Residual Partial Least Squares Learning:

Brain Cortical Thickness Simultaneously Predicts Eight Non-pairwise-correlated Behavioural and Disease Outcomes in Alzheimer's Disease

Oliver Y. Chén[1,2*], Duy Thanh Vũ[1,6*], Christelle Schneuwly Diaz[1,2], Julien S. Bodelet[1], Huy Phan[3], Gilles Allali[4], Viet-Dung Nguyen[5,6], Hengyi Cao[7,8], Xingru He[9], Yannick Müller[1], Bangdong Zhi[10], Haochang Shou[11], Haoyu Zhang[12], Wei He[9], Xiaojun Wang[10], Marcus Munafò[13], Nguyen Linh Trung[6], Guy Nagels[14,15], Philippe Ryvlin[16], Giuseppe Pantaleo[1], for the Alzheimer's Disease Neuroimaging Initiative.

1. Département Médecine de Laboratoire et Pathologie, Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland.
2. Faculté de Biologie et de Médecine, Université de Lausanne (UNIL), Lausanne, Switzerland.
3. Department of Electrical Engineering, KU Leuven, Leuven, Belgium.
4. Centre Leenaards de la Mémoire, CHUV, Lausanne, Switzerland.
5. Lab-STICC, École Nationale Supérieure de Techniques Avancées de Bretagne, Bretagne, France.
6. The Advanced Institute of Engineering and Technology, Vietnam National University, Hanoi, Vietnam.
7. Center for Psychiatric Neuroscience, Feinstein Institutes for Medical Research, Manhasset, NY, USA.
8. Division of Psychiatry Research, Zucker Hillside Hospital, Glen Oaks, NY, USA.
9. School of Public Health, He University, Shengyang, China.
10. Innovation and Healthcare Group, University of Bristol, Bristol, UK.
11. Department of Biostatistics, University of Pennsylvania, Philadelphia, PA, USA.
12. Division of Cancer Epidemiology and Genetics, National Institutes of Health, Bethesda, MD, USA.
13. School of Psychological Science, University of Bristol, Bristol, UK.
14. Department of Neurology, Universitair Ziekenhuis Brussel, Jette, Belgium.
15. Institute of Biomedical Engineering, University of Oxford, Oxford, UK.
16. Département des Neurosciences Cliniques, CHUV, Lausanne, Switzerland.

*Correspondence should be addressed to:

Duy Thanh Vũ
Département Médecine de Laboratoire et Pathologie
CHUV
1011 Lausanne, Switzerland
duy.vu@chuv.ch

Oliver Y. Chén
Département Médecine de Laboratoire et Pathologie, CHUV
Faculté de Biologie et de Médecine, UNIL
1011 Lausanne, Switzerland
olivery.chen@chuv.ch

Abstract

Alzheimer's Disease (AD) is the leading cause of dementia. It results in cortical thickness changes and is associated with a decline in cognition and behaviour. Such decline affects multiple important day-to-day functions, including memory, language, orientation, judgment and problem-solving. Recent research has made important progress in identifying brain regions associated with single outcomes, such as individual AD status and general cognitive decline. The complex projection from multiple brain areas to multiple AD outcomes, however, remains poorly understood. This makes the assessment and especially the prediction of multiple AD outcomes - each of which may unveil an integral yet different aspect of the disease - challenging, particularly when some are not strongly correlated. Here, uniting residual learning, partial least squares (PLS), and predictive modelling, we develop an explainable, generalisable, and reproducible method called the *Residual Partial Least Squares Learning* (the re-PLS Learning) to (1) chart the pathways between large-scale multivariate brain cortical thickness data (inputs) and multivariate disease and behaviour data (outcomes); (2) simultaneously predict multiple, non-pairwise-correlated outcomes; (3) control for confounding variables (*e.g.*, age and gender) affecting both inputs and outcomes and the pathways in-between; (4) perform longitudinal AD disease status classification and disease severity prediction. We evaluate the performance of the proposed method against a variety of alternatives on data from AD patients, subjects with mild cognitive impairment (MCI), and cognitively normal individuals ($n = 1{,}196$) from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Our results unveil pockets of brain areas in the temporal, frontal, sensorimotor, and cingulate areas whose cortical thickness may be *respectively* associated with declines in different cognitive and behavioural subdomains in AD. Finally, we characterise re-PLS' geometric interpretation and mathematical support for delivering meaningful neurobiological insights and provide an open software package (*re-PLS*) available at https://github.com/thanhvd18/rePLS.

Introduction

Alzheimer's Disease (AD) is a neurodegenerative disorder affecting 50 million people worldwide and is projected to affect as many as 152 million by 2050[1]. It is the cause of 60–70% of dementia cases[2]. An early symptom of classical AD is the difficulty in remembering recent events. Gradually, a patient may exhibit difficulties in language, orientation, mood swings, loss of motivation, self-neglect, and behavioural changes. Progressive cognitive decline is accompanied by gradual loss of bodily functions eventually leading to death[3]. An AD patient's typical life expectancy following diagnosis ranges from three to nine years[4].

As AD patients lose important cognitive, behavioural, and social abilities to perform basic daily activities, their quality of life (QoL) decreases[5,6]. The decline in QoL, unfortunately, is multifaceted[7], from losing a sense of purpose or pleasure[8,9], social isolation[10], to developing depression[11,12]. Because AD and related dementia are directly associated with disability and, eventually, death, the economic burden of the disease globally was staggering: from $2.8 trillion in 2019 to an expected $16.9 trillion in 2050[13].

Discovering markers associated with, and in particular predictive of AD, is therefore not only essential in understanding the pathology of the disease, but also crucial in identifying patients, assessing their disease progression, and thereby achieving timely management of the disease[14]. An important marker of AD is brain cortical thickness, also known as the AD cortical "signature"[15]. The change of cortical thickness is differentially expressed across brain areas and varies between pre-clinical dementia stages (*i.e.,* subjects with mild cognitive impairment (MCI)) and dementia[16,17]. In general, compared to cognitively normal subjects, individuals with MCI and AD have decreased cortical thickness in the medial temporal lobe region and parts of the frontal and parietal cortices[16–18]. As the disease progresses (from MCI to AD), thinning is observed across the entire cortex, especially in the lateral temporal lobe[16]. Besides structural distribution, cortical thickness of frontal, parietal and temporal lobes in AD is correlated with cognitive impairment[17], while regional thinning predicts (even mild) AD[19].

In addition to cortical thickness changes, the disease is accompanied by multiple cognitive and behavioural disruptions in memory, language, orientation, judgment or problem solving[20]. Furthermore, the disease affects several cognitive subdomains (*e.g.*, attention, executive functions and conceptual thinking) to various degrees[21]. Yet, despite advances in single outcome assessment and prediction, our understanding of the many-to-many (*i.e.*, many brain areas to many outcomes) relationship between the spatially varying cortical thickness changes and multiple symptoms or cognitive dysfunctions, remains limited. To improve our knowledge about and better manage the disease, it is crucial to identify and isolate brain regions each of whose cortical thickness may be differentially associated with a unique cognitive or behavioural outcome, chart the pathways between each set of brain areas and their corresponding outcome, as well as quantify the pathway effect. Equally important is to leverage these pathways and parameters of the identified regions to predict multiple, likely non-pairwise-correlated, cognitive and behavioural scores. That is, to use cortical thickness data from every set of identified regions to predict each corresponding outcome.

Such quests for neurobiological insights and predictive performances require methodological

innovations combined with biological knowledge, through which one can perform the following tasks. (i) Search for AD-specific brain features (*e.g.*, cortical thickness from specific brain areas); (ii) Uncover pathways between high-dimensional brain areas and multivariate disease and behaviour outcomes; (iii) Deal with confounding variables (*e.g.*, age and gender) affecting both features and outcomes and the pathway identification; (iv) Predict multivariate, potentially non-pairwise-correlated outcomes; (v) Predict outcomes cross-sectionally and longitudinally.

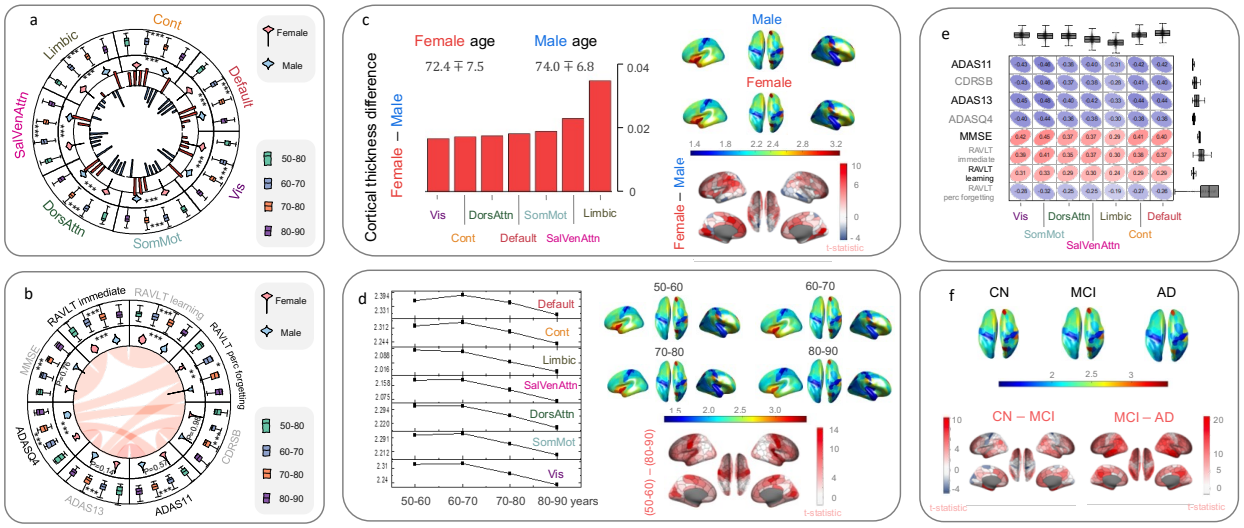

Figure 1. The effect of age and gender on cortical thickness and disease and behaviour outcomes in Alzheimer's Disease. **(a)** Cortical thickness from seven functional brain areas exhibits differences and variations across age and gender groups. The outer circle shows the cortical thickness by age group; the middle circle shows the cortical thickness by gender; the inner circle has eight bar charts per brain region, where each bar represents one of the eight outcomes and shows the averaged weights associated with each outcome prediction (red = positive values and blue = negative values). For visualization purposes, the weights of the bars are normalised to the range of $[-1,1]$. **(b)** Eight Alzheimer's disease (AD)-related outcomes from five examinations are not strongly pairwise-correlated. Additionally, they are differentially presented across age and gender groups. The outer circle shows the outcomes by age; the middle circles show the outcomes by gender; the inner circle shows the outcomes that are not highly correlated (connected lines indicate those that are highly correlated ($r > 0.6$)). **(c)** The mean cortical thickness varies between males and females across seven brain networks (left plot) and across brain regions (right plots). The female cortex is generally thicker than the male cortex. **(d)** The mean cortical thickness differs between age groups across seven brain networks (left plot) and across the brain regions (right plots). **(e)** Distributions of and associative analysis between eight AD-related outcomes and regional cortical thickness data. The boxplots in the top panel show distributions of cortical thickness across seven functional brain areas (indexed in the bottom); the boxplots on the right panel show distributions of cognitive and behaviour scores from eight tests (indexed on the left). The value in each ellipse represents the correlation between each score and cortical thickness data from a corresponding brain region. **(f)** The mean of cortical thickness between different disease groups. For panels (c), (d) and (f), the top rainbow colour bar indicates the normalised (mean/variance) cortical thickness values. The bottom red colour bar represents t-statistics for each region.

First, there is a need to search for *subsets of brain areas* respectively associated with different cognitive and behavioural outcomes. This is important for improving our understanding of the pathology of the disease by separating cortical areas potentially associated with different aspects of the disease and is useful for pathway estimation (see next point). **Second**, there is a need to chart the *pathways* between high-dimensional cortical thickness and multiple cognitive and behavioural outcomes. Cortical thickness in AD is observed across several functional brain regions, each likely

projecting to multiple cognitive and behavioural domains. Identifying the pathways between multiple brain regions and outcomes, therefore, may help to understand how cortical thickness from different areas is related to or affects outcomes. Estimating the size and sign of the pathway parameters could provide important insights into the hierarchy of brain areas that contribute to various cognitive and behavioural domains. Together, they may improve the prediction accuracy of the outcomes (see the last point). **Third**, there is a need to deal with *confounding variables* that affect both brain data, such as cortical thickness, as well as behaviour and disease outcomes and the pathways in between. Indeed, as confounding variables, age and gender affect *both* cortical thickness and outcomes (see **Figure** 1); if not properly treated (for example, ignoring them or only considering their association with the outcomes), they may bias the estimated pathways[23]. In disease analysis and prediction, neglecting confounding effects may yield clinic misinterpretations[24]. **Fourth**, there is a need to *predict multivariate, non-pairwise-correlated outcomes.* Although predictive models built for assessing single outcomes[25] have considerably advanced our understanding of the general aspect (such as disease status[26]) or a specific subdomain (such as the cognitive decline[27]) of AD, single outcome prediction[22] may not capture the multi-dimensional and -functional cognitive and behaviour degenerative landscapes of the disease. This is because, on the one hand, AD affects patients' cortical thickness to various degrees in different brain regions[16,17], and, on the other hand, the disease differentially affects individual cognitive subdomains, such as attention, executive functions, and conceptual thinking[21]. A single outcome prediction, therefore, does not provide a comprehensive global picture of the disease progression and may omit the local effects where the overall progression is manifested in different cognitive and behavioural domains. **Finally**, as a neurodegenerative disease that not only develops in time but also potentially progresses differentially along various cognitive and behavioural domains, it is important to trace the disease progression longitudinally. This may help evaluate or anticipate the cognitive decline and disease conversion early and manage the disease progression timely.

