Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin
Ollscoil Átha Cliath | The University of Dublin

Development and validation of objective measures of brain maintenance and cognitive reserve

Rory Boyle

School of Psychology and Trinity College Institute of Neuroscience
Trinity College Dublin, University of Dublin

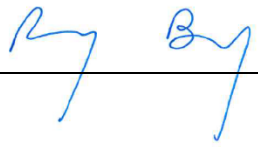
A thesis submitted to
Trinity College Dublin, University of Dublin
For the degree of
Doctor of Philosophy

2021

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Signed: Rory Boyle 

Date: 31/03/2021

Summary

Age-related cognitive decline is an increasingly important societal issue, given projected increases in the proportion of older adults in the coming decades. Early identification of cognitive decline will enable earlier interventions which have a greater likelihood of slowing decline, maintaining quality of life, and reducing the burden on caregivers and society. The application of machine learning to neuroimaging data is a promising strategy to detect age-related cognitive decline. However, there has been little emphasis on the development of measures of two related, yet separable, constructs – brain maintenance and cognitive reserve – both of which support cognitive function as we age. Accurate measurement of these two constructs may improve our ability to detect age-related cognitive decline.

Chapter 2 applied a machine learning method to structural MRI data in order to predict chronological age. Brain-predicted age difference scores were then created by subtracting chronological ages from the predicted age. A penalised regression with cross-validation was applied to generate the model using open-access structural MRI data. This model was then applied to structural MRI data in three independent datasets. Across these independent datasets, brain-predicted age differences were negatively correlated with measures of general cognitive status; semantic verbal fluency; executive function; and executive function (without processing speed). These results provided firm evidence of a robust relationship between increased brain-predicted age differences and reduced cognitive function in specific domains. As such, the findings established the validity of brain-predicted age difference scores as an operational measure of brain maintenance.

Chapter 3 applied a data-driven framework in order to establish the validity of different socio-behavioural variables as 'proxy' measures of cognitive reserve in a cross-sectional study of cognitively healthy older adults. To demonstrate face validity as a measure of cognitive reserve, candidate neuroimaging measures must be shown to correlate with a socio-behavioural proxy. Furthermore, socio-behavioural proxies are the most commonly used measures of cognitive reserve. However, there is little empirical evidence demonstrating the validity, and guiding the choice, of proxy variables as measures of cognitive reserve. The validity of five common proxies and all possible combinations of their composites were assessed across two community-dwelling older adult cohorts. Verbal intelligence was found to be the most robust socio-behavioural proxy measure of cognitive reserve.

Chapter 4 applied a novel machine learning method, connectome-based predictive modelling, to functional connectivity data in order to develop and validate an objective measure of cognitive reserve. This measure was developed using task-based functional connectivity data from one dataset and then applied to resting-state fMRI data from an independent dataset. Face validity of the measure was assessed by establishing its association with the most robust socio-behavioural proxy measure of CR, verbal intelligence, as identified in Chapter 3. The protective effects of the measure was assessed by establishing its association with cognitive function, independent of brain structure. The measure accurately predicted CR in the training set and was validated as a measure of CR as it demonstrated face validity and protective effects on cognition. However, the measure was not validated in the independent dataset when generated using resting-state data.

Overall, the findings demonstrate the value of machine learning for the development of robust and objective measures of brain maintenance and cognitive reserve using neuroimaging data. *Chapter 2* established that brain-predicted age difference scores can serve as a valid measure of brain maintenance across cohorts, and may prove to be useful biomarkers of cognitive ageing. *Chapter 3* identified verbal intelligence as the most robust socio-behavioural proxy of cognitive reserve and therefore recommended that researchers should use this variable when assessing the face validity of potential cognitive reserve neuroimaging measures. *Chapter 4* developed a functional neuroimaging measure of cognitive reserve based on task-based fMRI data but further research is needed to validate this measure using resting-state data. Further innovations to the models outlined in Chapters 2 and 4 may provide important insights into the development and enhancement of brain maintenance and cognitive reserve and will further improve our understanding of these constructs. These validated measures of brain maintenance and cognitive reserve could be used to improve the early identification of cognitive decline and to directly assess the efficacy of preventative interventions targeted at the enhancement of brain maintenance and/or cognitive reserve.

