by changes in brain regions with hub properties, and individual differences in topologically important regions may play a more significant role in predicting neuropsychological performance, especially in the early stage (Fischer, Wolf, & Fellgiebel, 2019).

In this study, we cross-validated the predictive model in both internal (holdout) and external (KSHAP) test dataset. We found that some neuropsychological tests (Trail Making Test A, Fluency – vegetable/supermarket) maintained similar predictive accuracy even when applied to the external study population. In contrast, most of the tests showed decreased predictive accuracy in the external dataset. This study provides novel evidence that the extent to which a brain-

behavior predictive model is replicable differs across specific tests. Although the matched sets of tests are highly homologous, even minor differences in test elements (e.g., digit number orders, semantic category, trail interference component, and story contents) may have significantly affected the qualitative attribute of the tests. This point warrants caution in developing replicable and generalizable brain predictive models. Although previous studies have verified predictive models with the internal cross-validation, they may not guarantee predictability in the novel dataset, and recent studies including both internal and external validation of connectome-based models have noted weak generalizability (Boeke, Holmes, & Phelps, 2019; Dinga et al., 2019; Fountain-Zaragoza, Samimy, Rosenberg, & Prakash, 2019; Yoo et al., 2018). Another possible factor that affects the generalizability is demographic heterogeneity. It has been well-documented that psychometric quality and the construct representativeness of the neuropsychological tests may differ across demographics and clinical status of the study population (Bertola et al., 2019; Mungas, Widaman, Reed, & Tomaszewski Farias, 2011; Siedlecki et al., 2010; Siedlecki, Honig, & Stern, 2008). Also, lifespan intellectual experience including educational attainment may alter brain-behavior relation (Franzmeier et al., 2016; O'Shea et al., 2018; Resende et al., 2018; Steffener et al., 2014). Further examination is needed to account for the modifiers of connectome-based models across stratified demographic characteristics (e.g., age range, years of education, and socioeconomic status).

In the previous studies, meaningful attempts were made using a similar predictive modeling approach (Lin et al., 2018; Meskaldji et al., 2016; Moradi, Hallikainen, Hänninen, & Tohka, 2017; Shim et al., 2017). While predictive patterns of neuropsychological test performance were often identified, studies typically summed various domains of functions into a total score or only examined a single domain (i.e., episodic memory function). Thus, the differential predictabilities across theoretically distinct cognitive domains were mostly unexamined. Identifying the generalizability or specificity of the predictive models will be crucial in developing formal theories of inter-subject neuromarker of cognitive functions (Jangraw et al., 2018; Rosenberg et al., 2015, 2018; Woo et al., 2017).

In examining the correlation between the predicted scores and the behavioral scores, the evidence of convergent and discriminant validity (i.e., the distinctiveness of model predictability) was unclear. While we hypothesized that each of the predictive models represents individual differences of the corresponding cognitive test that trained the model, we observed several scores that showed generally stronger predictability across other tests. This result may be due to the fact that a large proportion of neural correlate is based on the individual differences shared across cognitive domains. The latent factor structure of neuropsychological tests typically shows high correlations between distinct cognitive domains (Greenaway, Smith, Tangalos, Geda, & Ivnik, 2009; Park et al., 2012), and a relatively small portion of variance may be uniquely attributable to each cognitive test. It has been suggested that age-related decline in diverse domains of cognitive function are primarily driven by preceding changes in processing speed function (Salthouse, 1996a, 1996b), due to major late-life

neuropathological changes that are typically observed across diffuse brain areas (Fjell et al., 2014; Habes, Erus, et al., 2016). Accordingly, most of the identified brain-behavior relationship seems to be composed of common cognitive factors (S. Lee, Habeck, Razlighi, Salthouse, & Stern, 2016; Salthouse et al., 2015). Previous studies have shown that multivariate functional connectivity patterns tend to capture a strong covarying latent structure that commonly explains the variance of multiple cognitive functions (Perry et al., 2017; Smith et al., 2015). Although predictive modeling captures neural correlates of general cognitive functions, unique correlates of each domain can be masked if the given study population does not include subgroups that show heterogeneous cognitive impairment profiles (Delis, Jacobson, Bondi, Hamilton, & Salmon, 2003). The inclusion of various clinical subgroups may sharpen the unique specificity of each neuropsychological test (Corbetta et al., 2015; Sachdev et al., 2014).

This study also adds to a growing literature proposing that neuroimaging measures should be adaptable to translational goals (Arbabshirani, Plis, Sui, & Calhoun, 2017; Kapur, Phillips, & Insel, 2012; Woo et al., 2017). Regression models are suitable when neuropsychiatric disorders are essentially characterized by continuous dimensions of behavioral symptoms. In this way, the model can obtain subtle individual differences that would have been masked with binarized classification (Altman & Royston, 2006). Neurobehaviorally translated scores of the cognitive impairment may validly characterize behavioral significance and prognostic meaning of the preceding pathophysiological process of interest and play an important role in formalizing the dimensional construct of neurocognitive disorders (Bilder & Reise, 2019; Cuthbert & Insel, 2013). In the future, the predictive models will be able to gauge relative behavioral risk scores of dementia in generalized contexts similar to previously developed brain pathology risk scores (Habes, Erus, et al., 2016; Habes, Janowitz, et al., 2016). Nevertheless, our study result requires caution in applying predictive models to clinical usage. In the current study, while predictive models of category fluency and long-term episodic memory test captured incremental clinical information that the original score did not provide, the explanatory size was minimal and far smaller than behavioral scores. Despite the promising remarks of neuroimaging-based biomarkers in classifying diagnosis and prognosis of neuropsychiatric disorders (Orrù, Pettersson-Yeo, Marquand, Sartori, & Mechelli, 2012; Pellegrini et al., 2018; Rathore, Habes, Iftikhar, Shacklett, & Davatzikos, 2017), more evidence is needed to conclude that the identified neurobehavioral marker provides incrementally useful information (Jollans & Whelan, 2016; Moons et al., 2012).