Here, integrating residual learning[28,29], partial least squares (PLS)[30–32], and predictive modeling[25,33], we develop a new method called *Residual Partial Least Squares Learning* (re-PLS Learning) to identify brain areas whose cortical thickness may be associated with AD, chart the pathways from these brain areas to multivariate, potentially non-pairwise-correlated disease and behaviour outcomes, and predict these outcomes at population and individual levels as well as cross-sectionally and longitudinally. To evaluate and demonstrate the efficacy of the method, we first provide its mathematical foundation. Next, we apply the method to data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and discover potential pathways between high-dimensional cortical thickness data and multivariate disease and behaviour outcomes, while controlling for confounding age and gender variables. Subsequently, we use the trained model to predict multiple diseases and behaviour scores in unseen patients. Finally, we use the method to perform longitudinal AD prediction (see **Results** section).

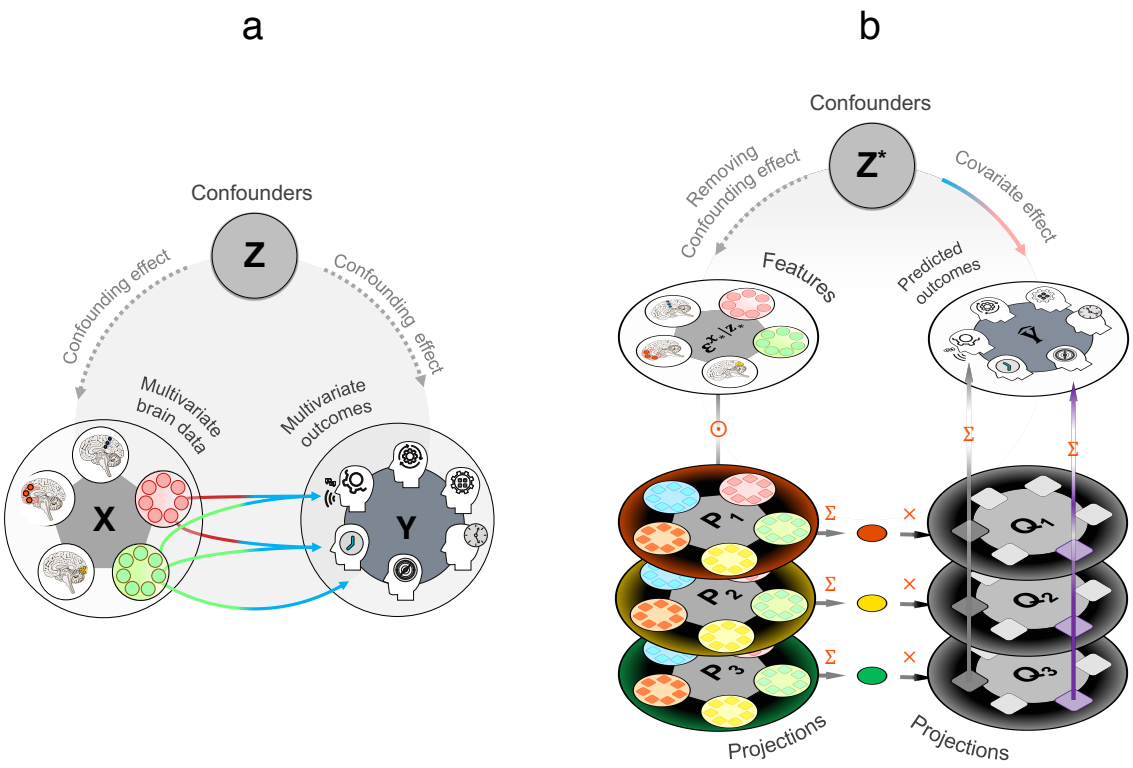

Figure 2. A schematic representation of the Residual Partial Least Squares Learning (re-PLS Learning). (a) A conventional way to predict multivariate outcomes using multivariate brain features. We use X to represent large-scale high-dimensional brain data. Each circle denotes a unique brain area highlighted by a specific colour. The smaller circles within each brain area represent (cortical thickness) data summarised from that brain region. We use Y to represent multivariate outcomes. Each icon represents scores obtained from a cognitive examination (*e.g.*, the Auditory Verbal Learning Test (RAVLT)). We use Z to represent confounders. For simplicity, the circle indicates all confounders (*e.g.*, age and gender) that may affect both the feature variable X, the outcome variable Y, and the pathways between X and Y. In classical prediction problems, one aims at looking for direct pathways between X and Y while controlling for confounding effects from Z. Data from the identified areas are potential biomarkers and are subsequently fed, via the pathways (coloured arrows), to make predictions on new subjects. **(b)** The Residual Partial Least Squares Learning (re-PLS Learning). After controlling for confounding effects from Z, re-PLS is conducted between the residuals of X and Y. Thus, the predictions are not performed directly between X and Y. Rather, the predictions are done via projections (P and Q) which not only extract information from both X and Y unaffected by the confounders Z (via residual learning) but also ensures that the pathways are obtained such that the correlation between X and Y is maximised (via PLS). See the text for mathematical details and neurobiological relevance regarding the re-PLS.

Results

We begin with a summary of five key points regarding our findings. **(1)** Both re-PLS Learning and other baseline models suggest that brain cortical thickness predicts multiple, non-pairwise-correlated behavioural and disease outcomes in AD (see **Figure** 3). **(2)** Re-PLS Learning and PLS yield higher estimation accuracy than other models while re-PLS additionally controls for the confounding variables (see **Supplementary Materials**). **(3)** After removing the age and gender

effects, cortical thickness changes that are significantly predictive of the eight cognitive and behavioural outcomes are mainly in the temporal, frontal, sensorimotor, and cingulate regions (see below for a discussion and **Figure** 4). **(4)** The method is useful for predicting longitudinal disease progression and seems, particularly, promising to chart the disease course for subjects who convert from mild cognitive impairment to AD over time (see **Figure** 5).

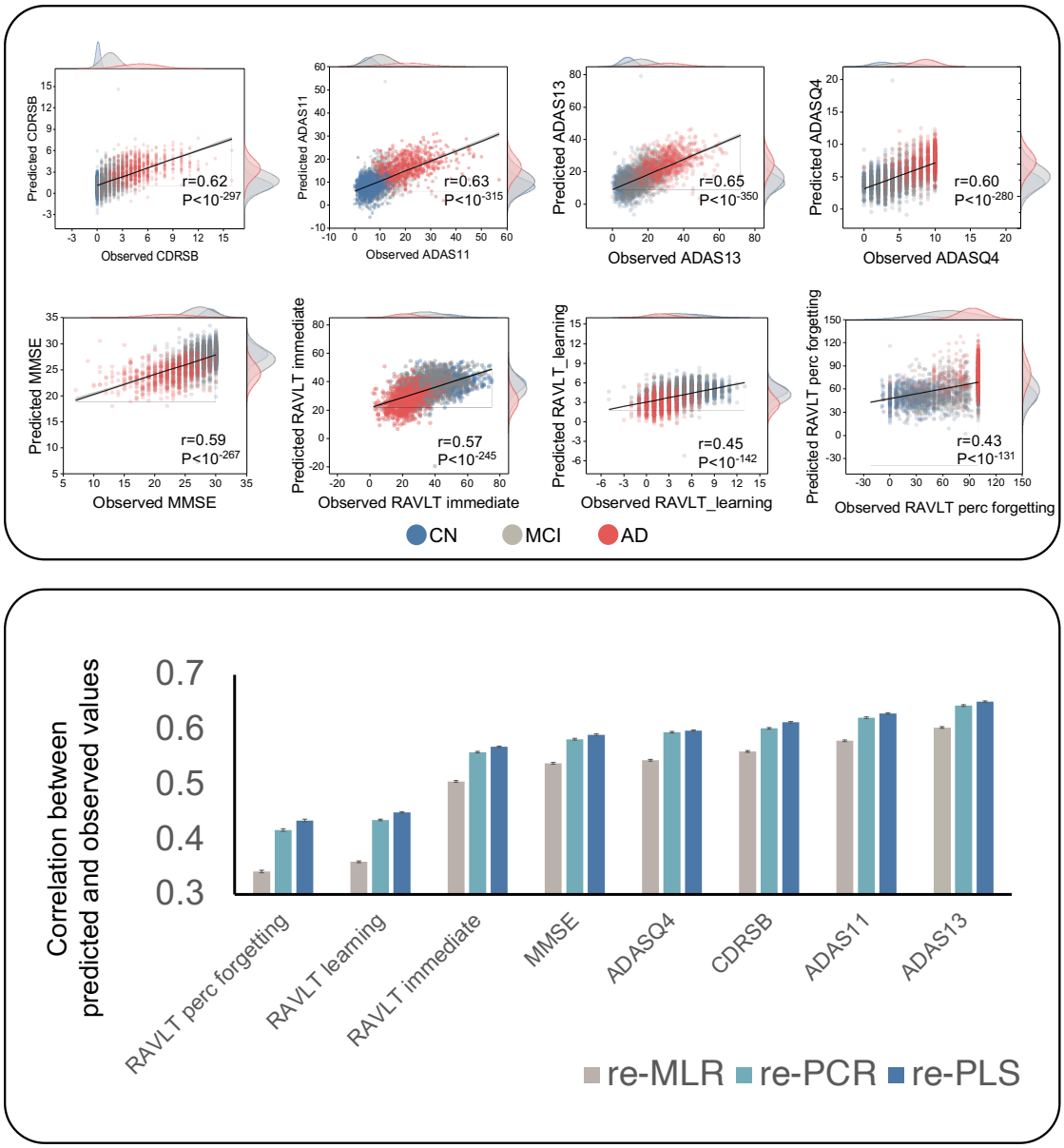

Figure 3. Model comparison between residual Partial Least Squares learning (re-PLS Learning) and prominent methods for predicting multivariate outcomes in previously unseen subjects. (a) Scatter plots of the predicted outcomes against the observed outcomes. Derived *PQ maps* (see **Eqs.** (2) and (3)) from the training sample are used to predict eight cognitive and behavioural outcomes in previously unseen patients using **Eq.** (6). Results are cross-validated (CV) using a 10-fold CV; only predictions made on unseen subjects are shown. **(b)** A comparison between the re-PLS with two common baseline methods. To avoid discrepancy, the confounders in all methods are treated using residual learning. Here, re-PCR refers to principal component regression with confounders controlled via residual learning and re-MLR refers to multivariate linear regression with confounders controlled via residual learning.

The plot shows the mean correlation coefficient of outputs obtained from 1000 bootstrapped samples. Overall, re-PLS Learning yields the best result across eight outcomes. Only results from out-of-sample predictions were shown.

The inputs to the algorithm are cortical thickness data derived from ADNI MRI imaging data (see **Data and preprocessing**). The observed outcomes are eight cognitive and behavioural scores from five AD examinations. These outcomes were chosen for two reasons. First, they are from standard AD examinations that quantify the functioning and dysfunction of various cognitive aspects of the disease, including the Clinical Dementia Rating (CDR), the Alzheimer's Disease Assessment Scale–Cognitive (ADAS-COG), the Mini Mental State Examination (MMSE), the Rey Auditory Verbal Learning Test (RAVLT), and the Montreal Cognitive Assessment (MoCA) (see **Table** 2 for details about these scores). Second, the cognitive metrics were selected such that some of them are highly correlated and some are not (see **Figure** 1e). This is to represent a real-world scenario where some of the outcomes may be significantly correlated while others are not, and to assess the performance of our method in dealing with general cases. Including non-pairwise-corrected outcomes is important because, first, from a practical point of view, it is unlikely that all disease outcomes are significantly correlated, and second, from a predictive point of view, if a set of (brain) data (*e.g.*, multivariate cortical thickness) can significantly predict an outcome, they can similarly predict a second outcome that is highly correlated with the first outcome (thus one only needs to make one prediction). By allowing non-pairwise-correlated outcomes, we can verify the method's performance in general cases and apply it in real-world case studies.

We summarise the experimental setup in **Figure** 2 and **Algorithm** 1. The analyses were performed using a custom *Python* package (*re-PLS; v 1.0*) [https://pypi.org/project/rePLS]. In **Figure** 3, we present the model's performance on AD outcome prediction. In **Figure** 4, we identify and present the brain areas whose cortical thickness is respectively predictive of each outcome. In **Figure** 5, we present the results of longitudinal AD prediction.

After demonstrating the *methodological properties* of the proposed framework, we report the *neurobiological findings*. In brief, our findings suggest a few new insights about five almost orthogonal "predictive AD markers" that are, respectively, predictive of multivariate, non-pairwise-correlated cognitive, behavioural, and disease outcomes for AD. Our findings also indicate potential "longitudinal AD markers" that are useful to depict the disease course over time, and, particularly, predictive of the longitudinal disease progression for subjects converting from mild cognitive impairment (MCI) to Alzheimer's disease. Additionally, our results provide converging evidence confirming some previous findings. In the following, we present these findings in detail.

Cross-sectional AD assessment

We first aim to identify and separate brain regions associated with and predictive of each AD-related cognitive and behavioural outcome under a cross-sectional setting where scans from each individual are treated as independent repeated measurements. This helps identify brain regions linked with each disease outcome. We then extend this to a longitudinal study in the next section where we identify brain regions potentially degenerate vis-à-vis disease outcome progression over

time. Throughout, when we refer to "predictive" or "prediction", we specifically mean out-of-sample prediction.

To that end, we extract the latent brain spaces (*P* maps in **Figure** 2) which correspond to the identified brain markers. We then look at the projections linking the latent brain spaces to the AD-related outcomes (*Q maps* in **Figure** 2). Overall, our results (i) unveil how brain regions whose cortical thickness may be functionally and anatomically linked to multiple AD outcomes (see **Figure** 4a-c); (ii) suggest that cortical thickness predicts multivariate non-pairwise-correlated outcomes (see **Figure** 3 and **Figure** 4d-e); (iii) show that re-PLS alleviates multicollinearity and confounding effects, exhibits stability to sample size variations, and improvs multivariate outcome prediction performance (see **Figure** 4e-f).

