Data Article

Dataset of relationship between longitudinal change in cognitive performance and functional connectivity in cognitively normal older individuals

Kumiko Oishi[a], Anja Soldan[b], Corinne Pettigrew[b], Johnny Hsu[c], Susumu Mori[c], Marilyn Albert[b], Kenichi Oishi[c,*], The BIOCARD Research Team

[a] Center for Imaging Science, Whiting School of Engineering, The Johns Hopkins University, Baltimore, MD, USA
[b] Department of Neurology, The Johns Hopkins University School of Medicine, Baltimore, MD, USA
[c] The Russell H. Morgan Department of Radiology and Radiological Science, The Johns Hopkins University School of Medicine, 208 Traylor Building, 720 Rutland Avenue, Baltimore, MD 21205, USA

ARTICLE INFO

ABSTRACT

The data show an association between measured and predicted changes in cognitive performance in older adults who are cognitively normal. Changes in cognitive performance over two years were assessed using the Cognitive Composite Score. The prediction of change in cognitive function was based on changes in pairwise functional connectivity between 80 gray matter regions examined by resting-state functional magnetic resonance imaging. A feature extraction process based on the Variable Importance Testing Approach (VITA) identified changes in 11 pairs of functional connections associated with the default mode network as features related to changes in cognitive performance. Linear and elastic net regression models were applied to these 11 features to predict changes in cognitive performance over two years. A relationship between the 11 features and the geriatric depression score was also shown. The dataset supplements the research findings in the "Changes in pairwise functional con-

https://doi.org/10.1016/j.dib.2022.108302

nectivity associated with changes in cognitive performance in cognitively normal older individuals: a two-year observational study" published in Oishi et al. (2022). The raw rs-fMRI correlation matrix and associated clinical data can be accessed upon request from the BIOCARD website (www.biocard-se.org) and can be reused for predictive model building.

Specifications Table

Subject	Neuroscience
Specific subject area	Functional neuroimaging, resting-state functional connectivity, aging, cognitive change
Type of data	Table
	Figure
How the data were acquired	The Imaging Core staff of the Johns Hopkins BIOCARD study team acquired high-resolution three-dimensional T1-weighted images and resting-state functional magnetic resonance imaging (rs-fMRI) using a 3T MR system (Philips Healthcare, Best, The Netherlands).
	Clinical assessments, cognitive evaluations, and APOE genotyping were performed by the Clinical Core staff of the Johns Hopkins BIOCARD study team.
	Variable importance measures were performed using VITA in R (v.4.0.2).
	Predictive modeling was performed using linear regression and an elastic net regression in the classification and regression training (caret) package in R (v.4.0.2).
	Evaluation of model performance was assessed with four-fold cross-validation, in which the data splitting was performed using the vtreat package in R (v.4.0.2).
	Depressive symptoms were rated using the Geriatric Depression Scale (GDS) [2].
	Correlation analysis was carried out with Pearson's product-moment correlation and Kendall's ranking correlation in R (v.4.0.2).
Data format	Analyzed
Description of data collection	Clinical assessments and cognitive evaluations for participants were performed during at least six visits, approximately one year apart. The MRI scans were obtained at the fourth visit (between 2015 and 2017, Time 1) and at the sixth visit (between 2017 and 2019, Time 2). Clinical evaluations were performed using methods consistent with the National Institute on Aging Alzheimer's Disease (AD) Research Centers program. The diagnostic procedures were in accordance with the recommendations for diagnosing mild cognitive impairment (MCI) and dementia due to AD contained in the report of the NIA/AA working group.
	The EPI sequence was used to acquire rs-fMRI data consisting of 140 functional volumes. The total scan duration for each session was 420 s. Forty-Eight axial slices were obtained to cover the whole brain. The field of view (FOV) was 2.12×2.12 cm^2; the voxel size was $3.3 \times 3.3 \times 3.3$ mm^3; the repetition time (TR) was 3000 ms; the echo time (TE) was 30 ms; the flip angle was 75°.
	For the anatomical reference, magnetization-prepared rapid gradient echo (MPRAGE) scans were acquired. There were 170 sagittal slices that covered the whole brain. The FOV was 240×256 mm^2; the TR was 6.8 ms; the TE was 3.1 ms; the shot interval was 3000 ms; the flip angle was 8°; the voxel size was $1 \times 1 \times 1.2$ mm^3; the scan duration was 5 min 59 s.
	The image processing, including postprocessing, segmentation, and quantification, was performed in MRICloud (www.MRICloud.org).