There are several limitations to the current study that should be noted. First, it is mostly unknown whether the FC modality can comprehensively represent various neural processes. Although FC is advantageous in bridging the theoretical gap between dynamic cognitive process and pathophysiology (Brier, Thomas, & Ances, 2014; Fox, 2018; Mill, Ito, & Cole, 2017; Tavor et al., 2016), its macroscale approach should be further specified with more direct and pathophysiological processes relevant to the current and future neuropsychological function. Combining other modalities and imaging biomarkers in the predictive model will clarify how much shared and unique

behavioral features are ingrained in each modality (Hedden, Schultz, Rieckmann, Mormino, & Buckner, 2016; Pellegrini et al., 2018; Sui et al., 2018). Second, resting-state FC based on low-frequency fluctuation may not be an optimal modality that purely captures the focal neural phenomenon. Initial regional change may induce a cascade of FC responses (Fornito et al., 2015), and both increased and decreased functional connectivities may be a result of secondary compensatory reorganization processes (Chiesa et al., 2019; Reijmer et al., 2015; Schultz et al., 2017). Combining the primary neuropathological signs in the predictive models will significantly enhance the interpretation of predictive models (Siegel et al., 2016). Finally, the current method only utilizes the brain regional definition as a group-norm atlas, rather than a personalized topology. However, recent studies underscore the individual difference in functional parcellation and its relevance to the behavioral traits (Cui et al., 2020). Future studies are necessary for delineating whether the prediction results are due to the strength of connectivities rather than the individual differences in functional topography.

We also note the cautious implication of our approach with regards to the validity of behavioral test measures. While we critically evaluated the possible reasons why some test performances are not accurately represented in the patterns of functional connectivities, the predictability of a brain model itself does not ensure the validity of behavioral measures. Reliable and robust correspondence to neural correlate may indicate what biological properties the test measures, but the specific and focal correspondence of neural correlate may instead compromise the comprehensive utility of the cognitive test to describe the real-world activities (i.e., ecological validity; Bilder & Reise, 2019; Kessels, 2019). Both the neuroscientific correspondence and ecological validity of the recently developed neurobehavioral markers need to be assessed alongside existing behavioral instruments.

ACKNOWLEDGMENTS

Data were provided in part by OASIS-3 Principal Investigators: T. Benzinger, D. Marcus, J. Morris; National Institutes of Health (NIH), P50AG00561, P30NS09857781, P01AG026276, P01AG003991, R01A G043434, UL1TR000448, and R01EB009352. This research is supported by the National Research Foundation of Korea (NRF-2017S1A3A2067165), funded by the Ministry of Education, Science and Technology. We thank W. Ahn, S. Hahn, J. Lee, and C. Woo for the constructive and critical feedback.

CONFLICT OF INTEREST

We declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data available on request due to privacy/ethical restrictions.

REFERENCES

Alstott, J., Breakspear, M., Hagmann, P., Cammoun, L., & Sporns, O. (2009). Modeling the impact of lesions in the human brain. *PLoS Computational Biology*, 5(6), e1000408. <https://doi.org/10.1371/journal.pcbi.1000408>

- Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ*, 332(7549), 1080. <https://doi.org/10.1136/bmj.332.7549.1080>
- Amieva, H., Mokri, H., Le Goff, M., Meillon, C., Jacqmin-Gadda, H., Foubert-Samier, A., ... Dartigues, J. F. (2014). Compensatory mechanisms in higher-educated subjects with Alzheimer's disease: A study of 20 years of cognitive decline. *Brain*, 137(4), 1167–1175. <https://doi.org/10.1093/brain/awu035>
- An, H., & Chey, J. (2004). A standardization study of the story recall test in the elderly Korean population. *The Korean Journal of Clinical Psychology*, 23(2), 435–454.
- Arbabshirani, M. R., Plis, S., Sui, J., & Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *NeuroImage*, 145, 137–165. <https://doi.org/10.1016/j.neuroimage.2016.02.079>
- Arenaza-Urquijo, E. M., & Vemuri, P. (2018). Resistance vs resilience to Alzheimer disease. *Neurology*, 90(15), 695–703. <https://doi.org/10.1212/WNL.0000000000005303>
- Avery, E. W., Yoo, K., Rosenberg, M. D., Greene, A. S., Gao, S., Na, D. L., ... Chun, M. M. (2019). Distributed patterns of functional connectivity predict working memory performance in novel healthy and memory-impaired individuals. *Journal of Cognitive Neuroscience*, 32, 1–15. https://doi.org/10.1162/jocn_a_01487
- Barbey, A. K. (2018). Network neuroscience theory of human intelligence. *Trends in Cognitive Sciences*, 22(1), 8–20. <https://doi.org/10.1016/j.tics.2017.10.001>
- Bayram, E., Caldwell, J. Z. K., & Banks, S. J. (2018). Current understanding of magnetic resonance imaging biomarkers and memory in Alzheimer's disease. *Alzheimer's & Dementia*, 4, 395–413. <https://doi.org/10.1016/j.trci.2018.04.007>
- Belleveille, S., Fouquet, C., Duchesne, S., Collins, D. L., & Hudon, C. (2014). Detecting early preclinical Alzheimer's disease via cognition, neuropsychiatry, and neuroimaging: Qualitative review and recommendations for testing. *Journal of Alzheimer's Disease*, 42, S375–S382. <https://doi.org/10.3233/JAD-141470>
- Bertola, L., Wei-Ming Watson, C., Avila, J. F., Zahodne, L. B., Angevaere, M., Schupf, N., & Manly, J. J. (2019). Predictors of episodic memory performance across educational Strata: Multiple-group comparisons. *Journal of the International Neuropsychological Society*, 25(9), 901–909. <https://doi.org/10.1017/s1355617719000717>
- Bilder, R. M., & Reise, S. P. (2019). Neuropsychological tests of the future: How do we get there from here? *The Clinical Neuropsychologist*, 33(2), 220–245. <https://doi.org/10.1080/13854046.2018.1521993>
- Boeke, E. A., Holmes, A. J., & Phelps, E. A. (2019). Toward robust anxiety biomarkers: A machine learning approach in a large-scale sample. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 5, 799–807. <https://doi.org/10.1016/j.bpsc.2019.05.018>
- Borsboom, D., Rhemtulla, M., Cramer, A. O. J., van der Maas, H. L. J., Scheffer, M., & Dolan, C. V. (2016). Kinds versus continua: A review of psychometric approaches to uncover the structure of psychiatric constructs. *Psychological Medicine*, 46(8), 1567–1579. <https://doi.org/10.1017/S0033291715001944>
- Brier, M. R., Thomas, J. B., & Ances, B. M. (2014). Network dysfunction in Alzheimer's disease: Refining the disconnection hypothesis. *Brain Connectivity*, 4(5), 299–311. <https://doi.org/10.1089/brain.2014.0236>
- Buckner, R. L., Krienen, F. M., & Yeo, B. T. T. (2013). Opportunities and limitations of intrinsic functional connectivity MRI. *Nature Neuroscience*, 16(7), 832–837. <https://doi.org/10.1038/nn.3423>
- Burgess, P. W., & Stuss, D. T. (2017). Fifty years of prefrontal cortex research: Impact on assessment. *Journal of the International Neuropsychological Society*, 23(9–10), 755–767. <https://doi.org/10.1017/S1355617717000704>
- Bzdok, D., & Ioannidis, J. P. A. (2019). Exploration, inference, and prediction in neuroscience and biomedicine. *Trends in Neurosciences*, 42(4), 251–262. <https://doi.org/10.1016/j.tins.2019.02.001>