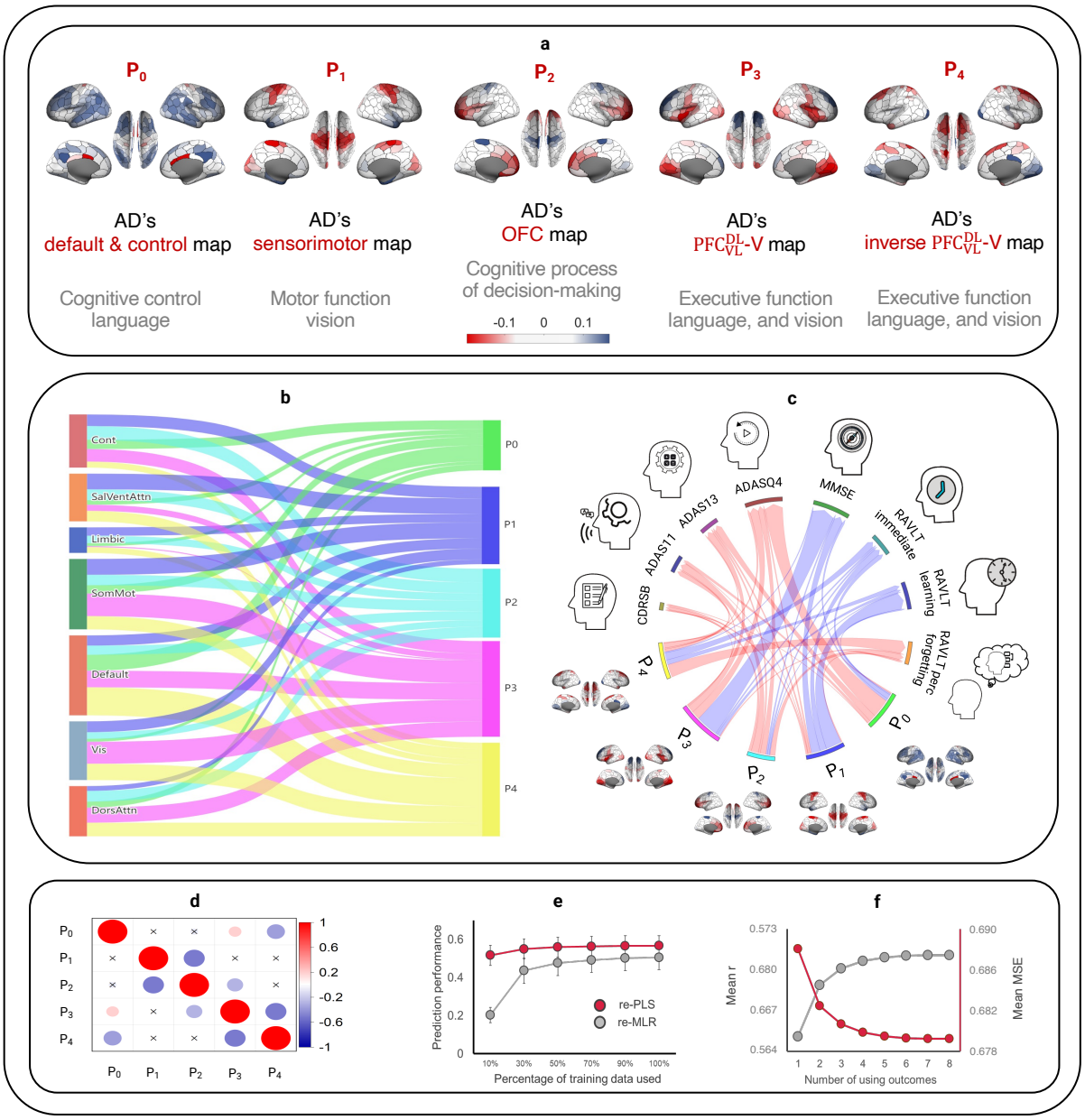

Figure 4. Discovering brain areas predictive of multivariate outcomes using the residual Partial Least Squares learning (re-PLS Learning). (a) Five latent brain spaces for cortical thickness were identified by re-PLS. The five brain spaces are selected using cross-validations. In the colour scheme, positive coefficients are represented in red, while negative coefficients are shown in blue. **(b)** The functional localization of the five projections in terms of cortical thickness. The link between functional brain regions and the five identified projections is based on the absolute value of the weights: each projection is the sum of absolute value contributions, with non-significant differences removed. **(c)** Relationship between the latent brain spaces and the eight behaviour and cognitive outcomes. The five projections (P_0 to P_4) are plotted in their corresponding brain space. The width of the lines from P_i ($i = 0,1, \dots, 4$) to each outcome is estimated using α times the Q *map* (see **Eqs.** (3) and (4)); it indicates the contribution each set of brain regions makes to the outcome prediction. Red links represent positive weights, and blue links represent negative weights. The width of the coloured bar next to each outcome is estimated using the sum of absolute weights. **(d)** The (in)dependency between the five latent brain spaces. The size of each circle is proportional to the correlation between a pair of latent brain spaces, with "x" marking insignificant results (P-value > 0.05). **(e)** Sample size studies for the re-PLS. During each bootstrap training, we compute a model using different percentages (from 10% to 100%) of the training samples and assess model performance on the testing samples. **(f)** Prediction accuracy improves as more outcomes are included. This is an inherited property of the re-PLS (through the learned projections) where each added outcome assists in the prediction of other outcomes.

First, we identify anatomical and functional regions whose cortical thickness is predictive of AD outcomes using the P maps (each entry of a P map corresponds to a set of brain regions). Specifically, P_0 encompasses the default mode and the control network; it consists predominantly of areas in the prefrontal, temporal, and cingulate cortex. P_1 is largely located in the sensorimotor area with small parts in V1 and V2. P_2 seems to highlight the cognitive brain – with most weights in the prefrontal cortex and a small region in the motor cortex. P_3 is positively linked to visual areas and negatively linked to the superior frontal lobe. Finally, P_4 resembles the opposite version of P_3, with positive values in the superior frontal lobe and negative areas in temporal and visual areas. We highlight that the positive and negative values in the P maps indicate whether an area is a *positive or negative predictor* of the outcome; the signs do not suggest an increase or decrease in cortical thickness.

Second, the latent brain spaces (P maps) provide important insights about potential AD markers. **(1)** Several P maps highlight the prominence of cortical thickness in the **temporal areas** (*e.g.*, BA20, BA21 and BA22 in P_0 and BA38 in P_1 and P_4) in predicting AD-related outcomes. Previous findings suggest that BA20 (inferior temporal) is involved with semantic memory processing[34] and BA21 (middle temporal) in processing language and higher-order audition processes[35]. Accumulated tau deposition in BA20 and BA21 has been shown to be associated with clinical impairments observed in Alzheimer's disease[58]. BA22 (superior temporal) in Wernicke's area is involved in the comprehension of written and spoken language and is shown to be engaged in tasks that necessitate cognitive switching[59]. BA38 participates in semantic processing, speech comprehension, and naming[62]. **(2)** Several P maps highlight the importance of cortical thickness in the **frontal areas** in predicting AD-related outcomes: the superior and middle frontal areas (P_0 and especially BA9, BA10 and BA46), middle frontal areas (P_2 and especially BA10 and BA11) and inferior frontal areas (P_3 and especially BA44 and BA45). BA9 (which contributes to the dorsolateral prefrontal cortex or DLPFC) is linked with short-term memory, spatial memory, working memory (WM), recognition, recall and calculation. BA9 is shown to cause profound WM impairment. Additionally, it is shown to be important in task inference such as the Stroop task,

Eriksen flanker task and task switching[60]. BA10 is related to working memory, episodic memory and mentalizing[45]. BA10 is known to support higher cognitive functions such as task management and planing[61]. DLPFC (BA46) in the right hemisphere is relevant for spatial working memory [53]; image-based representations of objects are mostly activated in the left BA46; right BA46 is related to working memory [55]. Evidence also suggests that BA46 is significantly involved in delayed-response spatial working memory tasks [54]. BA11 is involved in decision-making, processing rewards, and encoding new information [56, 57]. BA44 (phonological processing) and BA45 (semantic processing) are in Broca's area and are involved in speaking but also language comprehension[66]; additionally, BA45 is thought to be associated with increased activation in letter fluency tasks[67]. **(3)** Parts of the **parietal** lobe (especially BA39 and BA40 in P_0) are significantly predictive of the cognitive and memory scores. This confirms previous findings where AD patients undergo cortical thickness changes in the parietal cortices[16–18]. Importantly, BA39 (angular gyrus or AG) has been shown to correlate with longitudinal declines in verbal fluency[63]. BA40 (supramarginal gyrus) is thought to be "involved in reading both as regards meaning and phonology" and "damage to the left BA39 may result in dyslexia or semantic aphasia". This seems to be in line with the finding that P_0 is strongly linked to ADASQ4 (Delayed Word Recall task). The ventral parietal cortex (BA39 and BA40) is involved in recollective processing, including detailed retrieval of contextual information and vivid re-creation of past experiences tasks[64]. The angular gyrus is thought to be associated with cued recall deficits, particularly evident in cross-modal picture-sound pair tasks, though significant parietal lesion effects were also observed in unimodal word pair and picture pair tasks[64]. **(4)** Our results hint at the roles **sensorimotor areas** play in predicting AD-related outcomes. Although some have argued that sensory and motor changes may precede the cognitive symptoms of AD[46] , since the eight outcomes in this study measure various cognitive abilities, our findings cannot distinguish whether the changes in cortical thickness in sensory and motor areas (thus changes in sensory and motor functions) hinder the movement during the examinations, thereby affecting the performance on the eight scores, or if they contribute, in concert with other areas, to the performance during the tests. Further research needs to independently verify this. **(5)** Our results suggest that **ACC**, and dorsal **PCC** in BA31 (in P_0 and P_4) are predictive of AD-related outcomes. Previous studies found amyloid deposition and reduced metabolism in the PCC are early signs of AD[38–40] often present before definitive clinical diagnosis[38, 42]. In addition, previous studies have demonstrated an association between abnormal PCC and impaired performance on executive function tasks[63]. The role of ACC in AD is less clear [43]. **(6)** Our results suggest that cortical thickness in **the visual cortex** may be associated with attentional and visual memory-related word remembering. Particularly, BA18 (V2) is thought to be related to attentional modulation[68] and, arguably, to visual memory[69].

Latent brain spaces	Brodmann areas and relevant known functions	Prominently associated outcomes
P_0	**BA9** (short-term-, spatial-, and working- memories, recognition, recall and calculation[60]), **BA20** (semantic memory processing[34]), **BA21** (language and higher-order audition processes[35]), **BA39** and **BA40** (recollective processing[64]), **BA46** (spatial working memory[53]), **ACC**.	**ADAS13** (11 items assessing cognitive function plus delayed word recall and number cancellation), **ADASQ4** (delayed word recall).

P_1	**BA38** (semantic memory[62]), **BA17** (V1), **BA18** (V2, attentional modulation[68] and arguably visual memory[69]), **BA4** (voluntary movements), **BA6** (premotor area).	**RAVLT learning** (measuring the number of words remembered across all trials), **CDRSB** (total score regarding memory, orientation, judgment, problem-solving, personal care, home and hobbies, and community affairs), **RAVLT immediate** (test word recall using multiple trials after a time delay).
P_2	**BA10** (working memory, episodic memory and mentalizing[45]; higher cognitive functions such as task management and planing[61]), **BA11** (decision making, processing rewards, and encoding new information [56, 57]), **BA4** (voluntary movements).	**ADAS11** (11 items assessing cognitive function), **CDRSB**.
P_3	**BA44** and **BA45** (Broca's area; related to phonological and semantic processing, respectively. BA45 is associated with increased activation in letter fluency tasks[70]), **BA9** (short term-, spatial-, and working- memories, recognition, recall and calculation[60]), **BA17** (V1), **BA18** (V2, attentional modulation and arguably visual memory).	**MMSE** (orientation to time and place, recall, attention, calculation, and language), **ADAS13**.
P_4	**BA7** (visuomotor coordination[71]), **BA8** (management of uncertainty[72]), **BA38** (semantic memory), **right PCC** (executive function tasks[63]).	**RAVLT percent forgetting** (measures the percentage of words missed), **RAVLT immediate**.

Table 1. Relationship between identified latent brain spaces, their anatomical correspondence, and associated behaviour and cognitive outcomes. For tests that appear several times, we list their items once for brevity.

Third, we further investigate how the identified brain areas are distributed across seven network functional[52] regions. We notice that P_0 areas are located predominantly in the default and control regions; P_1 areas are prominently in the sensorimotor regions followed by attention regions; P_2 areas are primarily in the cognitive and attention regions; P_3 areas are in the dorsolateral and ventrolateral of the PFC regions followed by visual areas; and P_4 regions are similar to those of P_3 but of opposite signs (see **Figure** 4a and 4b). Furthermore, a correlation analysis between these five projections showed that they were nearly orthogonal to each other (see **Figure** 4d). This suggests that the *P* maps identify and isolate orthogonal functional brain areas that are predictive of multivariate outcomes. Based on the functional and anatomical separation (due in part to their orthogonality) of the *P* maps, we designate P_0 as AD's **default and control map** (P_0), P_1 as AD's **sensorimotor map**, P_2 as AD's **OFC map**, P_3 as AD's $\boldsymbol{PFC_{VL}^{DL}}$**-V map** (DL-PFC, VL-PFC, and Visual) and P_4 the **inverse** $\boldsymbol{PFC_{VL}^{DL}}$**-V** map (see **Figure** 4a).