	The identification of variables associated with changes in cognitive performance was based on 15 cross-validations and 10 permutations. Regression coefficients of the prediction model were obtained by fitting the model with 10-fold cross-validation [3]. Predictive model performance was evaluated by correlation analysis between predicted response and measured response using Pearson's product-moment correlation. Relationships between predicted response and measured response were visualized in scatterplots.
Data source location	School of Medicine, Johns Hopkins University Baltimore, Maryland United States
Data accessibility	Biomarkers for Older Controls at Risk for Dementia (BIOCARD) https://www.biocard-se.org/public/Data%20Access%20Procedures.html The raw datasheet containing the rs-fMRI correlation matrix and associated clinical data are available upon request. Send the required documents (BIOCARD Data Use Agreement_FILLABLE.pdf) via e-mail to BIOCARD@jhu.edu. Please clearly state that you are requesting the rs-fMRI datasheet in Oishi, K. et al., Data in Brief. Send any questions about the requests to this e-mail address as well. The rs-fMRI correlation matrix among structures was obtained from the nuisance-corrected BOLD time courses, followed by Fisher's z-transformation, which resulted in the z-correlations (z-cor) at each time point (Time1 and Time2). The other columns include age, sex, years of education, vascular health risk composite score, APOE gene status, and cognitive composite score when the MRI scan was assessed. If you have specific questions about the BIOCARD study, you may contact the Principal Investigator, Dr. Marilyn Albert, at malbert9@jhmi.edu, or (410) 614-3040.
Related research article	K. Oishi, A. Soldan, C. Pettigrew, J. Hsu, S. Mori, M. Albert, K. Oishi, BIOCARD Research Team, Changes in pairwise functional connectivity associated with changes in cognitive performance in cognitively normal older individuals: a two-year observational study, Neurosci Lett 10.1016/j.neulet.2022.136618 (2022) 136,618.https://www.ncbi.nlm.nih.gov/pubmed/35398188

Value of the Data

- This dataset is beneficial to researchers in the field of cognitive aging by providing a natural course of cognitive performance over five years of observation in cognitively normal older adults.
- This dataset is beneficial to functional connectivity researchers by providing an elastic net regression model that predicts changes in cognitive function from pairs of resting-state functional connectivity.
- This dataset can provide research guidance to data scientists who aim to extract clinically relevant information from signals originating from the brain.
- This dataset can be used as a reference with which to assess changes in cognitive performance seen in older adults who are cognitively normal.
- This data set can be used as a benchmark to evaluate advanced mathematical models aimed at predicting changes in cognitive performance.

1. Data Description

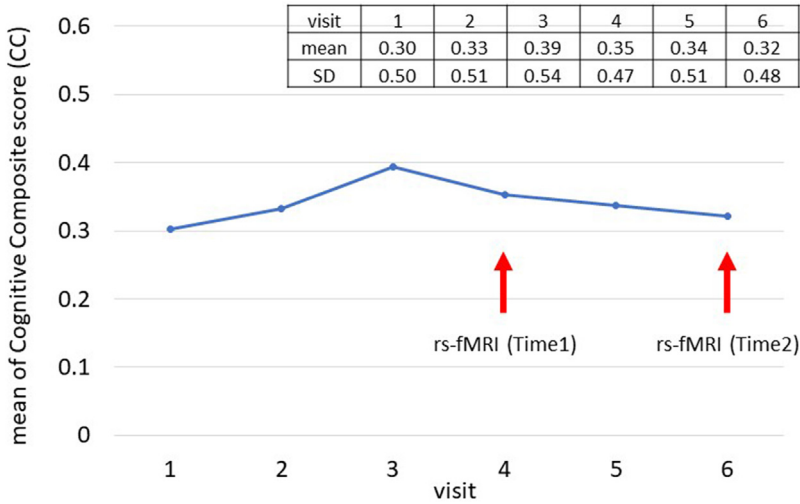| visit	1	2	3	4	5	6
mean	0.30	0.33	0.39	0.35	0.34	0.32
SD	0.50	0.51	0.54	0.47	0.51	0.48