- Cabeza, R., Albert, M., Belleville, S., Craik, F. I. M., Duarte, A., Grady, C. L., ... Rajah, M. N. (2018). Maintenance, reserve and compensation: The cognitive neuroscience of healthy ageing. *Nature Reviews Neuroscience*, 19(11), 701–710. <https://doi.org/10.1038/s41583-018-0068-2>
- Cabeza, R., Anderson, N. D., Locantore, J. K., & McIntosh, A. R. (2002). Aging gracefully: Compensatory brain activity in high-performing older adults. *NeuroImage*, 17(3), 1394–1402. <https://doi.org/10.1006/nimg.2002.1280>
- Carp, J., Gmeindl, L., & Reuter-Lorenz, P. A. (2010). Age differences in the neural representation of working memory revealed by multi-voxel pattern analysis. *Frontiers in Human Neuroscience*, 4(217), 1–10. <https://doi.org/10.3389/fnhum.2010.00217>
- Charroud, C. C. C. C., Le Bars, E., Deverdun, J. J. J. J., Steffener, J., Molino, F. F. F. F., Abdennour, M., ... Menjot de Champfleury, N. (2016). Working memory performance is related to intrinsic resting state functional connectivity changes in community-dwelling elderly cohort. *Neurobiology of Learning and Memory*, 132, 57–66. <https://doi.org/10.1016/j.nlm.2016.05.008>
- Chey, J. (2007). *Elderly memory disorder scale*. Seoul: Hakjisa.
- Chiesa, P. A., Cavedo, E., Vergallo, A., Lista, S., Potier, M. C., Habert, M. O., ... Younesi, E. (2019). Differential default mode network trajectories in asymptomatic individuals at risk for Alzheimer's disease. *Alzheimer's & Dementia*, 15, 940–950. <https://doi.org/10.1016/j.jalz.2019.03.006>
- Chouiter, L., Holmberg, J., Manuel, A. L., Colombo, F., Clarke, S., Annoni, J. M., & Spierer, L. (2016). Partly segregated cortico-subcortical pathways support phonologic and semantic verbal fluency: A lesion study. *Neuroscience*, 329, 275–283. <https://doi.org/10.1016/j.neuroscience.2016.05.029>
- Cole, M. W., Yarkoni, T., Repovš, G., Anticevic, A., & Braver, T. S. (2012). Global connectivity of prefrontal cortex predicts cognitive control and intelligence. *The Journal of Neuroscience*, 32(26), 8988–8999. <https://doi.org/10.1523/JNEUROSCI.0536-12.2012>
- Corbetta, M., Ramsey, L., Callejas, A., Baldassarre, A., Hacker, C. D., Siegel, J. S., ... Shulman, G. L. (2015). Common behavioral clusters and subcortical anatomy in stroke. *Neuron*, 85(5), 927–941. <https://doi.org/10.1016/j.neuron.2015.02.027>
- Cui, Z., & Gong, G. (2018). The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *NeuroImage*, 178(May), 622–637. <https://doi.org/10.1016/j.neuroimage.2018.06.001>
- Cui, Z., Li, H., Xia, C. H., Larsen, B., Adebimpe, A., Baum, G. L., ... Satterthwaite, T. D. (2020). Individual variation in functional topography of association networks in youth. *Neuron*, 106, 340–353.e8. <https://doi.org/10.1016/j.neuron.2020.01.029>
- Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine*, 11(1), 126. <https://doi.org/10.1186/1741-7015-11-126>
- Dadi, K., Rahim, M., Abraham, A., Chyzyk, D., Milham, M., Thirion, B., & Varoquaux, G. (2019). Benchmarking functional connectome-based predictive models for resting-state fMRI. *NeuroImage*, 192, 115–134. <https://doi.org/10.1016/j.neuroimage.2019.02.062>
- Davis, S. W., Kragel, J. E., Madden, D. J., & Cabeza, R. (2012). The architecture of cross-hemispheric communication in the aging brain: Linking behavior to functional and structural connectivity. *Cerebral Cortex*, 22(1), 232–242. <https://doi.org/10.1093/cercor/bhr123>
- Delis, D. C., Jacobson, M., Bondi, M. W., Hamilton, J. M., & Salmon, D. P. (2003). The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: Lessons from memory assessment. *Journal of the International Neuropsychological Society*, 9, 936–946. <https://doi.org/10.1017/S1355617703960139>
- Dinga, R., Schmaal, L., Penninx, B. W. J. H., van Tol, M. J., Veltman, D. J., van Velzen, L., ... Marquand, A. F. (2019). Evaluating the evidence for biotypes of depression: Methodological replication and extension of. *NeuroImage: Clinical*, 22, 101796. <https://doi.org/10.1016/j.nicl.2019.101796>
- Donders, J. (2019). The incremental value of neuropsychological assessment: A critical review. *The Clinical Neuropsychologist*, 34, 1–32. <https://doi.org/10.1080/13854046.2019.1575471>
- Dubois, J., & Adolphs, R. (2016). Building a science of individual differences from fMRI. *Trends in Cognitive Sciences*, 20(6), 425–443. <https://doi.org/10.1016/j.tics.2016.03.014>
- Dubois, J., Galdi, P., Paul, L. K., & Adolphs, R. (2018). A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 373(1756), 20170284. <https://doi.org/10.1098/rstb.2017.0284>
- Duchek, J. M., Balota, D., Thomas, J. B., Snyder, A. Z., Rich, P., Benzinger, T. L., ... Ances, B. M. (2013). Relationship between Stroop performance and resting state functional connectivity in cognitively normal older adults. *Neuropsychology*, 27(5), 516–528. <https://doi.org/10.1037/a0033402>
- Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, 14(1), 91–118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037>
- Ewers, M., Walsh, C., Trojanowski, J. Q., Shaw, L. M., Petersen, R. C., Jack, C. R., ... Hampel, H. (2012). Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiology of Aging*, 33(7), 1203–1214.e2. <https://doi.org/10.1016/j.neurobiolaging.2010.10.019>
- Ferreira, L. K., & Busatto, G. F. (2013). Resting-state functional connectivity in normal brain aging. *Neuroscience and Biobehavioral Reviews*, 37(3), 384–400. <https://doi.org/10.1016/j.neubiorev.2013.01.017>
- Fields, J. A., Ferman, T. J., Boeve, B. F., & Smith, G. E. (2011). Neuropsychological assessment of patients with dementing illness. *Nature Reviews Neurology*, 7(12), 677–687. <https://doi.org/10.1038/nrneurol.2011.173>
- Fields, J. A., Machulda, M., Aakre, J., Ivnik, R. J., Boeve, B. F., Knopman, D. S., ... Smith, G. E. (2010). Utility of the drs for predicting problems in day-to-day functioning. *The Clinical Neuropsychologist*, 24(7), 1167–1180. <https://doi.org/10.1080/13854046.2010.514865>
- Fillenbaum, G. G., Peterson, B., & Morris, J. (1996). Estimating the validity of the clinical Dementia Rating Scale: the CERAD experience. Consortium to Establish a Registry for Alzheimer's Disease. *Aging (Milan, Italy)*, 8(6), 379–385.
- Fischer, F. U., Wolf, D., & Fellgiebel, A. (2019). Connectivity and morphology of hubs of the cerebral structural connectome are associated with brain resilience in AD- and age-related pathology. *Brain Imaging and Behavior*, 13, 1650–1664. <https://doi.org/10.1007/s11682-019-00090-y>
- Fjell, A. M., McEvoy, L., Holland, D., Dale, A. M., Walhovd, K. B., & ADNI. (2014). What is normal in normal aging? Effects of aging, amyloid and Alzheimer's disease on the cerebral cortex and the hippocampus. *Progress in Neurobiology*, 117(2014), 20–40. <https://doi.org/10.1016/j.pneurobio.2014.02.004>
- Fornito, A., Zalesky, A., & Breakspear, M. (2015). The connectomics of brain disorders. *Nature Reviews Neuroscience*, 16(3), 159–172. <https://doi.org/10.1038/nrn3901>
- Fountain-Zaragoza, S., Samimy, S., Rosenberg, M. D., & Prakash, R. S. (2019). Connectome-based models predict attentional control in aging adults. *NeuroImage*, 186, 1–13. <https://doi.org/10.1016/j.neuroimage.2018.10.074>
- Fox, M. D. (2018). Mapping symptoms to brain networks with the human Connectome. *New England Journal of Medicine*, 379(23), 2237–2245. <https://doi.org/10.1056/NEJMra1706158>
- Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature*