Fourth, we link the latent brain spaces (*P* maps) with the eight behavioural and disease outcomes using the *Q maps*. We found that: **(1)** The outcomes embed different variability across subjects (see **Figure** 4c). Noticeably, ADASQ4 and MMSE show the most variability between subjects. The three RAVLT-related outcomes show similar variability, although they map to different *P* maps. Finally, CDRSB, ADAS11, and ADAS13 show small variability. **(2)** Among *P* maps, AD's

default and control map (P_0 map), AD's sensorimotor map (P_1 map), and AD's PFC_{VL}^{DL}-V map (P_3 map) show the most contribution to predict the eight outcomes. **(3)** Each P map shows distinctive importance for the outcomes. For predicting CDRSB, the most influential predictive brain areas are in the P_1 and P_2 regions. Brain areas in P_2 seem to be also strongly predictive of ADAS11. For predicting ADAS13, regions in P_0 and P_3 are playing important roles. ADASQ4 has dominant links with P_0, and MMSE scores are closely linked to P_3. Furthermore, RAVLT immediate recall is associated with P_1 and P_4; RAVLT learning performance is primarily linked to P_1; and RAVLT percent forgetting is attentively related to P_4. **(4)** The links between P maps and eight outcomes suggest how specific brain areas may neurobiologically underpin the cognitive and behavioural attributes of AD. From **Table** 1, one can see that the AD's **default and control map** (P_0) consists of regions in the temporal lobe (including BA20 and BA21), the default mode, the cognitive control network, and parts of the parietal cortex (BA39 and BA40). The main difference between P_0 and other maps is that P_0 consists of temporal lobes (BA20 and BA21), which contribute to memory processing[34] and language and higher-order audition processes[35], and the parietal lobe (BA39 and BA40), which contribute to recollective process[64]. These findings seem to suggest why the P_0 map is useful for predicting ADAS 13 (which assesses cognitive function plus delayed word recall and number cancellation) and ADASQ4 (which assesses delayed word recall). Additionally, previous studies have shown reduced numbers of somatostatin receptors[36] and synapse loss in BA21 in AD patients[37]. Previous studies have also shown that cortical thickness changes greatly in the lateral temporal lobe in AD patients[16–18]. Our results thus not only confirm the localization of cortical thinning in AD in previous studies but also demonstrate that cortical thickness data from these regions are predictive of multiple AD-related cognitive functions (or dysfunctions). AD's **sensorimotor map** (P_1) consists of a portion of the temporal lobe (BA38), a portion of the motor cortex (BA4 and BA6), and a portion of the visual cortex (BA17 and BA18). Importantly, BA38 is associated with semantic memory[62] and BA18 (V2) is related to attentional modulation[68] and arguably visual memory[69]. These seem to be in line with our findings of P_1's pronounced role in predicting RAVLT learning (which measures the number of words remembered across all trials), CDRSB (which evaluates, in part, memory tasks) and RAVLT immediate (which also examines memory by testing word recall using multiple trials after a time delay). AD's **OFC map** (P_2), which includes a portion of the primary motor cortex (BA4) and orbitofrontal cortex (BA10 and BA11), has an important impact in each of the eight outcomes, but noticeably for predicting ADAS11 and CDRSB, both related to testing general cognitive function. AD's $\boldsymbol{PFC_{VL}^{DL}}$**-V map** ($P_3$) consists of the superior (BA9) and inferior (BA44 and BA45) frontal cortex, and the visual area. The main differences between the P_3 map and other maps are in Broca's area (involved in speaking but also comprehensive of language). This seems to be in line with the fact that P_3 is strongly predictive of MMSE (evaluating, in part, short-term verbal memory and language). AD's **inverse** $\boldsymbol{PFC_{VL}^{DL}}$**-V map** ($P_4$) includes right PCC, a portion of the parietal lobe (BA7 and BA8), and a portion of the temporal lobe (BA38). The primary distinction between the P_4 map and other maps is in the parietal regions, which facilitate WM and serial ordering in WM[65]. Additionally, areas BA7 and BA8 are thought to be related to visuomotor coordination[71] and management of uncertainty[72], respectively; this seems to explain, in part, the predictability of the P_4 map and the RAVLT percent forgetting score which measures the percentage of words missed). **(5)** Although the P maps are almost orthogonal, there are some commonalities/overlaps between areas identified by the P maps and their pathways towards the eight outcomes. Noticeably, P_0 and P_3 maps share the dorsal

prefrontal cortex (BA9), involved in WM and executive functions. Both also have a strong correlation with ADAS13. P_1 and P_2 maps share the motor cortex (BA4), responsible for controlling voluntary movements; both have a high weight in predicting CDRSB. P_1 and P_4 maps share the temporal pole (BA38), involved in semantic memory (storing and retrieving factual knowledge); both have a strong correlation with RAVLT immediate recall.

Fifth, we probe into how re-PLS alleviates multicollinearity and confounding effects, exhibits stability to sample size variations, and improves multivariate outcome prediction performance. We investigate the impact of varying sample sizes on the performance of re-PLS as compared to a popular method, residual learning-aided multivariate linear regression (re-MLR). Our results suggest that overall, re-PLS outperforms re-MLR. More specifically, it is challenging for re-MLR to perform prediction, especially when the training data is small (see **Figure** 4e). In comparison, re-PLS seems to deliver better overall prediction accuracy across different training data sizes and is more consistent when training data sizes vary. Additionally, re-PLS seems to require less training data to achieve optimal prediction performance. For example, to achieve comparable results using 70% of the training data by re-PLS, re-MLR requires nearly 90% of the training data.

Sixth, we notice that re-PLS achieves higher prediction performance (both in terms of mean square error and in terms of correlation) as the number of outcomes increases (see **Figure** 4f). This is possibly due to the nature of re-PLS. The hidden projections aim to maximise the associations between the inputs (cortical thickness) and outcomes (disease scores) when controlling for covariates; thus, when making predictions, the prediction of each outcome is made by using the information of the inputs (cortical thickness), the covariates, and the projections (which also takes in the information of the outcomes). In other words, when predicting one outcome, re-PLS uses information from cortical thickness, covariates, and the *other outcomes*. Although the outcomes are not all pairwise correlated, each association between two (even modestly) correlated outcomes would make one a useful predictor of the other. Thus, the more outcomes, the better prediction performance. Certainly, in an extreme case, when all outcomes are identical, adding additional outcomes may not improve prediction performance.

Finally, re-PLS achieves higher prediction accuracy compared to conventional linear approaches. It is likely that the lower-dimensional "almost orthogonal" projections contain reduced noise in contrast to the original high-dimensional data. These refined projections (lower-dimensional features) seem to achieve more effective data representation in the latent space and lead to improved prediction performance. As a result, re-PLS not only assisted neurobiological explanation via the extracted latent brain spaces, but also requires a smaller dataset to achieve similar prediction performance compared to both re-PLR and re-PCR. This property may be useful in situations where one has limited data but multivariate complexity.

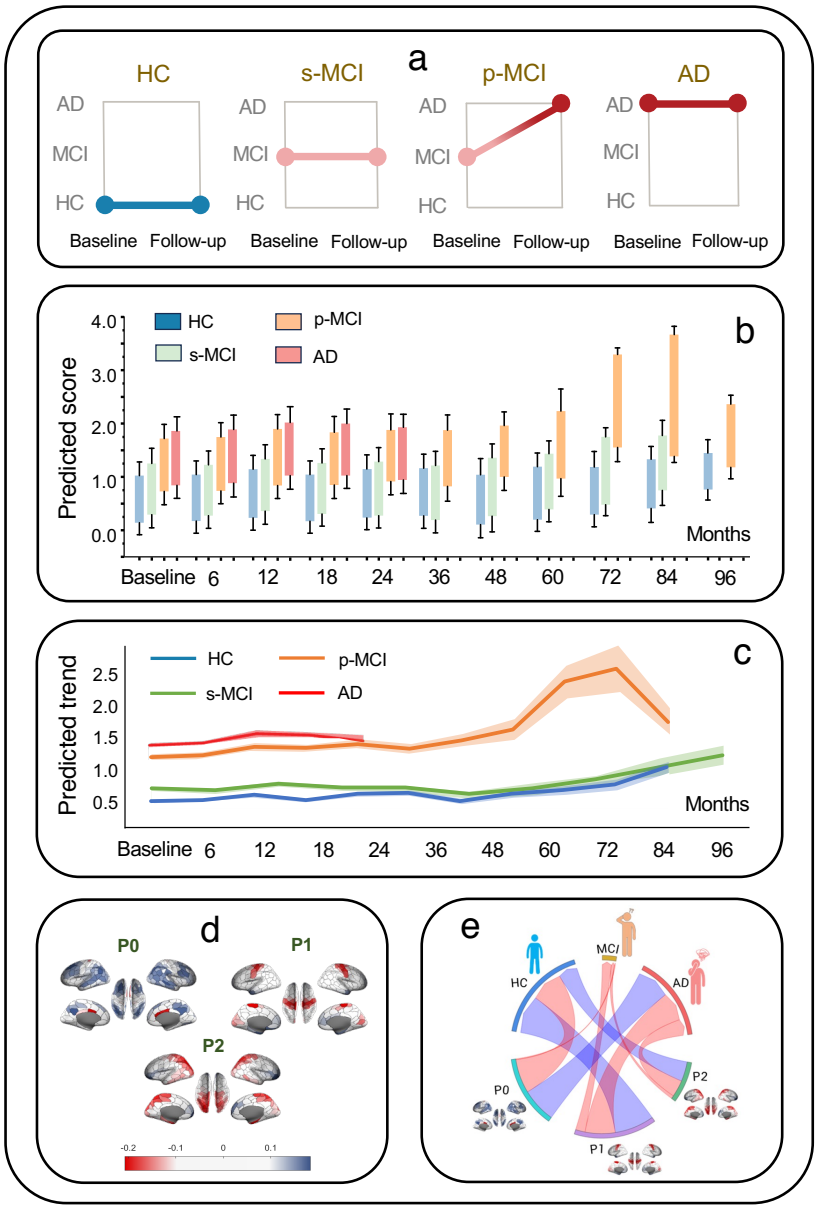

Figure 5: Longitudinal prediction of Alzheimer's disease. (a) Four types of subjects in the longitudinal study. CN = cognitively normal, sMCI = stable MCI (a subject assessed as an MCI during the first visit and continued to be diagnosed as an MCI during subsequent visits), pMCI = progressive MCI (a subject was an MCI during early visits and was later diagnosed with AD), and AD = Alzheimer's disease. **(b)** Predicted longitudinal AD score. The x-axis denotes the number of months since the initial visit, and the y-axis indicates the predicted scores using the re-PLS. Note that, to avoid overfitting and chance split, we perform a 10-fold cross-validation (CV) and only report the predicted disease scores for new subjects in the testing set. **(c)** Longitudinal trend prediction. The longitudinal curve for each group is estimated using the predicted mean group scores for new subjects at each time point. The width of the 95% confidence bands (shaded colour) is estimated using a repeated 10-fold CV (run 100 times). In general, the predicted longitudinal severity is AD > pMCI > sMCI > CN. The pMCI is predicted to worsen more than other groups over time. **(d)** Latent brain spaces are potentially related to longitudinal AD progression identified by the re-PLS. **(e)** Relationship between longitudinal latent brain spaces and diagnostic outcomes. The width of the lines between three latent brain spaces and three types of diagnostic outcomes indicates the size of the association. Blue lines represent negative coefficients, and red lines represent positive coefficients. The strength of connection was estimated using the magnitude of the coefficient, quantifying the contribution each brain space makes to predicting the target outcome.

Longitudinal AD assessment

Our analyses so far have focused on the cross-sectional aspect of the disease (where multiple samples from each subject are treated as repeated measures, *i.e.*, without temporal order). Whereas it is useful to identify brain areas that are, in general, related to AD and that are predictive of the overall multivariate cognitive and behavioural scores, it only provides a snapshot of the disease. As a neurodegenerative disease, AD progresses over time[3-4]. Naturally, one would ask if it were possible to expand the analyses to longitudinal settings to study disease progression. This is an important task, for two reasons. First, it is important to monitor and forecast disease progression to improve disease management and treatment. Second, it is important to identify brain areas whose degenerations are related to cognitive decline over time to gain insights into how the disease progresses and which brain regions contribute to disease progression over time, and if so, to what extent.

To that end, we use re-PLS to study longitudinal AD prediction. We do so in two settings. First, we extend the use of re-PLS from cross-sectional analysis to predict AD status over time (see **Figure** 5a-b). Second, we identify brain regions whose cortical thickness may be related to AD progression over time. We note that we conduct longitudinal disease prediction on disease status but not on the eight outcomes. This is because longitudinal multivariate disease prediction inspects the variability over time and across eight disease scores – it needs a much larger sample size to obtain reliable results. Further studies, however, will be able to examine longitudinal prediction on the multivariate regression task when larger data sets are available, and the flexibility of the re-PLS allows for performing such analysis.

Throughout, we refer to the statuses of AD, MCI, or lack thereof as diagnostic outcomes made by clinicians primarily based on clinical criteria. For more detailed information, see the ADNI2 Procedures Manual at: https://adni.loni.usc.edu/wp-content/uploads/2024/02/ADNI2_Procedures_Manual_28Feb2024.pdf. Specifically, every subject is diagnosed with one of the disease statuses: Cognitively Normal (CN), Mild Cognitive Impairment (MCI), or Alzheimer's Disease (AD), based on ADNI criteria. For modelling, we assign a disease status of 0, 1, or 2 to represent CN, MCI, and AD, respectively. In our analysis, we further group the individuals into four distinct longitudinal groups based on disease progression: Cognitively Normal (CN): individuals who were assessed as cognitively normal and maintained cognitively normal during subsequent visits; stable Mild Cognitive Impairment (sMCI): individuals who were assessed as MCIs during the first visit and continued to be diagnosed as an MCI during subsequent visits; progressive Mild Cognitive Impairment (pMCI): individuals who were assessed with MCI during early visits but diagnosed with AD during subsequent visits; Alzheimer's Disease (AD): individuals who were assessed having AD throughout the visits. As some subjects have missing data at baseline, we consider their earliest scans as baseline data and arrange their later scans accordingly.