List of Publications and Presentations

This thesis incorporates material already published or currently under review in the following manuscripts:

Boyle, R., Knight, S. P., De Looze, C., Carey, D., Scarlett, S., Stern, Y., ... , & Whelan, R. (2021). Verbal Intelligence, Not Level of Education, Robustly Assesses Cognitive Reserve. *Alzheimer's Research & Therapy (under review).* 10.21203/rs.3.rs-216364/v1

Boyle, R., Jollans, L., Rueda-Delgado, L.M., Rizzo, R., Yener, G.G., McMorrow, J.P., … , & Whelan, R. (2021). Brain-predicted age difference score is related to specific cognitive functions: a multi-site replication analysis. *Brain Imaging and Behavior* 15, 327–345 https://doi.org/10.1007/s11682-020-00260-3

The following are presentations arising from this work:

Boyle, R. Improving the moderation and independent effect criteria of cognitive reserve. Oral presentation at *2nd Workshop on Research Definitions for Reserve and Resilience in Cognitive Aging and Dementia 2020*.

Boyle, R., Scarlett, S., Knight, S.P., De Looze, C., Stern, Y., Robertson, I.H., … , & Whelan, R. Validation of composite proxy measures of cognitive reserve. Oral presentation at *Alzheimer's Association International Conference 2020.* https://doi.org/10.1002/alz.041824

Boyle, R., Scarlett, S., Knight, S.P., Stern, Y., Robertson, I.H., Kenny, R.A., & Whelan, R. A data-driven framework for validating composite measures of cognitive reserve. Poster presentation at *ResDem 2019: 2nd International Conference on Cognitive Reserve in Dementia and Other Disorders.* https://bit.ly/3ec2oz3

Boyle, R., Jollans, L., Rizzo, Rueda-Delgado, L.M., Fide, E., Emek-Savaş, D.D., … , & Whelan, R. Brain-predicted age difference as a neuromarker of cognitive function. Poster presentation at *Organization for Human Brain Mapping 2019.* https://bit.ly/2XmIdIN

Boyle, R., Jollans, L., Rizzo, Rueda-Delgado, L.M., Fide, E., Emek-Savaş, D.D., … , & Whelan, R. Brain-predicted age difference as an objective measure of cognitive function. Poster presentation at *British Neuroscience Association Festival of Neuroscience 2019.* https://bit.ly/2XsSMdc

Acknowledgements

I would like to sincerely thank my supervisor Prof. Robert Whelan for calmly guiding me through the research process from start to finish. I greatly appreciate the time you gave me and patience you showed with me throughout. Thank you very much for all of the advice, encouragement and support.

I would also like to thank everyone in the Whelan Lab, past and present. Specifically, I would like to thank Dr. Eimear McGlinchey, Dr. Lee Jollans, Michael Connaughton, and Dr. Rossella Rizzo for their help with the studies presented in this thesis. Thank you to Dr. Francesca Farina, Gabi Pragulbickaitė, Hannah Rice, Laura Brosnan, and Laura Rai for breezing through the older adult project. Dr. Zhipeng Cao, thank you for helping me with, and talking me through, different aspects of coding, data analysis, and machine learning. Dr. Kathy Ruddy, thank you very much for the help and training with SPM and neuroimaging preprocessing.