- Reviews. *Neuroscience*, 8(9), 700–711. <https://doi.org/10.1038/nrn2201>
- Franzmeier, N., Buerger, K., Teipel, S., Stern, Y., Dichgans, M., Ewers, M., & Alzheimer's Disease Neuroimaging Initiative (ADNI). (2016). Cognitive reserve moderates the association between functional network anti-correlations and memory in MCI. *Neurobiology of Aging*, 50, 152–162. <https://doi.org/10.1016/j.neurobiolaging.2016.11.013>
- Gao, S., Greene, A. S., Constable, R. T., & Scheinost, D. (2019). Combining multiple connectomes improves predictive modeling of phenotypic measures. *NeuroImage*, 201, 116038. <https://doi.org/10.1016/j.neuroimage.2019.116038>
- Genon, S., Reid, A., Langner, R., Amunts, K., & Eickhoff, S. B. (2018). How to characterize the function of a brain region. *Trends in Cognitive Sciences*, 22(4), 350–364. <https://doi.org/10.1016/j.tics.2018.01.010>
- Gratton, C., Nomura, E. M., Pérez, F., & D'Esposito, M. (2012). Focal brain lesions to critical locations cause widespread disruption of the modular Organization of the Brain. *Organization for Human Brain Mapping*, 24(6), 1275–1285. https://doi.org/10.1162/jocn_a_00222
- Greenaway, M. C., Smith, G. E., Tangalos, E. G., Geda, Y. E., & Ivnik, R. J. (2009). Mayo older Americans normative studies: Factor analysis of an expanded neuropsychological battery. *The Clinical Neuropsychologist*, 23(1), 7–20. <https://doi.org/10.1080/13854040801891686>
- Greene, A. S., Gao, S., Scheinost, D., & Constable, R. T. (2018). Task-induced brain state manipulation improves prediction of individual traits. *Nature Communications*, 9(1), 2807. <https://doi.org/10.1038/s41467-018-04920-3>
- Gu, S., Pasqualetti, F., Cieslak, M., Telesford, Q. K., Yu, A. B., Kahn, A. E., ... Bassett, D. S. (2015). Controllability of structural brain networks. *Nature Communications*, 6, 8414. <https://doi.org/10.1038/ncomms9414>
- Habeck, C., Razlighi, Q., Gazes, Y., Barulli, D., Steffener, J., & Stern, Y. (2016). Cognitive reserve and brain maintenance: Orthogonal concepts in theory and practice. *Cerebral Cortex*, 27, 3962–3969. <https://doi.org/10.1093/cercor/bhw208>
- Habeck, C., Steffener, J., Barulli, D., Gazes, Y., Razlighi, Q., Shaked, D., ... Stern, Y. (2015). Making cognitive latent variables manifest: Distinct neural networks for fluid reasoning and processing speed. *Journal of Cognitive Neuroscience*, 27(6), 1249–1258. https://doi.org/10.1162/jocn_a_00778
- Habes, M., Erus, G., Toledo, J. B., Zhang, T., Bryan, N., Launer, L. J., ... Davatzikos, C. (2016). White matter hyperintensities and imaging patterns of brain ageing in the general population. *Brain*, 139, 1164–1179. <https://doi.org/10.1093/brain/aww008>
- Habes, M., Janowitz, D., Erus, G., Toledo, J. B., Resnick, S. M., Doshi, J., ... Davatzikos, C. (2016). Advanced brain aging: Relationship with epidemiologic and genetic risk factors, and overlap with Alzheimer disease atrophy patterns. *Translational Psychiatry*, 6(September 2015), e775. <https://doi.org/10.1038/tp.2016.39>
- Han, J. W., Kim, T. H., Jhoo, J. H., Park, J. H., Kim, J. L., Ryu, S. H., ... Kim, K. W. (2010). A normative study of the mini-mental state examination for dementia screening (MMSE-DS) and its short form (SMMSE-DS) in the Korean elderly. *Journal of Korean Geriatric Psychiatry*, 14(1), 27–37.
- Hastie, T., J Qian, & Qian, J. (2016). Glmnet vignette. Retrieved from http://www.web.stanford.edu/~Hastie/Papers/Glmnet_Vignette.Pdf; <http://cran.us.r-project.org>.
- Hatoum, A. S., Reineberg, A. E., Kragel, P. A., Keller, M. C., Wager, T. D., Hewitt, J. K., & Friedman, N. P. (2019). Inferring the genetic influences on psychological traits using MRI connectivity predictive models: Demonstration with cognition. *BioRxiv*, 2019, 777821. <https://doi.org/10.1101/777821>
- Hayden, K. M., Jones, R. N., Zimmer, C., Plassman, B. L., Browndyke, J. N., Pieper, C., ... Welsh-Bohmer, K. A. (2011). Factor structure of the national Alzheimer's coordinating centers dataset neuropsychological battery: An evaluation of invariance between and within groups over time. *Alzheimer Disease and Associated Disorders*, 25, 128–137. <https://doi.org/10.1097/WAD.0b013e3181ffa76d>