Fig. 1. Trajectory of cognitive composite score in older adults who are cognitively normal. Cognitive performance was evaluated annually using the Cognitive Composite Score (CC). Each participant was assessed a total of at least six times. The mean CC of 92 cognitively normal participants is plotted against the number of visits. The rs-fMRI was scanned at visit 4 (= Time 1) and visit 6 (= Time 2). Mean CC and standard deviation (SD) are shown above the graph.
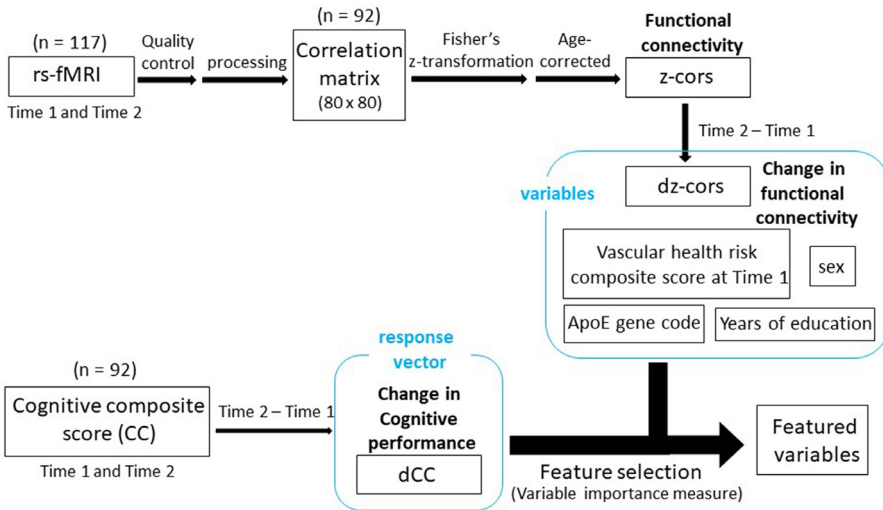

Fig. 2. Workflow for extracting features related to changes in cognitive performance from rs-fMRI signals and risk factors for cognitive decline.

Upper row: The process of calculating the change in pairwise functional connectivity between Time 1 and Time 2. For the evaluation of functional connectivity, the correlation matrix of the rs-fMRI data was Fisher's z-transformed and age-corrected (z-cors). The change in functional connectivity (dz-cors) was calculated from the difference in z-cors between Time 1 and Time 2.

Middle row: A set of variables that may be associated with changes in cognitive performance.