I would like to thank the team in TILDA for their help, support, and advice throughout this project, especially Prof. Rose Anne Kenny, Dr. Céline De Looze, Dr. Dan Carey, and Dr. Silvin Knight. I would like to thank Prof. Ian Robertson for the very valuable advice, guidance, and feedback throughout this project. I would like to thank Prof. Yaakov Stern for kindly sharing his data and providing insights and feedback at different stages throughout this project. Thank you also to Jayant Sakhardande and Dr. Sharon Sanz Simon, in Prof. Stern's group, for organising the data and helping me out with various different questions about the data. I would also like to thank Prof. Derya Durusu Emek-Savaş and Prof. Görsev Yener for kindly sharing their data. I would like to acknowledge all the participants who gave up their time to contribute this data.

I am very grateful to the Irish Research Council and Altoida AG for funding this PhD. In particular, thank you very much to Dr. Ioannis Tarnanas, of Altoida AG, for the support.

Mam, Dad, agus Geraldine, go raibh maith agaibh.

List of Figures

List of Tables

List of Abbreviations

AD	Alzheimer's Disease
AMNART	America National Adult Reading Test
APOE	Apolipoprotein E
Bagging	Bootstrap Aggregating
BLOT	Benton Judgement of Line Orientation Test
BM	Brain Maintenance
BR	Brain Reserve
BrainPAD	Brain Predicted Age Difference
CTT	Colour Trails Task
CPM	Connectome-based Predictive Modelling
CR	Cognitive Reserve
CRIq	Cognitive Reserve Index Questionnaire
CR/RANN	Cognitive Reserve/Reference Ability Neural Network study dataset
CRS	Cognitive Reserve Scale
CSF	Cerebrospinal Fluid
CV	Cross-validation
DARTEL	Diffeomorphic Anatomical Registration using Exponentiated Lie algebra
DEU	Dokuz Eylül University dataset
DMN	Default Mode Network
DOT	Dictionary of Occupational Titles
DRS	Mattis Dementia Rating Scale-2
DTI	Diffusion Tensor Imaging
EEG	Electroencephalography
EPI	Echo-Planar Imaging
FDG-PET	Fluorodeoxyglucose-Positron Emission Tomography
fMRI	Functional Magnetic Resonance Imaging
FOV	Field of View
FPN	Frontoparietal Network
FWD	Framewise Displacement
GM	Grey Matter
LASSO	Least Absolute Shrinkage and Selection Operator
LEQ	Lifetime Experience Questionnaire
LOOCV	Leave-One-Out-Cross-Validation
MAE	Mean Absolute Error
MCI	Mild Cognitive Impairment

MEG	Magnetoencephalography
MFN	Medial Frontal Network
MMSE	Mini-Mental State Examination
MNI	Montréal Neuroimaging Institute
MPRAGE	Magnetization-Prepared Rapid Gradient Echo
MRI	Magnetic Resonance Imaging
NART	National Adult Reading Test
OVMPT	Öktem Verbal Memory Processes Test
PCA	Principal Component Analysis
PET	Positron Emission Tomography
PVT	Psychomotor Vigilance Test
RAVLT	Rey Auditory Verbal Learning Test
ROI	Regions of Interest
SART	Sustained Attention to Response Task
SD	Standard Deviation
SPM	Statistical Parametric Mapping
SRT	Selective Reminding Test
TE	Echo Time
TILDA	The Irish Longitudinal Study on Ageing dataset
TIV	Total Intracranial Volume
TMT	Trail Making Test
TR	Repetition Time
WAIS-III	Wechsler Adult Intelligence Scale – Third Edition
WCST	Wisconsin Card Sorting Task
WM	White Matter
WMS-R	Wechsler Memory Scale – Revised Edition

Table of Contents

1 Chapter 1: General Introduction

1.1 Cognitive decline and the variability in cognitive decline

Dementia is a clinical syndrome characterised by a progressive decline in cognitive functioning, behavioural changes, and functional impairment that ultimately causes a loss of independence (Goyal et al., 2018; Van Der Flier & Scheltens, 2005). Dementia is associated with reduced quality of life (Goyal et al., 2018) and neuropsychiatric symptoms, such as anxiety (Seignourel et al., 2008) and depression (Enache et al., 2011). Dementia also places a huge burden on caregivers, who tend to experience poorer physical and mental health (Richardson et al., 2013; Sallim et al., 2015), and on society as a whole (Jutkowitz et al., 2017).