After training the longitudinal re-PLS model, we implement it to predict previously unseen individual subjects' disease status over time. Although we grouped every subject into one of the four groups – the group information for testing subjects was not used (to avoid information leakage); rather, the (four) group information was used to colour code the testing subjects to

evaluate the accuracy of the longitudinal prediction performance (see **Figure** 5b). We saw that the predicted overall mean scores increased from CN, sMCI, pMCI, to AD. This generally agrees with the actual diagnostic outcomes. Additionally, the predicted longitudinal trend for pMCI subjects (subjects who were MCIs during early visits and were later diagnosed with AD) seems to worsen noticeably more than the other groups (see **Figure** 5c). This is also consistent with their observed diagnostic longitudinal progression. The predicted trends for both CN and MCI groups are relatively stable, in line with their observed longitudinal diagnostic statuses, although our method predicts that both groups have a slight worsening sign after 60 months – presumptively because of a gentle cortical thinning due to ageing.

Next, we seek to unveil the brain regions whose longitudinal cortical thickness change may be potentially associated with and predictive of AD disease status over time. To that end, we extract the longitudinal latent brain spaces (the longitudinal version of the cross-sectional P maps), which uncover brain regions that may be associated with longitudinal disease progression over time. Concretely, we encode the disease status as a one-hot vector instead of scalars (0, 1, and 2) for the three possible outcomes (CN, MCI, and AD). Overall, we identify three longitudinal latent brain spaces. The longitudinal P_0 map is linked to the default mode and prefrontal regions, including the temporal and cingulate cortex; the longitudinal P_1 map corresponds to the motor cortex; and the longitudinal P_2 map is associated with parts of the temporal, parietal, and occipital areas, and a small portion of the prefrontal area. Interestingly, the first two longitudinal P maps in **Figure** 5d bear a striking similarity to the first two cross-sectional P maps in **Figure** 4a. This suggests that cortical thickness from these areas may be useful markers for both cross-sectional and longitudinal AD studies. Next, we use the *Q maps* to further inquire into how the identified "longitudinal brain markers" (using longitudinal P maps) may be predicting disease status over time. We find that the longitudinal P_0 and P_1 maps contribute significantly to the prediction of CN and AD, and the longitudinal P_1 map is most influential for predicting MCI. Additionally, we observe that CN is negatively associated with lower thickness values in the longitudinal P_1 and P_2 maps, while it is positively linked to higher thickness values in the longitudinal P_0 map; the pattern for AD is reversed (see **Figure** 5e).

Discussion

Identifying pathways between high-dimensional multivariate brain data and multivariate, non-pairwise-correlated behavioural, cognitive, and disease outcomes is central to advancing our knowledge about how spatial distribution and functional integration of cortical irregularities may give rise to neurodegenerative diseases. It is also critical to predicting disease progression that may manifest across different behavioural, cognitive, and disease domains. In this article, we develop the re-PLS Learning to (1) chart the pathways between high-dimensional multivariate brain cortical thickness data (inputs) and multivariate disease and behaviour data (outcomes); (2) simultaneously predict multiple, non-pairwise-correlated outcomes; (3) control for age and gender (confounding variables) affecting both the inputs, the outcomes, and the pathways in-between; (4) assess disease scores cross-sectionally and disease progression longitudinally.

The re-PLS framework first obtains the residuals containing information on cortical thickness and outcomes that are not affected by the confounders via residual learning. It then performs PLS between the brain data-specific and outcome-specific residuals to estimate feature weights that take both cortical thickness information and outcome information into account. The model finally uses the residuals, the confounders (now covariates), and the parameters estimated from the PLS to assess multivariate disease outcomes in new subjects.

We examine the method's efficacy using data from CN, individuals with MCI and AD patients from the ADNI, a multi-centre study aiming at developing biomarkers for AD[47]. Our results show that the re-PLS framework is promising for identifying, separating, and estimating unique pathways between high-dimensional cortical thickness data and multivariate cognitive and behavioural scores. The identified brain regions are mainly in the temporal, frontal, sensorimotor, and cingulate regions, supporting previous findings[16–18,38–42]. Additionally, our results have provided a few new insights: we identify several almost orthogonal "predictive AD markers" that are, respectively, predictive of multivariate outcomes related to different behavioural and cognitive traits of AD. Finally, extending the model to longitudinal settings, we discover potential "longitudinal AD markers" that are not only useful to unveil how AD may expand spatially in the cortical areas over time, but also promising to help assess longitudinal disease course and predict disease progression, for example, to estimate when subjects may be converting from mild cognitive impairment to Alzheimer's disease.

To showcase the generalisability and reproducibility of the re-PLS, we perform a 10-fold cross-validation (CV). The model is iteratively trained on nine folds of the data and tested on the remaining fold without further model fitting (note that no subjects from the training data are in the testing set). It then iterates, training the model on nine new golds and testing it on the new reaming fold, and so on. Although our results in **Figure** 4 highlight that parameters and pathways learned from the training data are useful to predict multivariate AD outcomes in previously unseen subjects, it remains possible that the model may not capture the data variability across folds. To that end, we perform ten additional analyses with different CV settings. Specifically, we set aside $x\%$ (where $x = 0, 10, 20, \dots, 90$) of the data for an additional step of out-of-sample test and run LOOCV on $(100 - x)\%$ of the data; when $x = 0$, one runs LOOCV on the entire ADNI data. To avoid a (un)lucky split (e.g., the training data contain many subjects with AD and MCI and the testing data contain many CNs), we perform stratified sampling. Taking $x = 70$ as an example, we randomly select 70% of AD subjects, 70% of the people with MCI, and 70% of CN – they form the training set, which is proportional to and representative of the entire data. The results show that the brain maps in the additional analyses, across various cross-validation settings, are generally consistent with those in **Figure** 4 via 10-fold CV (see **Supplementary Materials**). Additionally, the multivariate outcome predictive performance remains high and is consistent among different CV settings. The average correlation between predicted outcomes and observed outcomes, where the average is done on eight correlations from eight outcome predictions, ranges from 0.557 (LOOCV on 10% of the data and testing on the independent 90%), 0.658 (LOOCV on 70% of the data and testing on the remaining 30%), to 0.659 (LOOCV on 90% of the data and testing on the remaining 10%). Specifically, across all CV settings, the extracted brain maps (*P maps*) converge (the changes becoming increasingly small) as more data are used for training. Across all CV

settings, the default and control map (P_0) and the sensorimotor map (P_1) are almost identical between these 10 additional CVs and with the 10-fold CV. The OFC map (P_2) and the PFC_{VL}^{DL}-V map (P_3) are also consistent up to a sign (the brain areas are similar with similar weights of importance, but the signs of the weights may flip) and become increasingly stabilised as more data are used for training. The sign flip, however, does not affect interpretation and prediction. This is because (a) the method identified the same brain regions; (b) a sign flip does not affect prediction: if the P_i map has a sign flip, the corresponding Q_1 map has also a sign flip, thus $P_i \times Q_i$ remains the same for **Eq.** (6). The P_4 map shows more variability across CV settings. The P_4, however, explains the least amount of variability and may be subject to noise. Taken together, our additional cross-validations suggest the utility of the re-PLS in predicting multivariate outcomes and that the model performance and neurobiological explanation are consistent across different cross-validation mechanisms. Particularly, the consistency of the default and control map (P_0), the sensorimotor map (P_1), the OFC map (P_2) and the PFC_{VL}^{DL}-V map (P_3) as well as their convergence property as more data are used, suggest the strong plausibility of them being sensible predictive and explainable "neuromakers" for AD. In concert, these explorations further demonstrate the generalisability and reproducibility of the method in identifying brain regions predictive respectively of those non-pairwise-corrected outcomes.

There are several limitations to this study. First, the nature of the imaging and cognitive data suggests that the identified pathways are associative, although our methods selected brain regions whose cortical thickness is significantly predictive of multiple cognitive outcomes (which raises association to out-of-sample prediction). Future studies should examine if some of the identified brain markers and pathways between the high-dimensional neural data and multiple outcomes can be raised to causal relationships. A beginning can perhaps be made by studying individuals with cortical lesions in the identified AD-related areas and examining if they exhibit AD-like behaviour and cognitive symptoms; combining re-PLS and causal inference may be useful in this effort. Second, although re-PLS can perform longitudinal AD prediction, the evaluation of the algorithm was made on sparse time points. This was, in part, due to the nature of the disease (brain structure degenerates progressively at a relatively slow pace, so it is perhaps not necessary to have frequent assessments), and, in part, due to sparse measurements. But making a more semi-continuous assessment of cognitive impairment may help paint a refined, and perhaps more accurate, trajectory of the disease course, assist in monitoring symptom progression and, for patients under treatment, evaluate, more frequently and timely, the treatment efficacy. Future analysis may extend the re-PLS to more dense outcomes. Future analysis may further refine patients into early AD and advanced AD patients and make finer forecasts. In parallel, one can apply the re-PLS on MCI subjects in this study and then follow up and apply the re-PLS to data from the same subjects a few years later to study disease progression. Third, although our method unveils latent maps between brain regions and AD outcomes, the latent maps are not deep (in the sense of deep learning). In a separate analysis, we run a multilayer perceptron (MLP) with a single linear layer, and the results are similar to those from a multivariate linear regression. Additionally, we develop

a multilayer network and incorporate re-PLS into it (we call it the re-Net)[i]. In brief, the re-Net shows better results than MLP and CFNet[48]. In general, however, "deeper" models obtain similar results as the re-PLS and sometimes underperform the re-PLS. One major challenge with the "deeper" models is, while solving the many-to-many disease prediction problem, it is at present oftentimes difficult to make neurobiological sense of the identified brain areas when the weights of the (deep) hidden layers are projected on the brain space. As one of our goals here is to introduce a methodologically sound and neurobiologically meaningful method that delivers both predictive power and can identify brain areas and pathways that may shed light on neuropathology and neurology, we have not run a comprehensive comparison study with deep learning methods and shall reserve it for future works. One potential direction is to consider the re-Net with insights from recent developments in explainable deep learning methods[49] to improve the predictions as well as explanations. Fourth, the definition of AD is only based on symptoms and the clinical diagnosis of patients only assigns them a categorical label of "AD". Although using re-PLS, we can further stratify the patients into different groups based on their continuous (non-categorical) predicted disease scores or using the predicted multivariate cognitive and behaviour scores. One can even build a new, finer continuous AD total score leveraging the multivariate cognitive and behaviour scores (as different subjects have differential degeneration across those multivariate cognitive and behaviour subdomains); an example of a simple score can be a weighted sum of the predicted multivariate scores. This may offer new insights about how to provide a finer prediction of the disease, but we cannot ascertain the validity using current data. Indeed, as a noticeable proportion of AD patients will end up with another diagnosis, such as FTD, LATE, PART, and vascular dementia, it is important to validate if re-PLS can further predict AD patients into these groups. Future work can train re-PLS on subjects with FTD, LATE, PART, and vascular dementia to verify this possibility. Finally, although the goal here is to address a many-to-many problem, the re-PLS can also be applied to predicting single outcomes (as univariate outcomes are, in essence, special cases of multivariate outcomes). We have, however, not studied its performance in predicting single outcomes relative to other methods as it departs from the central aim of the paper.

Taken together, our analyses demonstrated the possibility of identifying and isolating the many-to-many pathways, both cross-sectionally and longitudinally, between high-dimensional multivariate brain data and multiple, non-pairwise-correlated cognitive and behaviour outcomes and using the former to predict the latter in face of confounding variables. We have also provided an open Python package (*re-PLS*) for users to perform their individual analyses via the method. Our readers could use our model and the package on independent datasets to verify if the method can be extended to study other diseases, datasets, or fields.

Methods and Materials

Subject information. The data set used in this article was from the ADNI. ADNI was funded by 20 companies and two foundations through the Foundation for the National Institutes of Health

[i] The network has four MLP blocks with the mean squared error loss (MSE) for regressing Z from X, regression Z from Y, estimating Y residuals from X residuals, and predicting the outcomes. The total loss function is formed by weighted sum of the four MSE losses.

and the National Institute on Aging[50]. Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see www.adni-info.org.

The MRI data release had 1,196 subjects in total. Among them, 45 subjects are in their 50s, 305 subjects are in their 60s, 620 subjects are in their 70s, and 226 subjects are above 80 years old. At the baseline, 184 were diagnosed as AD, 429 with late mild cognitive impairment (LMCI), 234 with early mild cognitive impairment (EMCI), 28 with subjective memory complaint (SMC), and 321 were cognitively normal. The statuses of AD, MCI, or CN are diagnostic outcomes made by clinicians primarily based on clinical criteria (see the ADNI2 Procedures Manual at: https://adni.loni.usc.edu/wp-content/uploads/2024/02/ADNI2_Procedures_Manual_28Feb2024.pdf). During the follow-ups, 12 CNs converted to MCIs, 1 CN converted to AD, 2 SMCs converted to MCIs, and 170 MCIs converted to ADs. Additionally, 9 subjects with either EMCIs or LMCIs reverted to CNs, 26 subjects with SMC reverted to CNs, and 2 AD patients reverted to MCIs. For the cross-sectional study, we used data from all 1,196 subjects. For the longitudinal study, we define CN = cognitively normal, sMCI = stable MCI (a subject assessed as an MCI during the first visit and continued to be diagnosed as an MCI during subsequent visits), pMCI = progressive MCI (a subject was an MCI during early visits and was later diagnosed with AD), and AD = Alzheimer's disease. We excluded 52 subjects from the longitudinal study because they were either labelled as SMC at baseline (28 subjects), converted from CN to MCI (12 subjects) or from CN to AD (1 subject), from AD to MCI (2 subjects), or from EMCI or LMCI to CN (9 subjects); they do not fall into one of the four major groups (CNs, sMCI, pMCI, and AD) and their sub-sample sizes were too small to support meaningful analysis. Thus, the longitudinal study consists of 1,144 subjects, including 308 CNs, 484 sMCI, 170 pMCI and 182 AD.