Lower row: Evaluation of change in cognitive performance between Time 1 and Time 2 (dCC) and selection of important variables related to dCC. The change in cognitive performance (dCC) was calculated from the difference in CC between Time1 and Time 2. Among 3160 dz-cors, vascular health risk, years of education, sex, and APOE code, important features associated with dCC were selected using VITA.
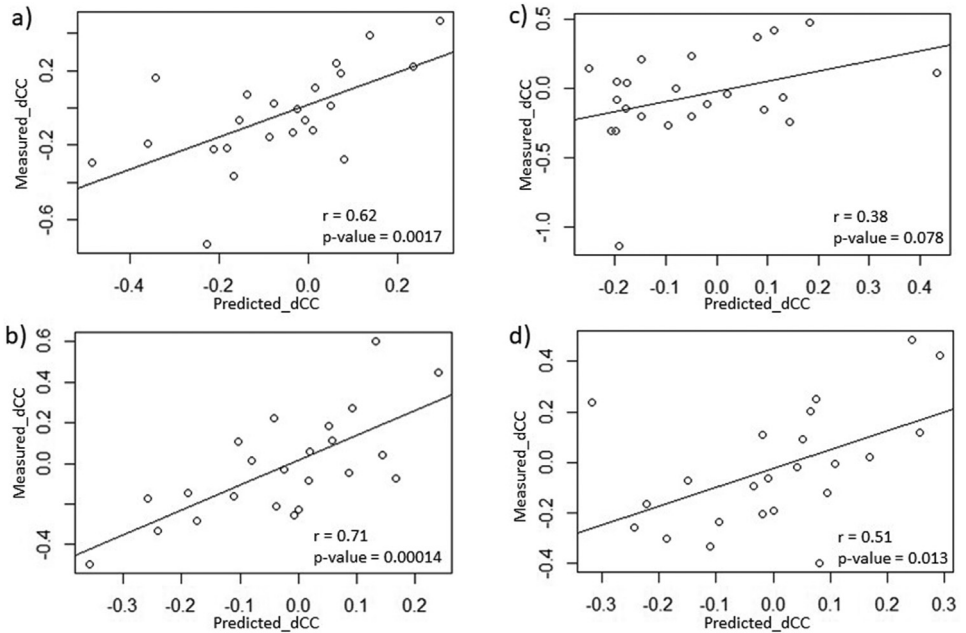
Fig. 3. Scatterplots showing the relationship between measured dCC (y-axis) and dCC predicted by the linear regression model (x-axis). With the variables selected by the procedure detailed in Fig. 2 and sex as inputs, a linear regression model was applied to create a predictive model for dCC. Four-fold cross-validation was used for the validation. That is, the data were randomly divided into four equal-sized subsets (subsets a–d); all samples not included in the selected subset were used as training data, and samples included in the selected subset were used as test data. Predictive model performance was evaluated by coefficient of determination between measured dCC and predicted dCC and the p-values, which are embedded in each graph. This data supplements Table 3 in the original publication [1] to assess the predictive model's performance.

2. Experimental Design, Materials and Methods

2.1. Participants

Data of 92 cognitively normal older individuals (52–82 years of age (SD, 6.22 years; male/female = 31/61) were extracted from the Biomarker for Older Controls at Risk for Dementia (BIOCARD) cohort. All participants signed informed consent to complete examinations and evaluate previously obtained blood and cerebrospinal fluid (CSF) samples and brain MRIs.

2.2. Clinical Assessments

The Clinical Core staff of the Johns Hopkins BIOCARD study team made a consensus diagnoses. First, a syndromic diagnosis of normal, mild cognitive impairment (MCI), or dementia was made. This diagnosis was based on clinical data about medical, neurological, and psychiatric status; reports of cognitive changes by the individual and his or her companion; and cognitive decline based on a review of longitudinal tests in multiple cognitive domains. If cognitive decline was determined to be present, the probable etiology of the syndrome was then determined based on medical, neurological, and psychiatric information collected at each visit and from medical records obtained from the individual. Each individual might have multiple etiologies, such as AD and vascular disease. Vascular health risk factors and APOE gene status were assessed as in the original publication [1].

2.3. Cognitive Composite Score (CC) and Geriatric Depression Scale (GDS) Score

The CC is the mean of the following four cognitive test scores converted to z-scores [4]. The cognitive tests used were Paired Associates Immediate Recall and Logical Memory Delayed Recall (Story A) from the Wechsler Memory Scale-Revised [5], Boston Naming [6], and Digit-Symbol Substitution of the Wechsler Adult Intelligence Scale-Revised [7]. The requirement for measuring cognitive composite scores was the presence of at least two of the four scores at each visit [8], and if two or more scores were missing, the CC for that visit was considered missing. The two-year change in cognitive performance (dCC) was obtained from the difference in CC between Time 1 and Time 2 (Fig. 1). Depressive symptoms were assessed using GDS [2] (Min = 0, Max = 15) (results are described in the original publication in [1]) (Table 1).