Globally, 43.8 million individuals were estimated to be living with dementia in 2016, an increase of 117% from the estimate of 20.2 million in 1990 (Nichols et al., 2019). This rate of increase is likely to continue as age is the most important risk factor for dementia (Van Der Flier & Scheltens, 2005) and the proportion of the population over the age of 60 is projected to rapidly increase in the coming decades (Bloom et al., 2015). The projected growth in the number of cases and associated increasing burden on caregivers and society means there is an urgent need to cure and/or prevent dementia. It also emphasises the need to slow or reduce the cognitive decline associated with dementia in order to maintain the quality of life and independence of adults living with dementia.

To-date, there is no effective disease-modifying cure or treatment for dementia (Gauthier et al., 2016). Compounding this problem, there are a low number of pharmacological treatments in development, especially when compared to other disorders (Gauthier et al., 2016). Given the general failure of, and pessimistic outlook for, pharmacological interventions (Cummings et al., 2014), there has been a growing importance placed on research into the prevention of dementia (Livingston et al., 2020; Solomon et al., 2014). Broadly, preventative strategies focus on either avoiding dementia-related neuropathology or identifying cognitive decline as early as possible and slowing its progression (Hodes et al., 2019). To achieve these goals, preventative interventions target various modifiable lifestyle risk factors (Kivipelto et al., 2018). However, effective prevention requires identification of individuals with dementia at a very early stage of the disease time course, before neurodegeneration results in the emergence of symptoms, i.e., cognitive decline (McDade et al., 2020). Early identification is also critical to enable treatment at the earliest stage possible, in order to maximise the effect of any interventions (Mortimer et al., 2005).

The identification of individuals at risk for dementia and indeed the design and evaluation of interventions is complicated by the considerable variability in cognitive decline across individuals (R. S. Wilson et al., 2002). This variation cannot be fully attributed to neuropathology (Boyle et al., 2013) or brain structure (Hedden et al., 2014). This phenomenon was famously described by Katzman et al. (1988) who reported that ten individuals, with normal levels of cognitive function, possessed the neuropathological features of Alzheimer's disease (AD), when their brains were examined post-mortem. Further post-mortem studies have confirmed these findings (Bennett, Schneider, Arvanitakis, et al., 2006) and it is widely accepted that cognitive function is not completely dependent on brain structural health (Nilsson & Lövdén, 2018; Stern, 2002).

The lack of a one-to-one relationship between cognitive function and brain structure, means that screening tools using cognitive assessments may not correctly identify individuals with the underlying pathologies or structural damage at an early stage (Mortimer et al., 2005). Conversely, solely measuring brain structure or pathology may mistakenly identify individuals with normal cognitive function but underlying neuropathology, like those described by Katzman et al. (1988), as being at significant risk of cognitive impairment. Targeting interventions at such individuals could be an ineffective use of resources and could needlessly expose these individuals to potential side effects of pharmacological therapeutics (De Jager, 2005). The variability in this relationship further complicates the design of intervention efficacy because interventions targeted at modifying brain structure or pathology might not translate to effects on cognition. Similarly, the unexplained variability in this relationship can obscure the perceived efficacy of an intervention (Liyanage et al., 2018).