All participants provided written informed consent. Participants were recruited across North America and agreed to complete a variety of imaging and clinical assessments[50]. All sites are managed by the ADNI Clinical Core and the Data and Publications Committee (DPC) vets all publications using ADNI data[51]. Full details regarding the initiative and the datasets are available at https://adni.loni.usc.edu/methods/documents.

Age	Age group and size	Gender	Total samples
73.39 ± 7.17	50-60: 45 60-70: 305 70-80: 620 80-90: 226	M/F: 662 / 534	2,862

Baseline	Follow-ups
CN: 321 subjects	**CN → CN: 308 subjects**

	CN → MCI: 12 subjects
	CN → AD: 1 subject
MCI: 663 subjects	MCI → CN: 9 subjects
	MCI → MCI: 484 subjects (sMCI)
	MCI → AD: 170 subjects (pMCI)
AD: 184 subjects	AD → MCI: 2 subjects
	AD → AD: 182 subjects

Status	Age at baseline	Gender
CN (308 subjects)	74.50 ± 5.70	M/F: 153 / 155 (49.7% M)
sMCI (484 subjects)	72.57 ± 7.59	M/F: 290 / 194 (59.9% M)
pMCI (170 subjects)	73.27 ± 7.12	M/F: 99 / 71 (58.2% M)
AD (182 subjects)	74.70 ± 7.66	M/F: 96 / 86 (52.7% M)

Test	Abbreviation	Meaning
ADAS-COG	ADAS 11	It is the original ADAS-COG test including 11 items assessing cognitive function. 1. Spoken language ability. 2. Comprehension of spoken language. 3. Recall of test instructions. 4. Word-finding difficulty in spontaneous speech. 5. Following commands. 6. Naming objects and fingers. 7. Constructional praxis. 8. Ideational praxis. 9. Orientation. 10. Word-recall task. 11. Word-recognition task.
	ADAS 13	ADAS 13 (or ADAS-COG 13-item) test includes 11 original ADAS-COG items plus Delayed Word Recall and Number Cancellation.
	ADAS Q4	Q4 task is the Delayed Word Recall task in ADAS13.
The Clinical Dementia Rating (CDR)	CDR SB	The CDR Scale Sum of Boxes (CDRSB) score is obtained by summing the evaluator's rating from six domains: Memory, Orientation, Judgment and Problem Solving, Personal Care, Home and Hobbies, and Community Affairs.
The Mini Mental State Examination (MMSE)	MMSE	The MMSE assessment evaluates orientation to time and place, recall, attention, calculation, and language.
The Rey Auditory Verbal Learning Test (RAVLT)	RAVLT Immediate	The RAVLT is a list of learning tasks that test word recall using multiple trials after a time delay. The RAVLT immediate score measures participant word recall after the first list learning trial.
	RAVLT Learning	The score measures the number of words remembered across all trials.
	RAVLT percent forgetting	The RAVLT percent forgetting score measures the number of words from the original word list missed over all trials in percentage.

Table 2. Demographic and test information for the studied sample.

This paper considers eight disease and behavioural outcomes from the Clinical Dementia Rating (CDR), the Alzheimer's Disease Assessment Scale–Cognitive (ADAS-COG), the Mini Mental

State Examination (MMSE), the Rey Auditory Verbal Learning Test (RAVLT), and the Montreal Cognitive Assessment (MoCA). More specifically, the **CDR** is a score that is derived from the summation of scores from each of the six categories Memory (M), Orientation (O), Judgment and Problem Solving (JPS), Community Affairs (CA), Home and Hobbies (HH) and Personal Care (PC). **ADAS-COG** assesses learning and memory, language production, language comprehension, constructional praxis, ideational praxis, and orientation. It includes tasks/tests such as Word Recall task, Naming task, Word Recognition task, Remembering Test, Word-Finding, and Spoken Language Ability. The MMSE is a brief cognitive screening test used to assess cognitive impairment and cognitive decline. A higher score on the MMSE indicates better cognitive function, while a lower score may suggest the presence of cognitive impairment or dementia. The **RAVLT** assesses abilities like immediate memory, delayed recall, and recognition memory across five immediate learning trials. Further explanations regarding the scores we used in the analysis are in **Table** 2; full explanation of ADNI scores and procedures manual documents at https://adni.loni.usc.edu/methods/documents.

Data acquisition and preprocessing. We used the preprocessed MRI images in ADNI. The functional imaging data (ADNI MRI) were acquired at 2mm isotropic on 1.5T and 3T with different scanner protocols in each phase (ADNI 1, ADNI 2, ADNI GO, and ADNI 3). For this study, we used T1-MRI processed data in ADNI. All image preprocessing was performed using the CAT12 toolbox (http://dbm.neuro.uni-jena.de/cat) with default parameters. In brief, first, the input data were inhomogeneity corrected. The next step was voxel-based morphometry (VBM), the image was spatially registered to a reference brain template and segmented into grey and white matter and CSF. Segmentations were modulated to remain the same grey matter as the original image. Then surface-based morphometry (SBM) was applied to estimate cortical thickness. Regional tissue volumes and cortical thickness for different volumes and surface-based atlas maps were estimated using a high-dimensional spatial registration to map the atlas and the individual brain. Regional tissue volumes were calculated using the Schaefer-Yeo 7 networks atlase[52] with 200-parcel parcellation. Secondary data analysis, including the re-PLS Learning, was conducted using a customised *Python* package which is available at https://github.com/thanhvd18/rePLS.

Notations and data organizations. We begin by defining the notations used throughout this article. The raw imaging data $\boldsymbol{X}$ is a data cubic of $N \times P \times T$, where N, P, and T denote the number of subjects, brain areas, and time points, respectively. Let $\overline{\boldsymbol{X}}_{N\times P}$ be the imaging data averaged over time. For simplicity, without confusion, we refer to $\overline{\boldsymbol{X}}$ as $\boldsymbol{X}$ henceforth. Let $\boldsymbol{Y}$ and $\boldsymbol{Z}$ be the $N \times J$ and $N \times R$ outcome and confounder matrices, respectively. More specifically, each subject has J disease outcomes and R confounders. Denote y_{ij} and z_{ir} as the j^{th} outcome and the r^{th} confounding variable, respectively, for $1 \le j \le J$, $1 \le r \le R$, and $1 \le i \le N$

Residual Partial Least Squares Learning (re-PLS Learning). Here, we outline the basics of re-PLS Learning. First, we obtain residuals $\boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}}$ and $\boldsymbol{\varepsilon}^{\boldsymbol{Y}|\boldsymbol{Z}}$ by regressing X and Y on confounders Z. Specifically,

$$\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}} := \boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}}|_{\boldsymbol{X}=\boldsymbol{x},\ \boldsymbol{Z}=\boldsymbol{z}} = \boldsymbol{x} - \widehat{\boldsymbol{H}}(\boldsymbol{z})\boldsymbol{x}$$

$$\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}} := \boldsymbol{\varepsilon}^{\boldsymbol{Y}|\boldsymbol{Z}}|_{\boldsymbol{Y}=\boldsymbol{y},\ \boldsymbol{Z}=\boldsymbol{z}} = \boldsymbol{y} - \widehat{\boldsymbol{H}}(\boldsymbol{z})\boldsymbol{y} \qquad (1)$$

where $\widehat{\boldsymbol{H}}(\boldsymbol{A}) = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$.

Note that the confounders $\boldsymbol{Z}$ no longer affect the residuals. By construction, $\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}$ and $\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}}$ centre around $\mathbf{0}_P$ and $\mathbf{0}_J$ (column zeros), respectively. They, in practice, may not be strictly zero-cantered; one could zero-centre them without loss of generality.

Algorithm 1: The Residual Partial Least Squares Learning (the re-PLS Learning)

Step 0 (Data organization). Reshape sample data $\boldsymbol{x}$ to be of size $N \times P$, where P is the number of features (brain areas). Arrange sample outcomes and confounding variables $\mathbf{y}$ and $\boldsymbol{z}$ to be of sizes $N \times J$ and $N \times R$, respectively.

Step 1 (Obtaining residuals $\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}$ and $\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}}$). Removing confounding effect by obtaining residuals $\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}} = \boldsymbol{x} - \widehat{\boldsymbol{H}}(\boldsymbol{z})\boldsymbol{x}$ and $\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}} = \boldsymbol{y} - \widehat{\boldsymbol{H}}(\boldsymbol{z})\boldsymbol{y}$.

Step 2 (PLS on residuals $\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}$ and $\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}}$).

(2.0) If $\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}$ and $\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}}$ are not zero centred, zero-centre them.

(2.1) For $s = 1$.

Initiate $\boldsymbol{u}_s^{(0)} = \boldsymbol{y}_j$ for some j. Set:

$$\boldsymbol{p}_s^{(i)} := (\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}})^T\boldsymbol{u}_s^{(i)} / \left\| (\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}})^T\boldsymbol{u}_s^{(i)} \right\|$$
$$\boldsymbol{t}_s^{(i)} := \boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\boldsymbol{p}_s^{(i)}$$
$$\boldsymbol{q}_s^{(i)} := (\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}})^T\boldsymbol{t}_s^{(i)} / \left\| (\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}})^T\boldsymbol{t}_s^{(i)} \right\|$$
$$\boldsymbol{u}_s^{(i+1)} := \boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}}\boldsymbol{q}_s^{(i)}$$

Stop until $\boldsymbol{t}_1$ converges.

(2.2) For $1 < s \leq q$.

Set $\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}} := \boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}} - \boldsymbol{t}_s\boldsymbol{p}_s^T$ and $\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}} := \boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}} - \boldsymbol{u}_s\boldsymbol{q}_s^T$.

Repeat Step 2.1

(2.3) Store the results as projection matrices $\widehat{\boldsymbol{P}}_{train}$ and $\widehat{\boldsymbol{Q}}_{train}$ of dimension $P \times q$ and $J \times q$, respectively. Also store the score matrices $\boldsymbol{T}$ and $\boldsymbol{U}$ both of dimension $N \times q$, where q can be determined by cross-validation. Regress $\boldsymbol{U}$ on $\boldsymbol{T}$: $\boldsymbol{U} = \boldsymbol{T}\boldsymbol{\alpha} + \boldsymbol{E}$, and store the estimated parameters $\widehat{\boldsymbol{\alpha}}_{train}$.

Step 3 (Out-of-sample prediction). For a new subject with feature data $\boldsymbol{X} = \boldsymbol{x}_*$ and confounders $\boldsymbol{Z} = \boldsymbol{z}_*$. The predicted multivariate outcome is: $\widehat{\boldsymbol{y}}_* = \boldsymbol{\varepsilon}^{\boldsymbol{x}_*|\boldsymbol{z}_*}\widehat{\boldsymbol{P}}_{train}\widehat{\boldsymbol{\alpha}}_{train}\widehat{\boldsymbol{Q}}_{train}^T + \boldsymbol{z}_*\boldsymbol{z}_{train}^{-1}\widehat{\boldsymbol{H}}(\boldsymbol{z}_{train})\boldsymbol{y}_{train}$, where $\boldsymbol{\varepsilon}^{\boldsymbol{x}_*|\boldsymbol{z}_*} = \boldsymbol{x}_* - \widehat{\boldsymbol{H}}(\boldsymbol{z}_*)\boldsymbol{x}_*$.

Next, we perform PLS on residuals $\boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}}$ and $\boldsymbol{\varepsilon}^{\boldsymbol{Y}|\boldsymbol{Z}}$, a process which we term *Residual PLS Learning* (re-PLS learning). The two key points of performing the re-PLS Learning are: (a) After removing the confounding effect, the residuals $\boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}}$ and $\boldsymbol{\varepsilon}^{\boldsymbol{Y}|\boldsymbol{Z}}$ are likely to provide better insights about the potential relationship (see **Supplementary Materials**) between the multivariate features $\boldsymbol{X}$ and outcomes $\boldsymbol{Y}$ (compared to the case when confounder effect exists) as the residuals still contain information about $\boldsymbol{X}$ and $\boldsymbol{Y}$ but are independent of $\boldsymbol{Z}$. (b) After removing the effect of $\boldsymbol{Z}$ on $\boldsymbol{X}$, we consider $\boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}}$ as the new, transformed input variable (or transformed features), and the initial confounding effect of $\boldsymbol{Z}$ on $\boldsymbol{Y}$ now becomes a covariate effect (note that $\boldsymbol{Z}$ has an effect on $\boldsymbol{Y}$, $\boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}}$ has an effect on $\boldsymbol{Y}$, but $\boldsymbol{Z}$ does not have any effect on $\boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}}$). This observation is valuable for performing out-of-sample prediction.

In the following, we outline the steps of Residual PLS Learning. Specifically, we decompose $\boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}}$ and $\boldsymbol{\varepsilon}^{\boldsymbol{Y}|\boldsymbol{Z}}$ and link their scores as:

$$\boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}} = \boldsymbol{T}\boldsymbol{P}^{T} \qquad (2)$$

$$\boldsymbol{\varepsilon}^{\boldsymbol{Y}|\boldsymbol{Z}} = \boldsymbol{U}\boldsymbol{Q}^{T} \qquad (3)$$

$$\boldsymbol{U} = \boldsymbol{T}\boldsymbol{\alpha} \qquad (4)$$

such that the decomposition takes into information from both $\boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}}$ and $\boldsymbol{\varepsilon}^{\boldsymbol{Y}|\boldsymbol{Z}}$. Here, $\boldsymbol{P}$ and $\boldsymbol{Q}$ are projection matrices of dimension $P \times q$ and $J \times q$, respectively, for some $q \leq min(P, J)$; $\boldsymbol{T} = \boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}}\boldsymbol{P}$ and $\boldsymbol{U} = \boldsymbol{\varepsilon}^{\boldsymbol{Y}|\boldsymbol{Z}}\boldsymbol{Q}$ are score matrices both of dimension $n \times q$; $\boldsymbol{\alpha}$ is the ordinary least square fit.