2.4. MRI Scans

An echo-planar imaging sequence was used to acquire rs-fMRI data consisting of 140 functional volumes. The total scan duration for each session was 420 s. Forty-Eight axial slices were obtained to cover the whole brain. The field of view (FOV) was 2.12×2.12 cm^2; the voxel size was $3.3 \times 3.3 \times 3.3$ mm^3; the repetition time (TR) was 3000 ms; the echo time (TE) was 30 ms; the flip angle was 75°. For the anatomical reference, magnetization-prepared rapid gradient echo (MPRAGE) scans were acquired. There were 170 sagittal slices that covered the whole brain. The FOV was 240×256 mm^2; the TR was 6.8 ms; the TE was 3.1 ms; the shot interval was 3000 ms; the flip angle was 8°; the voxel size was $1 \times 1 \times 1.2$ mm^3; and the scan duration was 5 min 59 s.

2.5. Image Processing

rs-fMRI images were automatically post-processed, segmented, and qualified in MRICloud (www.MRICloud.org). The multi-atlas label fusion approach was applied to parcellate the MPRAGE image into 283 anatomical structures that included 80 gray matter structures used in this study [9]. The JHU multi-atlas library [10] version 9, containing 30 atlases covering an age-range of 50–90 years, was adopted as the reference library. The label fusion step was performed based on [11], adjusted by the PICSL algorithm [12]. The MPRAGE and corresponding parcellation map were co-registered to the motion and slice timing-corrected EPI [13], and the time course of the BOLD signal was extracted from the 72 cortical (including the hippocampus) and eight sub-cortical gray matter regions [14].

2.6. rs-fMRI Analysis and Feature Selection

The correlation matrix among structures was obtained from processed rs-fMRI images. Fisher's z-transformed correlation was used as functional connectivity (z-cor). z-cors were age-corrected [15]. The two-year changes in functional connectivity (dz-cors) were obtained from the differences of z-cors between Time 1 and Time 2. Featured variables related to dCC were assessed using VITA [16] with a cross-validated permutation variable importance measure (CVPVI) from 3164 variables, including 3160 dz-cors and risk factors (sex, vascular health risk, years of education, and APOE gene status) at Time 1 (detailed in the original publication in [1]) (Fig. 2).

2.7. Prediction Modeling

After feature selection using VITA, we applied the generalized model to build the prediction model with selected functional connections as variables for a change in cognition. We used four-

Table 1

Regression coefficients and the intercept obtained from an elastic net regression model. With the variables selected by the procedure detailed in Fig. 2 and sex as inputs, elastic net regression was applied to generate a predictive model for dCC. Four-fold cross-validation was used for the validation. That is, the data were randomly divided into four equal-sized subsets (subsets 1–4 in Table 1); all samples not included in the selected subset were used as training data, and samples included in the selected subset were used as test data. Predictive model performance was evaluated by coefficient of determination between measured dCC and predicted dCC and the p-value. AG_L, the left angular gyrus; AG_R, the right angular gyrus; subgenual_ACC_R, the left subgenual anterior cingulate cortex; MTG_L_pole, the left middle temporal gyrus pole; MTG_R, the right middle temporal gyrus; MTG_L, the left middle temporal gyrus; MTG_L_pole, pole of the left middle temporal gyrus; PrCu_R, the right precuneus; STG_R_pole, pole of the right superior temporal gyrus; Insula_L, the left insula; RG_R, the right rectal gyrus; GP_L, the left globus pallidus; Caud_L, the left caudate; LFOG_R, the right lateral fronto-orbital gyrus; MFG_L, the left middle frontal gyrus; MFG_DPFC_R, the right dorsolateral prefrontal aspect of the middle frontal gyrus; SFG_L, the left superior frontal gyrus; SFG_PFC_R, the right prefrontal aspect of the superior frontal gyrus; SPG_L, the left superior parietal gyrus. The areas involved in the default mode network are boxed; the salient network is underlined; and the lateral prefrontal areas are double underlined. This data supplements Table 3 of the original publication in [1], in which the results of a simple linear model are presented instead of the elastic net penalized regression model.