- Hedden, T., Mormino, E. C., Amariglio, R. E., Younger, A. P., Schultz, A. P., Becker, J. A., ... Rentz, D. M. (2012). Cognitive profile of amyloid burden and white matter Hyperintensities in cognitively Normal older adults. *Journal of Neuroscience*, 32(46), 16233–16242. <https://doi.org/10.1523/JNEUROSCI.2462-12.2012>
- Hedden, T., Schultz, A. P., Rieckmann, A., Mormino, E. C., Johnson, K., Sperling, R., & Buckner, R. L. (2016). Multiple brain markers are linked to age-related variation in cognition. *Cerebral Cortex*, 26(4), 1388–1400. <https://doi.org/10.1093/cercor/bhu238>
- Hohman, T. J., McLaren, D. G., Mormino, E. C., Gifford, K. A., Libon, D. J., Jefferson, A. L., & Alzheimer's Disease Neuroimaging Initiative. (2016). Asymptomatic Alzheimer disease: Defining resilience. *Neurology*, 87, 2443–2450. <https://doi.org/10.1212/WNL.0000000000003397>
- Jack, C. R., Wiste, H. J., Therneau, T. M., Weigand, S. D., Knopman, D. S., Mielke, M. M., ... Petersen, R. C. (2019). Associations of amyloid, tau, and Neurodegeneration biomarker profiles with rates of memory decline among individuals without dementia. *JAMA: The Journal of the American Medical Association*, 321, 2316–2325. <https://doi.org/10.1001/jama.2019.7437>
- Jangraw, D. C., Gonzalez-Castillo, J., Handwerker, D. A., Ghane, M., Rosenberg, M. D., Panwar, P., & Bandettini, P. A. (2018). A functional connectivity-based neuromarker of sustained attention generalizes to predict recall in a reading task. *NeuroImage*, 166(October 2017), 99–109. <https://doi.org/10.1016/j.neuroimage.2017.10.019>
- Jiang, J., Paradise, M., Liu, T., Armstrong, N. J., Zhu, W., Kochan, N. A., ... Wen, W. (2018). The association of regional white matter lesions with cognition in a community-based cohort of older individuals. *NeuroImage: Clinical*, 19, 14–21. <https://doi.org/10.1016/j.nicl.2018.03.035>
- Jollans, L., & Whelan, R. (2016). The clinical added value of imaging: A perspective from outcome prediction. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 1(5), 423–432. <https://doi.org/10.1016/j.bpsc.2016.04.005>
- Jollans, L., & Whelan, R. (2018). Neuromarkers for mental disorders: Harnessing population neuroscience. *Frontiers in Psychiatry*, 9, 1–14. <https://doi.org/10.3389/fpsy.2018.00242>
- Kang, Y., Chin, J., Na, D. L., Lee, J., & Park, J. S. (2000). A normative study of the Korean version of controlled Oral word association test (COWAT) in the elderly. *Korean Journal of Clinical Psychology*, 19(2), 385–392.
- Kang, Y., Jang, S., & Na, D. (2012). *Seoul neuropsychological screening battery (SNSB)*. Seoul: Human Brain Research and Consulting.
- Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *The Boston naming test*. Philadelphia: Lea and Febiger.
- Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular Psychiatry*, 17(12), 1174–1179. <https://doi.org/10.1038/mp.2012.105>
- Kessels, R. P. C. (2019). Improving precision in neuropsychological assessment: Bridging the gap between classic paper-and-pencil tests and paradigms from cognitive neuroscience. *The Clinical Neuropsychologist*, 33(2), 357–368. <https://doi.org/10.1080/13854046.2018.1518489>
- Kievit, R. A., Frankenhuys, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology*, 4, 513. <https://doi.org/10.3389/fpsyg.2013.00513>
- Kim, S. (2015). ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Communications for Statistical Applications and Methods*, 22(6), 665–674. <https://doi.org/10.5351/csam.2015.22.6.665>
- Laird, A. R., Eickhoff, S. B., Rottschy, C., Bzdok, D., Ray, K. L., & Fox, P. T. (2013). Networks of task co-activations. *NeuroImage*, 80, 505–514. <https://doi.org/10.1016/j.neuroimage.2013.04.073>
- LaMontagne, P. J., Keefe, S., Lauren, W., Xiong, C., Grant, E. A., Moulder, K. L., ... Marcus, D. S. (2018). OASIS-3: Longitudinal

- neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer's disease. *Alzheimer's & Dementia*, 14(7), P1097. <https://doi.org/10.1016/j.jalz.2018.06.1439>
- Lebreton, M., Bavard, S., Daunizeau, J., & Palminteri, S. (2019). Assessing inter-individual differences with task-related functional neuroimaging. *Nature Human Behaviour*, 3, 897–905. <https://doi.org/10.1038/s41562-019-0681-8>
- Lee, J., Lee, W. J., Kim, H. C., Choi, W., Lee, J., Sung, K., ... Youm, Y. (2014). The Korean social life, health and aging project-health examination cohort. *Epidemiology and Health*, 36, e2014003. <https://doi.org/10.4178/epih/e2014003>
- Lee, S., Habeck, C., Razlighi, Q., Salthouse, T., & Stern, Y. (2016). Selective association between cortical thickness and reference abilities in normal aging. *NeuroImage*, 142, 293–300. <https://doi.org/10.1016/j.neuroimage.2016.06.041>
- Lin, Q., Rosenberg, M. D., Yoo, K., Hsu, T. W., O'Connell, T. P., & Chun, M. M. (2018). Resting-state functional connectivity predicts cognitive impairment related to Alzheimer's Disease. *Frontiers in Aging Neuroscience*, 10(April), 1–10. <https://doi.org/10.3389/fnagi.2018.00094>
- Lynch, C. A., Walsh, C., Blanco, A., Moran, M., Coen, R. F., Walsh, J. B., & Lawlor, B. A. (2005). The clinical dementia rating sum of box score in mild dementia. *Dementia and Geriatric Cognitive Disorders*, 21, 40–43. <https://doi.org/10.1159/000089218>
- Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2007). Open access series of imaging studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience*, 19(9), 1498–1507. <https://doi.org/10.1162/jocn.2007.19.9.1498>
- Masouleh, K. S., Eickhoff, S. B., Hoffstaedter, F., & Genon, S. (2019). Empirical examination of the replicability of associations between brain structure and psychological variables. *eLife*, 8, 1–25. <https://doi.org/10.7554/eLife.43464>
- Meskaldji, D. E., Preti, M. G., Bolton, T. A., Montandon, M. L., Rodriguez, C., Morgenthaler, S., ... Van De Ville, D. (2016). Prediction of long-term memory scores in MCI based on resting-state fMRI. *NeuroImage: Clinical*, 12, 785–795. <https://doi.org/10.1016/j.nicl.2016.10.004>
- Michell, J. (2012). Alfred Binet and the concept of heterogeneous orders. *Frontiers in Psychology*, 3, 261. <https://doi.org/10.3389/fpsyg.2012.00261>
- Mill, R. D., Ito, T., & Cole, M. W. (2017). From connectome to cognition: The search for mechanism in human functional brain networks. *NeuroImage*, 160(January), 124–139. <https://doi.org/10.1016/j.neuroimage.2017.01.060>
- Moms, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., ... Clark, C. (1989). The consortium to establish a registry for Alzheimer's Disease (CERAD). Part I. clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, 39(9), 1159–1159. <https://doi.org/10.1212/WNL.39.9.1159>
- Moons, K. G. M. M., Kengne, A. P., Woodward, M., Royston, P., Vergouwe, Y., Altman, D. G., & Grobbee, D. E. (2012). Risk prediction models: I. development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*, 98(9), 683–690. <https://doi.org/10.1136/heartjnl-2011-301246>
- Moradi, E., Hallikainen, I., Hänninen, T., & Tohka, J. (2017). Rey's auditory verbal learning test scores can be predicted from whole brain MRI in Alzheimer's disease. *NeuroImage: Clinical*, 13, 415–427. <https://doi.org/10.1016/j.nicl.2016.12.011>
- Morris, J. (1993). The clinical dementia rating (CDR): Current version and scoring rules. *Neurology*, 43(11), 2412–2414.
- Morris, J. C., Ernesto, C., Schafer, K., Coats, M., Leon, S., Sano, M., ... Woodbury, P. (1997). Clinical dementia rating training and reliability in multicenter studies: The Alzheimer's Disease cooperative study experience. *Neurology*, 48(6), 1508–1510. <https://doi.org/10.1212/WNL.48.6.1508>
- Morris, J. C. (1997). Clinical dementia rating: A reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *International Psychogeriatrics*, 9(S1), 173–176. <https://doi.org/10.1017/S1041610297004870>
- Mortamais, M., Ash, J. A., Harrison, J., Kaye, J., Kramer, J., Randolph, C., ... Ritchie, K. (2017). Detecting cognitive changes in preclinical Alzheimer's disease: A review of its feasibility. *Alzheimer's & Dementia*, 13(4), 468–492. <https://doi.org/10.1016/j.jalz.2016.06.2365>
- Mungas, D., Widaman, K. F., Reed, B. R., & Tomaszewski Farias, S. (2011). Measurement invariance of neuropsychological tests in diverse older persons. *Neuropsychology*, 25(2), 260–269. <https://doi.org/10.1037/a0021090>
- Nagel, I. E., Preuschhof, C., Li, S.-C., Nyberg, L., Bäckman, L., Lindenberger, U., & Heekeren, H. R. (2009). Performance level modulates adult age differences in brain activation during spatial working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 106(52), 22552–22557. <https://doi.org/10.1073/pnas.0908238106>
- Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, 86(3), 214–255. <https://doi.org/10.1037/0033-295X.86.3.214>
- Nielsen, A. N., Barch, D. M., Petersen, S. E., Schlaggar, B. L., & Greene, D. J. (2019). Machine learning with neuroimaging: Evaluating its applications in psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 5, 791–798. <https://doi.org/10.1016/j.bpsc.2019.11.007>
- Nostro, A. D., Müller, V. I., Varikuti, D. P., Pläschke, R. N., Hoffstaedter, F., Langner, R., ... Eickhoff, S. B. (2018). Predicting personality from network-based resting-state functional connectivity. *Brain Structure and Function*, 223(6), 2699–2719. <https://doi.org/10.1007/s00429-018-1651-z>
- O'Shea, D. M., Langer, K., Woods, A. J., Porges, E. C., Williamson, J. B., O'Shea, A., & Cohen, R. A. (2018). Educational attainment moderates the association between hippocampal volumes and memory performances in healthy older adults. *Frontiers in Aging Neuroscience*, 10(361), 1–9. <https://doi.org/10.3389/fnagi.2018.00361>
- Orrù, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., & Mechelli, A. (2012). Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neuroscience and Biobehavioral Reviews*, 36(4), 1140–1152. <https://doi.org/10.1016/j.neubiorev.2012.01.004>
- Park, L. Q., Gross, A. L., McLaren, D. G., Pa, J., Johnson, J. K., Mitchell, M., & Manly, J. J. (2012). Confirmatory factor analysis of the ADNI neuropsychological battery. *Brain Imaging and Behavior*, 6(4), 528–539. <https://doi.org/10.1007/s11682-012-9190-3>
- Park, M., & Chey, J. (2003). A normative study of the modified trail making test for elderly Korean people. *Korean Journal of Clinical Psychology*, 22(1), 247–259.
- Parks, C. M., Iosif, A. M., Farias, S., Reed, B., Mungas, D., & DeCarli, C. (2011). Executive function mediates effects of white matter hyperintensities on episodic memory. *Neuropsychologia*, 49(10), 2817–2824. <https://doi.org/10.1016/j.neuropsychologia.2011.06.003>
- Pellegrini, E., Ballerini, L., del Hernandez, M. C. V., Chappell, F. M., González-Castro, V., Anblagan, D., ... Wardlaw, J. M. (2018). Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review. *Alzheimer's and Dementia*, 10, 519–535. <https://doi.org/10.1016/j.dadm.2018.07.004>
- Perry, A., Wen, W., Kochan, N. A., Thalamuthu, A., Sachdev, P. S., & Breakspear, M. (2017). The independent influences of age and education on functional brain networks and cognition in healthy older adults. *Human Brain Mapping*, 38(10), 5094–5114. <https://doi.org/10.1002/hbm.23717>

- Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., ... Weiner, M. W. (2010). Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology*, *74*(3), 201–209. <https://doi.org/10.1212/WNL.0b013e3181cb3e25>
- Poldrack, R. A. (2015). Is "efficiency" a useful concept in cognitive neuroscience? *Developmental Cognitive Neuroscience*, *11*, 12–17. <https://doi.org/10.1016/j.dcn.2014.06.001>
- Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., ... Petersen, S. E. (2011). Functional network Organization of the Human Brain. *Neuron*, *72*(4), 665–678. <https://doi.org/10.1016/j.neuron.2011.09.006>
- Rabin, J. S., Perea, R. D., Buckley, R. F., Neal, T. E., Buckner, R. L., Johnson, K. A., ... Hedden, T. (2019). Global white matter diffusion characteristics predict longitudinal cognitive change independently of amyloid status in clinically Normal older adults. *Cerebral Cortex*, *29*(3), 1251–1262. <https://doi.org/10.1093/cercor/bhy031>
- Ramirez-Gomez, L., Zheng, L., Reed, B., Kramer, J., Mungas, D., Zarow, C., ... Chui, H. (2017). Neuropsychological profiles differentiate Alzheimer Disease from subcortical ischemic vascular dementia in an autopsy-defined cohort. *Dementia and Geriatric Cognitive Disorders*, *44*, 1–11. <https://doi.org/10.1159/000477344>
- Rathore, S., Habes, M., Iftikhar, M. A., Shacklett, A., & Davatzikos, C. (2017). A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *NeuroImage*, *155*(March), 530–548. <https://doi.org/10.1016/j.neuroimage.2017.03.057>
- Reijmer, Y. D., Schultz, A. P., Leemans, A., O'Sullivan, M. J., Gurol, M. E., Sperling, R., ... Hedden, T. (2015). Decoupling of structural and functional brain connectivity in older adults with white matter hyperintensities. *NeuroImage*, *117*, 222–229. <https://doi.org/10.1016/j.neuroimage.2015.05.054>
- Reitan, R. M., & Wolfson, D. (1993). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation* (2nd ed.). Tucson: Neuropsychology Press.
- Resende, E. P. F., Rosen, H. J., Chiang, K., Staffaroni, A. M., Allen, I., Grinberg, L. T., ... Caramelli, P. (2018). Primary school education may be sufficient to moderate a memory-hippocampal relationship. *Frontiers in Aging Neuroscience*, *10*, 381. <https://doi.org/10.3389/fnagi.2018.00381>
- Rosenberg, M. D., Finn, E. E. S., Scheinost, D., Constable, R. R. T., & Chun, M. M. M. (2017). Characterizing attention with predictive network models. *Trends in Cognitive Sciences*, *21*(4), 290–302. <https://doi.org/10.1016/j.tics.2017.01.011>
- Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2015). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*, *19*(1), 165–171. <https://doi.org/10.1038/nn.4179>
- Rosenberg, M. D., Hsu, W.-T., Scheinost, D., Todd Constable, R., & Chun, M. M. (2018). Connectome-based models predict separable components of attention in novel individuals. *Journal of Cognitive Neuroscience*, *30*(2), 160–173. https://doi.org/10.1162/jocn_a_01197
- Sachdev, P. S., Blacker, D., Blazer, D. G., Ganguli, M., Jeste, D. V., Paulsen, J. S., & Petersen, R. C. (2014). Classifying neurocognitive disorders: The DSM-5 approach. *Nature Reviews Neurology*, *10*(11), 634–642. <https://doi.org/10.1038/nrneurol.2014.181>
- Sala-Llonch, R., Bartrés-Faz, D., & Junqué, C. (2015). Reorganization of brain networks in aging: A review of functional connectivity studies. *Frontiers in Psychology*, *6*(May), 663. <https://doi.org/10.3389/fpsyg.2015.00663>
- Salthouse, T. A. (1996a). General and specific speed mediation of adult age differences in memory. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *51B*, P30–P42. <https://doi.org/10.1093/geronb/51B.1.P30>
- Salthouse, T. A. (1996b). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403–428. <https://doi.org/10.1037/0033-295X.103.3.403>
- Salthouse, T. A., Habeck, C., Razlighi, Q., Barulli, D., Gazes, Y., & Stern, Y. (2015). Breadth and age-dependency of relations between cortical thickness and cognition. *Neurobiology of Aging*, *36*, 1–9. <https://doi.org/10.1016/j.neurobiolaging.2015.08.011>
- Schneider-Garces, N. J., Gordon, B., Brumback-Peltz, C. R., Shin, E., Lee, Y., Sutton, B. P., ... Fabiani, M. (2010). Span, CRUNCH, and beyond: Working memory capacity and the aging brain. *Journal of Cognitive Neuroscience*, *22*(4), 655–669. <https://doi.org/10.1162/jocn.2009.21230>
- Schultz, A. P., Chhatwal, J. P., Hedden, T., Mormino, E. C., Hanseeuw, B. J., Sepulcre, J., ... Sperling, R. A. (2017). Phases of Hyperconnectivity and Hypoconnectivity in the default mode and salience networks track with amyloid and tau in clinically Normal individuals. *The Journal of Neuroscience*, *37*(16), 4323–4331. <https://doi.org/10.1523/JNEUROSCI.3263-16.2017>
- Seghier, M. L., & Price, C. J. (2018). Interpreting and Utilising Intersubject variability in brain function. *Trends in Cognitive Sciences*, *22*(6), 517–530. <https://doi.org/10.1016/j.tics.2018.03.003>
- Seo, E. H., Lee, D. Y., Kim, K. W., Lee, J. H., Jhoo, J. H., Youn, J. C., ... Woo, J. I. (2006). A normative study of the trail making test in Korean elders. *International Journal of Geriatric Psychiatry*, *21*(9), 844–852. <https://doi.org/10.1002/gps.1570>
- Shaw, E. E., Schultz, A. P., Sperling, R. A., & Hedden, T. (2015). Functional connectivity in multiple cortical networks is associated with performance across cognitive domains in older adults. *Brain Connectivity*, *5*, 505–516. <https://doi.org/10.1089/brain.2014.0327>
- Sheline, Y. I., & Raichle, M. E. (2013). Resting state functional connectivity in preclinical Alzheimer's Disease. *Biological Psychiatry*, *74*(5), 340–347. <https://doi.org/10.1016/j.biopsych.2012.11.028>
- Shen, X., Tokoglu, F., Papademetris, X., & Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage*, *82*, 403–415. <https://doi.org/10.1016/j.neuroimage.2013.05.081>
- Shim, G., Choi, K.-Y., Kim, D., Suh, S., Lee, S., Jeong, H.-G., & Jeong, B. (2017). Predicting neurocognitive function with hippocampal volumes and DTI metrics in patients with Alzheimer's dementia and mild cognitive impairment. *Brain and Behavior: A Cognitive Neuroscience Perspective*, *7*(9), e00766. <https://doi.org/10.1002/brb3.766>
- Shine, J. M., Bissett, P. G., Bell, P. T., Koyejo, O., Balsters, J. H., Gorgolewski, K. J., ... Poldrack, R. A. (2015). The dynamics of functional brain networks: Integrated network states during cognitive function. *Neuron*, *92*(2), 544–554. <https://doi.org/10.1016/j.neuron.2016.09.018>
- Siedlecki, K. L., Honig, L. S., & Stern, Y. (2008). Exploring the structure of a neuropsychological battery across healthy elders and those with questionable dementia and Alzheimer's Disease. *Neuropsychology*, *22*(3), 400–411. <https://doi.org/10.1037/0894-4105.22.3.400>
- Siedlecki, K. L., Manly, J. J., Brickman, A. M., Schupf, N., Tang, M.-X., & Stern, Y. (2010). Do neuropsychological tests have the same meaning in Spanish speakers as they do in English speakers? *Neuropsychology*, *24*(3), 402–411. <https://doi.org/10.1037/a0017515>
- Siegel, J. S., Ramsey, L. E., Snyder, A. Z., Metcalf, N. V., Chacko, R. V., Weinberger, K., ... Corbetta, M. (2016). Disruptions of network connectivity predict impairment in multiple behavioral domains after stroke. *Proceedings of the National Academy of Sciences*, *113*(30), E4367–E4376. <https://doi.org/10.1073/pnas.1521083113>
- Smith, S. M., Nichols, T. E., Vidaurre, D., Winkler, A. M., Behrens, T. E., Glasser, M. F., ... Miller, K. L. (2015). A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Neuroscience*, *18*(September), 1–7. <https://doi.org/10.1038/nn.4125>