We outline the procedure to find the decompositions in **Algorithm** 1 and give mathematical and geometrical interpretations for them in **Supplementary Materials**.

Linking **Eqs.** (2)-(4), we have

$$\boldsymbol{\varepsilon}^{\boldsymbol{Y}|\boldsymbol{Z}} = \boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}}\boldsymbol{P}\boldsymbol{\alpha}\boldsymbol{Q}^{T} \qquad (5)$$

which connects $\boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}}$ and $\boldsymbol{\varepsilon}^{\boldsymbol{Y}|\boldsymbol{Z}}$, with parameters $\boldsymbol{P}$ and $\boldsymbol{Q}$ that maximises the correlations between the residuals (see **Supplementary Materials**).

Predict multivariate outcomes in new subjects. Consider a new subject with feature data $\boldsymbol{X} = \boldsymbol{x}_{*}$ and confounders $\boldsymbol{Z} = \boldsymbol{z}_{*}$. Using Fact 2 below (see Interpretation of the re-PLS Learning) and combining **Eqs.** (1) and (5), the predicted outcome, $\hat{\boldsymbol{y}}_{*}$, without further model fitting, for the new subject is:

$$\hat{\boldsymbol{y}}_{*} = \boldsymbol{\varepsilon}^{\boldsymbol{x}_{*}|\boldsymbol{z}_{*}}\widehat{\boldsymbol{P}}_{train}\widehat{\boldsymbol{\alpha}}_{train}\widehat{\boldsymbol{Q}}_{train}^{T} + \boldsymbol{z}_{*}\boldsymbol{z}_{train}^{-1}\widehat{\boldsymbol{H}}(\boldsymbol{z}_{train})\boldsymbol{y}_{train} \qquad (6)$$

where $\boldsymbol{\varepsilon}^{\boldsymbol{x}_{*}|\boldsymbol{z}_{*}} = \boldsymbol{x}_{*} - \widehat{\boldsymbol{H}}(\boldsymbol{z}_{*})\boldsymbol{x}_{*}$ and the parts with subscript "train" are learnt from the training data.

Note that after removing the confounding effect of $\boldsymbol{z}_*$ on $\boldsymbol{x}_*$, the residuals $\boldsymbol{\varepsilon}^{\boldsymbol{x}_*|\boldsymbol{z}_*}$ are no longer affected by $\boldsymbol{z}_*$. The effect on the outcomes, therefore, is now a covariate effect, namely the second part in **Eq.** (6).

Interpretation of the re-PLS Learning

Fact 1. $(\boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}}, \boldsymbol{\varepsilon}^{\boldsymbol{Y}|\boldsymbol{Z}}) \perp \boldsymbol{Z}$.

Proof. This directly follows from the regression. ■

Fact 1 suggests that since the residuals contain information about features $\boldsymbol{X}$ and outcomes $\boldsymbol{Y}$ but are no longer affected by the confounders $\boldsymbol{Z}$, the pathways identified using the residuals are not biased by $\boldsymbol{Z}$, and, therefore, may provide better insights about the relationship between features $\boldsymbol{X}$ and outcomes $\boldsymbol{Y}$. Fact 1 is the reason we consider performing PLS on residuals $\boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}}$ and $\boldsymbol{\varepsilon}^{\boldsymbol{Y}|\boldsymbol{Z}}$.

Fact 2. After regressing the effect of $\boldsymbol{Z}$ on $\boldsymbol{X}$ out, the confounding effect of $\boldsymbol{Z}$ on $\boldsymbol{Y}$ becomes a covariate effect.

Proof. To see this, note that $\boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}}$ affects $\boldsymbol{Y}$, $\boldsymbol{Z}$ affects $\boldsymbol{Y}$ but $\boldsymbol{Z}$ is independent from $\boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}}$. ■

Fact 2 is the basis for making out-of-sample predictions using $\boldsymbol{\varepsilon}^{\boldsymbol{X}|\boldsymbol{Z}}$ as exposures and $\boldsymbol{Z}$ as covariates in re-PLS learning.

Fact 3. The projections ($\boldsymbol{P}$ and $\mathbf{Q}$ in **Eqs.** (2) and (3)) maximise both residual variances and the correlation between the projected residuals. Said differently, they take into consideration information of $\boldsymbol{X}$, of $\boldsymbol{Y}$, and between $\boldsymbol{X}$ and $\boldsymbol{Y}$ that is not affected by confounders $\boldsymbol{Z}$.

Proof. From Step 2.1 of **Algorithm** 1, we see $\boldsymbol{p} = (\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}})^T\boldsymbol{u}/\left\|(\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}})^T\boldsymbol{u}\right\| = (\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}})^T\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}}\boldsymbol{q}/\left\|(\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}})^T\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}}\boldsymbol{q}\right\| = (\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}})^T\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}}(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}})^T\boldsymbol{t}/\left\|(\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}})^T\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}}(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}})^T\boldsymbol{t}\right\| = (\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}})^T\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}}(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}})^T\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\boldsymbol{p}/\left\|(\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}})^T\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}}(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}})^T\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\boldsymbol{p}\right\| = 1/\lambda \times ((\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}})^T\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}})^T(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}})^T\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\boldsymbol{p}$ for some λ; *i.e.*, $\boldsymbol{p}$ is the eigenvector of covariance matrix $(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}})^T\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}$.

Lemma 1. Given a symmetric matrix $\boldsymbol{M}$, the solution $\tilde{\boldsymbol{x}}$ to:

$$\tilde{\boldsymbol{x}} = \underset{\boldsymbol{x}:\|\boldsymbol{x}\|=1}{\operatorname{argmax}}\{\boldsymbol{x}^T\boldsymbol{M}\boldsymbol{x}\}$$

is the one such that $\tilde{\boldsymbol{x}}^T\boldsymbol{M}\tilde{\boldsymbol{x}} = \max_i\{\lambda_i\}$, where $\tilde{\boldsymbol{x}}$ is the eigenvector of $\boldsymbol{M}$ corresponding to $\max_i\{\lambda_i\}$.

Proof of Lemma 1. Let $\boldsymbol{M} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^T$ be the spectral decomposition of $\boldsymbol{M}$. Let $\tilde{\boldsymbol{x}} = \boldsymbol{\Phi}^T\boldsymbol{x}$, for any $\boldsymbol{x}$ with $\|\boldsymbol{x}\| = 1$. Then $\boldsymbol{x}^T\boldsymbol{M}\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^T\boldsymbol{x} = \tilde{\boldsymbol{x}}^T\boldsymbol{\Lambda}\tilde{\boldsymbol{x}} = \sum_i \lambda_i\tilde{x}_i^2$.

Since $\min_i\{\lambda_i\}\sum_i \tilde{x}_i^2 \le \sum_i \lambda_i\tilde{x}_i^2 \le \max_i\{\lambda_i\}\sum_i \tilde{x}_i^2 = \max_i\{\lambda_i\}\tilde{\boldsymbol{x}}^T\tilde{\boldsymbol{x}} = \max_i\{\lambda_i\}(\boldsymbol{\Phi}^T\boldsymbol{x})^T\boldsymbol{\Phi}^T\boldsymbol{x} = \max_i\{\lambda_i\}\boldsymbol{x}^T\boldsymbol{x} = \max_i\{\lambda_i\}$, then $\boldsymbol{x}^T\boldsymbol{M}\boldsymbol{x} \le \max_i\{\lambda_i\}$. This proves the first part. Now suppose $\tilde{\boldsymbol{x}}$ is the i^{th} eigenvector of $\boldsymbol{M}$ or $\boldsymbol{\phi}_i$, then if $\max_i\{\lambda_i\} = \boldsymbol{x}^T\boldsymbol{M}\boldsymbol{x} = \boldsymbol{\phi}_i{}^T\boldsymbol{\Phi}\mathrm{diag}\{\lambda_i\}\boldsymbol{\Phi}^T\boldsymbol{\phi}_i$, it must be that $\boldsymbol{\phi}_i$ corresponding to $\max_i\{\lambda_i\}$. ■

By Lemma 1, $\boldsymbol{p}$ is the solution to:

$$\underset{\boldsymbol{p}:\|\boldsymbol{p}\|=1}{\operatorname{argmax}}\{\boldsymbol{p}^T(\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}})^T\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}}(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}})^T\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\boldsymbol{p}\}$$
$$= \underset{\boldsymbol{p}:\|\boldsymbol{p}\|=1}{\operatorname{argmax}}\left\{\left((\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}})^T\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\boldsymbol{p}\right)^T\left((\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}})^T\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\boldsymbol{p}\right)\right\} = \underset{\boldsymbol{p}:\|\boldsymbol{p}\|=\mathbf{1}}{\operatorname{argmax}}\left\{cov\left(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}},\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\boldsymbol{p}\right)^T cov\left(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}},\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\boldsymbol{p}\right)\right\}$$
$$= \underset{\boldsymbol{p}:\|\boldsymbol{p}\|=1}{\operatorname{argmax}}\left\{\begin{array}{l}\left(\sqrt{var(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}})}corr(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}},\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\boldsymbol{p})\sqrt{var(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}}\boldsymbol{p})}\right)^T \\ \times\sqrt{var(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}})}corr(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}},\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\boldsymbol{p})\sqrt{var(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}}\boldsymbol{p})}\end{array}\right\}$$
$$= \underset{\boldsymbol{p}:\|\boldsymbol{p}\|=1}{\operatorname{argmax}}\left\{\begin{array}{l}var(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}})\left(corr(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}},\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\boldsymbol{p})\right)^T \\ corr(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}},\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\boldsymbol{p})var\left(\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\boldsymbol{p}\right)\end{array}\right\},$$

which maximises the residual variance of the outcomes $var\left(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}}\right)$, the variance of the projected features residuals $var\left(\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\boldsymbol{p}\right)$, and the correlation between the two $corr(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}},\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\boldsymbol{p})$. A similar argument would show that $\boldsymbol{q}$ maximises the residual variance of the features $var\left(\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\right)$, the variance of the projected outcomes residuals $var\left(\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}}\boldsymbol{q}\right)$, and the correlation between the two $corr(\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}},\boldsymbol{\varepsilon}^{\boldsymbol{y}|\boldsymbol{z}}\boldsymbol{q})$. ■

Fact 3 provides an interpretation of the weights in the re-PLS learning. It also highlights the difference between the PCR (principal component regression), another attractive method that could simultaneously predict multivariate outcomes, and re-PLS learning. The former only aims at maximizing the variance of the projected feature residuals $var\left(\boldsymbol{\varepsilon}^{\boldsymbol{x}|\boldsymbol{z}}\boldsymbol{p}\right)$, whereas the latter, in addition, takes information (more precisely the correlation) between features (more precisely projected residuals of features) and outcomes (more precisely the residual outcomes) into account. We think this is the reason why re-PLS yields better out-of-sample prediction.

Code and package availability.

The *re-PLS* python package is available at https://github.com/thanhvd18/rePLS.

Data availability.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Author contributions.

O.Y.C. conceptualised the study. **D.T.V.** and **O.Y.C.** developed the methods. **D.T.V.** wrote the codes, performed data processing and analysis, and wrote and maintained the *re-PLS* package. **N.L.T.** and **O.Y.C.** provided support and guidance for **D.T.V. H.C.**, **G.A.**, **P.R.**, and **G.N.** provided neurobiological support. **C.S.D.**, **J.S.B.**, **H.P.**, **V.D.N.**, **H.S.**, and **H.Z.** provided statistical and machine learning support. **X.H.**, **B.Z.**, **W.H.**, **X.W.**, and **M.M.**, provided insights into disease prediction. **Y.M.** and **G.P.** provided clinical and medical insights. **O.Y.C.** and **D.T.V.** wrote the manuscript, with comments from all other authors.

Acknowledgement.

D.T.V. was funded by Vingroup JSC and the Master and PhD Scholarship Program from the Institute of Big Data of Vingroup Innovation Foundation (VINIF) [VINIF.2021.ThS.15].

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Conflicts of interest.

Non-declared.