Variables	Regression coefficients				
	subset1	subset2	subset3	subset4	mean
Sex (female = 1, male = 0)	-0.075	-0.029	-0.034	-0.021	-0.04
AG_L - RG_R	0.13	0.018	0.18	0.053	0.097
AG_L - STG_R_pole	0.067	0.11	0.087	0.097	0.091
subgenual_ACC_R - Insula_L	0.14	0.05	0.15	0.13	0.12
MTG_R - Caud_L	0.041	0.12	-0.05	0.067	0.045
AG_R - MTG_L_pole	0.15	0.12	0.068	0.11	0.11
MTG_L_pole - GP_L	0.11	0.079	0.15	0.13	0.12
MTG_L - LFOG_R	0.16	0.0083	0.15	0.17	0.12
SPG_L - MFG_DPFC_R	0.071	0.062	0.13	0.084	0.088
AG_L - SFG_PFC_R	0.023	0.022	0.054	0.16	0.065
PrCu_R - SFG_L	-0.35	-0.25	-0.32	-0.27	-0.3
MTG_R - MFG_L	0.085	0.049	0.099	0.055	0.072
intercept	0.14	0.039	0.060	0.047	0.071
r^2 (predicted dCC vs dCC)	0.39	0.44	0.15	0.26	0.31
p-value	0.0015	0.00051	0.065	0.016	0.021

fold cross-validation for the evaluation of the model, in which the data set was split into training data ($n = 69$) and test data ($n = 23$). The kWayStratifiedY function in the vtreat package in R (v.4.0.2) was used to introduce a stratified sampling to address the internal structure imbalance problem of the data splitting. Namely, it allows the splitting of data to maintain the internal distribution of dCC in the original data. The glmnet for elastic net regression [3] in the classification and regression training (caret) package in R (v.4.0.2) was used. Model fitting for regression co-

Table 2

Correlation between change in functional connectivity (dz-cors, see Fig. 2) and the Geriatric Depression Scale (GDS). The dz-cors were selected by the procedures described in Fig. 2. AG_L, the left angular gyrus; AG_R, the right angular gyrus; subgenual_ACC_R, the left subgenual anterior cingulate cortex; MTG_L_pole, the left middle temporal gyrus pole; MTG_R, the right middle temporal gyrus; MTG_L, the left middle temporal gyrus; MTG_L_pole, pole of the left middle temporal gyrus; PrCu_R, the right precuneus; STG_R_pole, pole of the right superior temporal gyrus; Insula_L, the left insula; RG_R, the right rectal gyrus; GP_L, the left globus pallidus; Caud_L, the left caudate; LFOG_R, the right lateral fronto-orbital gyrus; MFG_L, the left middle frontal gyrus; MFG_DPFC_R, the right dorsolateral prefrontal aspect of the middle frontal gyrus; SFG_L, the left superior frontal gyrus; SFG_PFC_R, the right prefrontal aspect of the superior frontal gyrus; SPG_L, the left superior parietal gyrus. The areas involved in the default mode network are boxed; the salient network is underlined; and the lateral prefrontal areas are double underlined. This dataset was supplementary to the original publication in [1].

	r (vs GDS)	p-value
AG_L - RG_R	0.007	0.93
AG_L - STG_R_pole	-0.0026	0.98
subgenual_ACC_R - Insula_L	0.079	0.33
MTG_R - Caud_L	0.10	0.21
AG_R - MTG_L_pole	0.049	0.55
MTG_L_pole - GP_L	0.0089	0.91
MTG_L - LFOG_R	0.13	0.099
SPG_L - MFG_DPFC_R	-0.024	0.77
AG_L - SFG_PFC_R	0.098	0.23
PrCu_R - SFG_L	-0.011	0.89
MTG_R - MFG_L	0.073	0.37

efficients was performed with ten-fold cross-validation in all four subsets of training data. Each model was applied to each subset of the test dataset to obtain the predicted dCC. The model performance for the prediction of the dCC was evaluated by the correlation analysis between the measured dCC and the output of the model (predicted dCC) obtained from four subsets of the test data. The correlation between the measured and predicted dCC was calculated using Pearson's product-moment correlation in R (v.4.0.2). The regression coefficients obtained from each of the four models and the *p*-values were demonstrated (Figs. 3 and 4).

2.8. Correlation Between Functional Connectivity and Geriatric Depression Scale (GDS) Score

GDS score at baseline (Time 1) was assessed in the correlation analysis with the dz-cor of featured functional connections (Table 2). Correlation analysis was performed using Kendall's rank correlation in R (v.4.0.2).