- Stam, C. J. (2014). Modern network science of neurological disorders. *Nature Reviews Neuroscience*, 15(10), 683–695. <https://doi.org/10.1038/nrn3801>
- Steffener, J., Barulli, D., Habeck, C., O'Shea, D., Razlighi, Q., & Stern, Y. (2014). The role of education and verbal abilities in altering the effect of age-related gray matter differences on cognition. *PLoS One*, 9(3), e91196. <https://doi.org/10.1371/journal.pone.0091196>
- Sui, J., Qi, S., van Erp, T. G. M., Bustillo, J., Jiang, R., Lin, D., ... Calhoun, V. D. (2018). Multimodal neuromarkers in schizophrenia via cognition-guided MRI fusion. *Nature Communications*, 9(1), 3028. <https://doi.org/10.1038/s41467-018-05432-w>
- Sutterer, M. J., & Tranel, D. (2017). Neuropsychology and cognitive neuroscience in the fMRI era: A recapitulation of localizationist and connectionist views. *Neuropsychology*, 31(8), 972–980. <https://doi.org/10.1037/neu0000408>
- Tavor, I., Jones, O. P., Mars, R. B., Smith, S. M., Behrens, T. E., & Jbabdi, S. (2016). Task-free MRI predicts individual differences in brain activity during task performance. *Science*, 352(6282), 216–220. <https://doi.org/10.1126/science.aad8127>
- Vaidya, A. R., Pujara, M. S., Petrides, M., Murray, E. A., & Fellows, L. K. (2019). Lesion studies in contemporary neuroscience. *Trends in Cognitive Sciences*, 23(8), 653–671. <https://doi.org/10.1016/j.tics.2019.05.009>
- Van Petten, C. (2004). Relationship between hippocampal volume and memory ability in healthy individuals across the lifespan: Review and meta-analysis. *Neuropsychologia*, 42(10), 1394–1413. <https://doi.org/10.1016/j.neuropsychologia.2004.04.006>
- Varoquaux, G. (2018). Cross-validation failure: Small sample sizes lead to large error bars. *NeuroImage*, 180, 68–77. <https://doi.org/10.1016/j.neuroimage.2017.06.061>
- Vemuri, P., Weigand, S. D., Przybelski, S. A., Knopman, D. S., Smith, G. E., Trojanowski, J. Q., ... Jack, C. R. (2011). Cognitive reserve and Alzheimer's disease biomarkers are independent determinants of cognition. *Brain*, 134(5), 1479–1492. <https://doi.org/10.1093/brain/awr049>
- Wechsler, D. (1987a). *Manual for Wechsler adult intelligence scale [revised]*. New York, NY: The Psychological Corporation.
- Wechsler, D. (1987b). *Manual for Wechsler memory scale [revised]*. New York, NY: The Psychological Corporation.
- Weintraub, S., Salmon, D., Mercaldo, N., Ferris, S., Graff-Radford, N. R., Chui, H., ... Morris, J. C. (2009). The Alzheimer's Disease Centers' uniform data set (UDS). *Alzheimer Disease & Associated Disorders*, 23(2), 91–101. <https://doi.org/10.1097/wad.0b013e318191c7dd>
- Whelan, R., Watts, R., Orr, C. a., Althoff, R. R., Artiges, E., Banaschewski, T., ... Garavan, H. (2014). Neuropsychosocial profiles of current and future adolescent alcohol misusers. *Nature*, 512(7513), 185–189. <https://doi.org/10.1038/nature13402>
- Woo, C.-W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better biomarkers: Brain models in translational neuroimaging. *Nature Neuroscience Reviews*, 20(3), 365–377. <https://doi.org/10.1038/nn.4478>
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12(6), 1100–1122. <https://doi.org/10.1177/1745691617693393>
- Yoo, K., Rosenberg, M. D., Hsu, W. T., Zhang, S., Li, C. S. R., Scheinost, D., ... Chun, M. M. (2018). Connectome-based predictive modeling of attention: Comparing different functional connectivity features and prediction methods across datasets. *NeuroImage*, 167(November 2017), 11–22. <https://doi.org/10.1016/j.neuroimage.2017.11.010>
- Youm, Y., Laumann, E. O., Ferraro, K. F., Waite, L. J., Kim, H. C., Park, Y. R., ... Lee, J. A. (2014). Social network properties and self-rated health in later life: Comparisons from the Korean social life, health, and aging project and the national social life, health and aging project. *BMC Geriatrics*, 14, 102. <https://doi.org/10.1186/1471-2318-14-102>
- Younes, L., Albert, M., Moghekar, A., Soldan, A., Pettigrew, C., & Miller, M. I. (2019). Identifying changepoints in biomarkers during the preclinical phase of Alzheimer's disease. *Frontiers in Aging Neuroscience*, 11, 74. <https://doi.org/10.3389/fnagi.2019.00074>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Kwak S, Kim H, Kim H, Youm Y, Chey J. Distributed functional connectivity predicts neuropsychological test performance among older adults. *Hum Brain Mapp.* 2021;42:3305–3325. <https://doi.org/10.1002/hbm.25436>