References

1. Breijyeh, Z. & Karaman, R. Comprehensive review on Alzheimer's disease: Causes and treatment. *Molecules* **25**, 5789 (2020).
2. World Health Organization. Dementia. (2021). Available at: https://www.who.int/en/news-room/fact-sheets/detail/dementia.
3. National Institute on Aging. Alzheimer's disease fact sheet. Available at: https://www.nia.nih.gov/health/alzheimers-disease-fact-sheet.
4. Querfurth, H. W. & LaFerla, F. M. Alzheimer's disease. *N. Engl. J. Med.* **362**, 329–344 (2010).
5. Kahle-Wrobleski, K., Ye, W., Henley, D., Hake, A.M., Siemers, E., Chen, Y.F. and Liu-Seifert, H. Assessing quality of life in Alzheimer's disease: Implications for clinical trials. *Alzheimer's Dement. Diagnosis, Assess. Dis. Monit.* **6**, 82–90 (2017).
6. Lawton, M. P. Quality of life in Alzheimer disease. *Alzheimer Dis. Assoc. Disord.* **29**, 138–150 (1994).
7. Logsdon, R. G., Gibbons, L. E., McCurry, S. M. & Teri, L. Quality of life in Alzheimer's disease: Patient and caregiver reports. *J. Ment. Health Aging* **5**, 21–32 (1999).
8. Teri, L. & Logsdon, R. G. Identifying pleasant activities for Alzheimer's disease patients: The pleasant events schedule-AD. *Gerontologist* **31**, 124–127 (1991).
9. Tappen, R. M., Williams, C., Fishman, S. & Touhy, T. Persistence of self in advanced Alzheimer's disease. *J. Nurs. Scholarsh.* **31**, 121–125 (1999).
10. El Haj, M., Jardri, R., Larøi, F. & Antoine, P. Hallucinations, loneliness, and social isolation in Alzheimer's disease. *Cogn. Neuropsychiatry* **21**, 1–13 (2016).
11. Lyketsos, C. G. & Olin, J. Depression in Alzheimer's disease: Overview and treatment. *Biol. Psychiatry* **52**, 243–252 (2002).
12. Logsdon, R. G. & Teri, L. Depression in Alzheimer's disease patients: Caregivers as surrogate reporters. *J. Am. Geriatr. Soc.* **43**, 150–155 (1995).
13. Nandi, A., Counts, N., Chen, S., Seligman, B., Tortorice, D., Vigo, D. and Bloom, D.E. Global and regional projections of the economic burden of Alzheimer's disease and related dementias from 2019 to 2050: A value of statistical life approach. *eClinicalMedicine* **51**, 101580 (2022).
14. Querbes, O., Aubry, F., Pariente, J., Lotterie, J.A., Démonet, J.F., Duret, V., Puel, M., Berry, I., Fort, J.C., Celsis, P. and Alzheimer's Disease Neuroimaging Initiative. Early diagnosis of Alzheimer's disease using cortical thickness: Impact of cognitive reserve. *Brain* **132**, 2036–2047 (2009).
15. Dickerson, B.C., Bakkour, A., Salat, D.H., Feczko, E., Pacheco, J., Greve, D.N., Grodstein, F., Wright, C.I., Blacker, D., Rosas, H.D. and Sperling, R.A. The cortical signature of Alzheimer's disease: Regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cereb. Cortex* **19**, 497–510 (2009).
16. Singh, V., Chertkow, H., Lerch, J.P., Evans, A.C., Dorr, A.E. and Kabani, N.J. Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer's disease. *Brain* **129**, 2885–2893 (2006).
17. Du, A.T., Schuff, N., Kramer, J.H., Rosen, H.J., Gorno-Tempini, M.L., Rankin, K., Miller, B.L. and Weiner, M.W. Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia. *Brain* **130**, 1159–1166 (2007).
18. Lerch, J.P., Pruessner, J.C., Zijdenbos, A., Hampel, H., Teipel, S.J. and Evans, A.C. Focal

decline of cortical thickness in Alzheimer's disease identified by computational neuroanatomy. *Cereb. Cortex* **15**, 995–1001 (2005).

19. Bakkour, A., Morris, J. C. & Dickerson, B. C. The cortical signature of prodromal AD: Regional thinning predicts mild AD dementia. *Neurology* **72**, 1048–1055 (2009).
20. Morris, J. C. Clinical dementia rating: A reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *Int. Psychogeriatrics* **9**, 173–176 (1997).
21. Cano, S.J., Posner, H.B., Moline, M.L., Hurt, S.W., Swartz, J., Hsu, T. and Hobart, J.C. The ADAS-cog in Alzheimer's disease clinical trials: Psychometric evaluation of the sum and its parts. *J. Neurol. Neurosurg. Psychiatry* **81**, 1363–1368 (2010).
22. Preziosa, P., Rocca, M.A., Pagani, E., Stromillo, M.L., Enzinger, C., Gallo, A., Hulst, H.E., Atzori, M., Pareto, D., Riccitelli, G.C. and Copetti, M. Structural MRI correlates of cognitive impairment in patients with multiple sclerosis: A multicenter study. *Hum. Brain Mapp.* **37**, 1627–1644 (2016).
23. Tian, J. & Pearl, J. A general identification condition for causal effects. in *Proceedings of the National Conference on Artificial Intelligence.* AAAI Press/The MIT Press: Menlo Park, CA, 567–573 (2002). Available at: https://www.aaai.org/Papers/AAAI/2002/AAAI02-085.pdf
24. Skelly, A., Dettori, J. & Brodt, E. Assessing bias: The importance of considering confounding. *Evid. Based. Spine. Care. J.* **3**, 9–12 (2012).
25. Chén, O. Y. The roles of statistics in human neuroscience. *Brain Sci.* **9**, 194 (2019).
26. Payan, A. & Montana, G. Predicting Alzheimer's disease: A neuroimaging study with 3D convolutional neural networks. *arXiv* 1502.02506 (2015).
27. Velayudhan, L., Proitsi, P., Westman, E., Muehlboeck, J.S., Mecocci, P., Vellas, B., Tsolaki, M., Kłoszewska, I., Soininen, H., Spenger, C. and Hodges, A. Entorhinal cortex thickness predicts cognitive decline in Alzheimer's disease. *J. Alzheimer's Dis.* **33**, 755–766 (2013).
28. Saad, Y. *Iterative methods for sparse linear systems*. (SIAM, 200AD).
29. Zhang, K., Zuo, W., Chen, Y., Meng, D. & Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.* **26**, 3142–3155 (2017).
30. Garthwaite, P. H. An interpretation of partial least squares. *J. Am. Stat. Assoc.* **89**, 122–127 (1994).
31. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **58**, 109–130 (2001).
32. Trygg, J. & Wold, S. Orthogonal projections to latent structures (O-PLS). *J. Chemom.* **16**, 119–128 (2002).
33. Chén, O.Y. *et al*. The statistical analysis of the varying brain. *IEEE Statistical Signal Processing*, 700–704 (2023).
34. Mummery, C.J., Patterson, K., Price, C.J., Ashburner, J., Frackowiak, R.S. and Hodges, J.R. A voxel-based morphometry study of semantic dementia: Relationship between temporal lobe atrophy and semantic memory. *Ann. Neurol.* **47**, 36–45 (2000).
35. Strotzer, M. One century of brain mapping using Brodmann areas. *Clin. Neuroradiol.* **19**, 179–186 (2009).
36. Beal, M.F., Mazurek, M.F., Tran, V.T., Chattha, G., Bird, E.D. and Martin, J.B. Reduced numbers of somatostatin receptors in the cerebral cortex in Alzheimer's disease. *Science* **229**, 289–291 (1985).

37. Scheff, S. W. & Price, D. A. Synapse loss in the temporal lobe in Alzheimer's disease. *Ann Neurol.* **33**, 190–199 (1993).
38. Johnson, K.A., Jones, K., Holman, B.L., Becker, J.A., Spiers, P.A., Satlin, A. and Albert, M.S. Preclinical prediction of Alzheimer's disease using SPECT. *Neurology* **50**, 1563–1571 (1998).
39. Greicius, M. D., Srivastava, G., Reiss, A. L. & Menon, V. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *Proc. Natl. Acad. Sci.* **101**, 4637–4642 (2004).
40. Buckner, R.L., Snyder, A.Z., Shannon, B.J., LaRossa, G., Sachs, R., Fotenos, A.F., Sheline, Y.I., Klunk, W.E., Mathis, C.A., Morris, J.C. and Mintun, M.A. Molecular, structural, and functional characterization of Alzheimer's disease: Evidence for a relationship between default activity, amyloid, and memory. *J. Neurosci.* **25**, 7709–7717 (2005).
41. Meadows, J. C. Disturbed perception of colours associated with localized cerebral lesions. *Brain* **97**, 615–32 (1974).
42. Minoshima, S., Giordani, B., Berent, S., Frey, K.A., Foster, N.L. and Kuhl, D.E. Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. *Ann. Neurol.* **42**, 85–94 (1997).
43. Shinno, H., Inagaki, T., Miyaoka, T., Okazaki, S., Kawamukai, T., Utani, E., Inami, Y. and Horiguchi, J. A decrease in N-acetylaspartate and an increase in myoinositol in the anterior cingulate gyrus are associated with behavioral and psychological symptoms in Alzheimer's disease. *J. Neurol. Sci.* **206**, 132–138 (2007).
44. Liu, X., Chen, W., Hou, H., Chen, X., Zhang, J., Liu, J., Guo, Z. and Bai, G. Decreased functional connectivity between the dorsal anterior cingulate cortex and lingual gyrus in Alzheimer's disease patients with depression. *Behav. Brain Res.* **326**, 132–138 (2017).
45. Gilbert, S.J., Spengler, S., Simons, J.S., Steele, J.D., Lawrie, S.M., Frith, C.D. and Burgess, P.W. Functional specialization within rostral prefrontal cortex (Area 10): A meta-analysis. *J. Cogn. Neurosci.* **18**, 932–948 (2006).
46. Albers, M.W., Gilmore, G.C., Kaye, J., Murphy, C., Wingfield, A., Bennett, D.A., Boxer, A.L., Buchman, A.S., Cruickshanks, K.J., Devanand, D.P. and Duffy, C.J. At the interface of sensory and motor dysfunctions and Alzheimer's disease. *Alzheimer's Dement.* **11**, 70–98 (2015).
47. Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Liu, E. and Morris, J.C. The Alzheimer's disease neuroimaging initiative: A review of papers published since its inception. *Alzheimer's Dement.* **9**, e111–e194 (2013).
48. Zhao, Q., Adeli, E. & Pohl, K. M. Training confounder-free deep learning models for medical applications. *Nat. Commun.* **11**, 1–9 (2020).
49. Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R. and Chatila, R. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **58**, 82–115 (2020).
50. ADNI. https://adni.loni.usc.edu/about (2017).
51. Weiner, M.W., Aisen, P.S., Jack Jr, C.R., Jagust, W.J., Trojanowski, J.Q., Shaw, L., Saykin, A.J., Morris, J.C., Cairns, N., Beckett, L.A. and Toga, A. The Alzheimer's disease neuroimaging initiative: Progress report and future plans. *Alzheimer's Dement.* **6**, 202–211 (2010).

52. Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.N., Holmes, A.J., Eickhoff, S.B. and Yeo, B.T. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
53. Wager, Tor D., and Edward E. Smith. Neuroimaging studies of working memory. *Cognitive, Affective, & Behavioral Neuroscience* **3**, 255-274 (2003).
54. Leung, H-C., John C. Gore, and Patricia S. Goldman-Rakic. Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. J. Cogn. Neurosci. **14**, 659-671 (2002).
55. American Association for Research into Nervous and Mental Diseases, Courtney, S.M., Petit, L., Haxby, J.V. and Ungerleider, L.G. The role of prefrontal cortex in working memory: examining the contents of consciousness. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **353**, 1819-1828 (1998.56).
56. Frey, S. and Petrides, M. Orbitofrontal cortex and memory formation. *Neuron* **36**, 171-176 (2002).
57. Rolls, E.T. and Grabenhorst, F. The orbitofrontal cortex and beyond: From affect to decision-making. *Progress in neurobiology* **86**, 216-244 (2008).
58. Johnson, K.A., Schultz, A., Betensky, R.A., Becker, J.A., Sepulcre, J., Rentz, D., Mormino, E., Chhatwal, J., Amariglio, R., Papp, K. and Marshall, G. Tau positron emission tomographic imaging in aging and early Alzheimer disease. *Ann. Neurol* **79**, 110-119 (2016).
59. Jurado, M.B. and Rosselli, M. The elusive nature of executive functions: A review of our current understanding. *Neuropsychol. Rev.* **17**, 213-233 (2007).
60. Rushworth, M.F.S., Walton, M.E., Kennerley, S.W. and Bannerman, D.M. Action sets and decisions in the medial frontal cortex. *Trends Cogn. Sci.* **8**, 410-417 (2004).
61. Koechlin, E., Corrado, G., Pietrini, P. and Grafman, J. Dissociating the role of the medial and lateral anterior prefrontal cortex in human planning. *PNAS* **97**, 7651-7656 (2000).
62. Ardila, A., Bernal, B. and Rosselli, M. The elusive role of the left temporal pole (BA38) in language: A preliminary meta-analytic connectivity study. *International Journal of Brain Science 2014,* 2014.
63. Armstrong, N.M., An, Y., Shin, J.J., Williams, O.A., Doshi, J., Erus, G., Davatzikos, C., Ferrucci, L., Beason-Held, L.L. and Resnick, S.M. Associations between cognitive and brain volume changes in cognitively normal older adults. *Neuroimage* **223**, 117289 (2020)
64. Ben-Zvi, S., Soroker, N. and Levy, D.A. Parietal lesion effects on cued recall following pair associate learning. *Neuropsychologia* **73**, 176-194 (2015).
65. Ye, Z., Zhang, G., Li, S., Zhang, Y., Xiao, W., Zhou, X. and Muente, T.F. Age differences in the fronto-striato-parietal network underlying serial ordering. *Neurobiol. Aging* **87**, 115-124 (2020).
66. Gough, P.M., Nobre, A.C. and Devlin, J.T., 2005. Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *J. Neurosci*, **25**, pp.8010-8016.
67. Heim, S., Eickhoff, S.B. and Amunts, K., 2009. Different roles of cytoarchitectonic BA 44 and BA 45 in phonological and semantic verbal fluency as revealed by dynamic causal modelling. *Neuroimage*, **48**, pp.616-624.
68. Friston, K.J. and Büchel, C., 2000. Attentional modulation of effective connectivity from V2 to V5/MT in humans. *Proceedings of the National Academy of Sciences*, **97**, pp.7591-7596.

69. Bussey, T.J. and Saksida, L.M., 2007. Memory, perception, and the ventral visual-perirhinal-hippocampal stream: Thinking outside of the boxes. *Hippocampus*, **17**, pp.898-908.
70. Heim, S., Eickhoff, S.B. and Amunts, K., 2009. Different roles of cytoarchitectonic BA 44 and BA 45 in phonological and semantic verbal fluency as revealed by dynamic causal modelling. *Neuroimage*, **48**, pp.616-624.
71. Le, A., Vesia, M., Yan, X., Crawford, J.D. and Niemeier, M., 2017. Parietal area BA7 integrates motor programs for reaching, grasping, and bimanual coordination. *J. Neurophysiol,* **117***,* pp.624-636*.*
72. Volz, K.G., Schubotz, R.I. and von Cramon, D.Y., 2005. Variants of uncertainty in decision-making and their neural correlates. *Brain Res. Bull.*, **67**, pp.403-412.