Ethics Statements

The BIOCARD study was approved by the JHU Institutional Review Board,
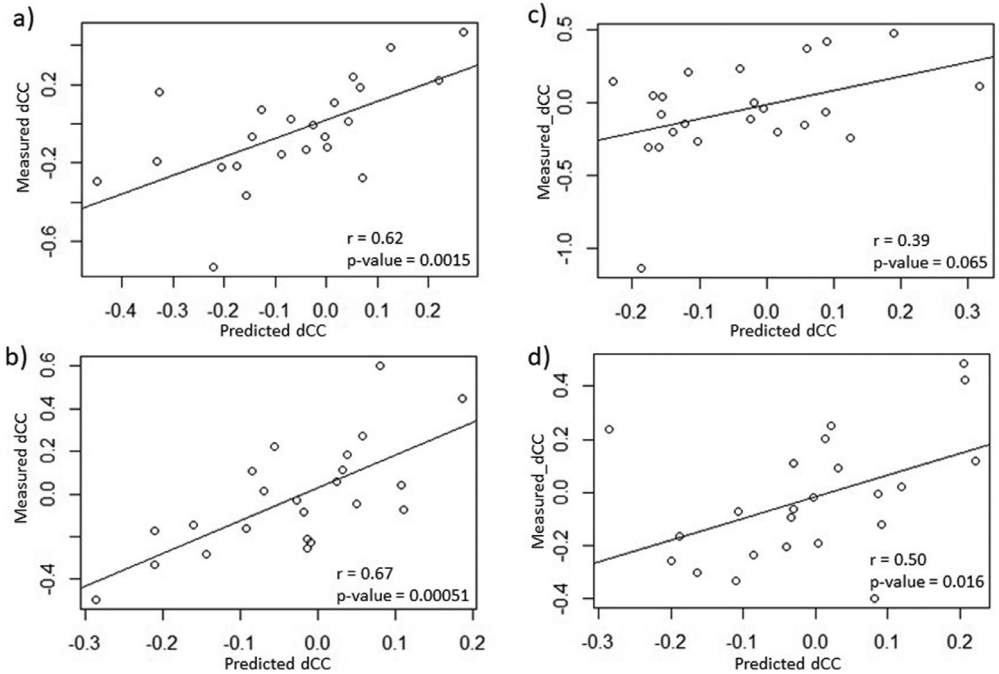
Fig. 4. Scatterplots showing the relationship between measured dCC (y-axis) and dCC predicted by the elastic net penalized model (x-axis). The (a–d) represent the results of subset 1, subset 2, subset 3, and subset 4 described in Table 1. Correlation and *p*-values are embedded in each graph. This data supplements Table 3 in the original publication [1] to assess the predictive model's performance.

The Johns Hopkins School of Medicine IRB, No. NA_00027232. All participants provided written, informed consent in accordance with the Declaration of Helsinki.

Declaration of Competing Interest

SM is co-founder and CEO, and KO is a consultant for "AnatomyWorks" and "Corporate-M." This arrangement is being managed by the Johns Hopkins University in accordance with its conflict-of-interest policies.

Data Availability

BIOCARD MRI (Original data) (BIOCARD).

CRediT Author Statement

Kumiko Oishi: Software, Visualization, Investigation, Data curation, Writing – original draft; **Anja Soldan:** Investigation, Data curation; **Corinne Pettigrew:** Investigation, Data curation; **Johnny Hsu:** Data curation; **Susumu Mori:** Supervision; **Marilyn Albert:** Supervision; **Kenichi Oishi:** Writing – review & editing, Supervision.

Acknowledgments

References

[1] K. Oishi, A. Soldan, C. Pettigrew, J. Hsu, S. Mori, M. Albert, K. Oishi, B.R. Team, Changes in pairwise functional connectivity associated with changes in cognitive performance in cognitively normal older individuals: a two-year observational study, Neurosci. Lett. (2022) 136618, doi:10.1016/j.neulet.2022.136618.

[2] J.I. Sheikh, J.A. Yesavage, Geriatric Depression Scale (GDS): recent evidence and development of a shorter version, Clin. Gerontol. 5 (1986) 9, doi:10.1300/J018v05n01_09.

[3] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, J. Stat. Softw. 33 (1) (2010) 1–22.

[4] M. Albert, A. Soldan, R. Gottesman, G. McKhann, N. Sacktor, L. Farrington, M. Grega, R. Turner, Y. Lu, S. Li, M.C. Wang, O. Selnes, Cognitive changes preceding clinical symptom onset of mild cognitive impairment and relationship to APOE genotype, Curr. Alzheimer Res. 11 (8) (2014) 773–784, doi:10.2174/156720501108140910121920.

[5] D. Wechsler, Manual for the Wechsler Memory Scale-Revised A., Psychological Corporation, 1987.

[6] E. Kaplan, H. Goodglass, S. Weintraub, Boston Naming Test, Lea & Febiger, Philadelphia, 1983.

[7] D. Wechsler, Adult Intelligence Scale - Revised Manual, Psychological Corporation, 1981.

[8] A. Soldan, C. Pettigrew, Q. Cai, M.C. Wang, A.R. Moghekar, R.J. O'Brien, O.A. Selnes, M.S. Albert, B.R. Team, Hypothetical preclinical alzheimer disease groups and longitudinal cognitive change, JAMA Neurol. 73 (6) (2016) 698–705, doi:10.1001/jamaneurol.2016.0194.

[9] P. Liang, L. Shi, N. Chen, Y. Luo, X. Wang, K. Liu, V.C. Mok, W.C. Chu, D. Wang, K. Li, Construction of brain atlases based on a multi-center MRI dataset of 2020 Chinese adults, Sci. Rep. 5 (2015) 18216, doi:10.1038/srep18216.

[10] D. Wu, T. Ma, C. Ceritoglu, Y. Li, J. Chotiyanonta, Z. Hou, J. Hsu, X. Xu, T. Brown, M.I. Miller, S. Mori, Resource atlases for multi-atlas brain segmentations with multiple ontology levels based on T1-weighted MRI, Neuroimage 125 (2016) 120–130, doi:10.1016/j.neuroimage.2015.10.042.

[11] X. Tang, K. Oishi, A.V. Faria, A.E. Hillis, M.S. Albert, S. Mori, M.I. Miller, Bayesian parameter estimation and segmentation in the multi-atlas random orbit model, PLoS ONE 8 (6) (2013) e65591, doi:10.1371/journal.pone.0065591.

[12] H. Wang, P.A. Yushkevich, Multi-atlas segmentation with joint label fusion and corrective learning-an open source implementation, Front. Neuroinform. 7 (2013) 27, doi:10.3389/fninf.2013.00027.

[13] T.J.R. Rezende, B.M. Campos, J. Hsu, Y. Li, C. Ceritoglu, K. Kutten, M.C. Franca Junior, S. Mori, M.I. Miller, A.V. Faria, Test-retest reproducibility of a multi-atlas automated segmentation tool on multimodality brain MRI, Brain Behav. 9 (10) (2019) e01363, doi:10.1002/brb3.1363.

[14] A.V. Faria, S.E. Joel, Y. Zhang, K. Oishi, P.C. van Zijl, M.I. Miller, J.J. Pekar, S. Mori, Atlas-based analysis of resting-state functional connectivity: evaluation for reproducibility and multi-modal anatomy-function correlation studies, Neuroimage 61 (3) (2012) 613–621, doi:10.1016/j.neuroimage.2012.03.078.

[15] J. Dukart, M.L. Schroeter, K. Mueller, I. Alzheimer's Disease Neuroimaging, Age correction in dementia–matching to a healthy brain, PLoS ONE 6 (7) (2011) e22193, doi:10.1371/journal.pone.0022193.

[16] S. Janitza, E. Celik, A.L. Boulesteix, A computationally fast variable importance test for random forests for high-dimensional data, Adv. Data Anal. Classif. 12 (2018) 885–915, doi:10.1007/s11634-016-0276-4.