1.2 Sources of variability in cognitive decline

Three main constructs have been described that may account for the variability in cognitive decline that is not attributable to brain ageing or pathology. The first construct, brain reserve (BR), describes the neurobiological capital of the brain (Stern et al., 2020). Brain reserve can be conceptualised as the 'hardware', as it solely refers to the structural properties of the brain (Medaglia et al., 2017). Indicators of brain reserve include brain volume, dendritic branching, number of neurons, and number of synapses (Stern, 2009; Stern et al., 2020). Individuals are held to have a specific critical threshold, and once BR is depleted below that threshold, cognitive deficits emerge (Stern et al., 2020). Accordingly, individuals with greater BR are able to tolerate greater amounts of brain ageing or pathology, before suffering cognitive impairment. For example, an individual with a large brain reserve capacity might endure a certain level of brain injury or pathology but once this does not deplete their brain reserve capacity beyond the critical threshold, they would not sustain any

cognitive impairments. In contrast, an individual with a smaller brain reserve capacity might endure the same level of brain injury or pathology, which would deplete their brain reserve capacity beyond the threshold. Consequently, this individual would suffer cognitive impairments. As no processes are invoked in response to brain injury or pathology, BR has been described as a passive model of reserve (Stern, 2002, 2012).

The second reserve construct, brain maintenance (BM), reflects the reduced development over time of brain ageing and pathology as a result of genetic and lifestyle factors (Stern et al., 2020). Essentially, BM holds that variability in the development of age-related brain changes or pathology can explain variability in cognitive ageing (Nilsson & Lövdén, 2018). Individuals with greater BM have less age-related brain changes which preserves the integrity of the brain and therefore have reduced cognitive decline (Nilsson & Lövdén, 2018). BM is a modifiable construct and may be influenced by genetics, as well as life experiences and lifestyle factors (Nyberg et al., 2012). For example, increased BM has been associated with greater amounts of physical activity (Steffener et al., 2016) and more years of education (Gazzina et al., 2019; Steffener et al., 2016 but cf. Mungas et al., 2018; Zahodne et al., 2019). While BM and BR are related, there are important distinctions between the constructs (Stern et al., 2020). First, BM refers to the maintenance, or structural preservation, of the brain over time whereas BR refers to the status, or neurobiological capital, of the brain at a single point in time (Habeck et al., 2017; Stern et al., 2020). Second, BM protects against the accumulation of age-related brain changes or pathology whereas BR protects against the impact of age-related brain changes or pathology (Stern et al., 2020). Greater BM may support higher BR (Stern et al., 2020).

The third reserve construct, cognitive reserve (CR), is defined as the adaptability of cognitive or functional brain processes that explain individual differences in cognition in response to brain ageing or pathology (Stern et al., 2020). Individuals with more adaptable cognitive and functional brain processes are believed to be better able to cope with brain ageing or pathology and maintain normal cognitive function (Stern et al., 2020). As these processes may be invoked in response to brain ageing or pathology, CR has been described as an active model, in contrast to the passive model of BR (Stern, 2002). Moreover, whereas BR is concerned with the structural properties of the brain and is thus considered the “hardware” of reserve, CR is concerned with the processes and therefore can be thought of as the “software” of reserve (Medaglia et al., 2017; Stern, 2002). Like BM, CR is a modifiable construct that is thought to be influenced by genetics as well as life experiences (Stern et al., 2020), such as educational attainment (Malek-Ahmadi et al., 2017) or occupational complexity (Boots et al., 2015). Although similar factors may influence both BM and CR, they have been shown to be separable, or orthogonal, constructs (Habeck et al., 2017).

CR may be implemented via different mechanisms, including neural reserve, neural compensation, and generic CR networks (Steffener & Stern, 2012). Neural reserve holds that greater efficiency or capacity of neural networks enables successful cognitive performance in the face of age-related brain changes or pathology (Steffener & Stern, 2012). Neural compensation holds that greater ability to recruit alternative neural networks to perform a task, when the primary neural network for that task is disrupted, enables successful cognitive performance in the face of age-related brain changes or pathology (Steffener & Stern, 2012). Evidence for neural reserve and neural compensation has been reported from functional magnetic resonance imaging (fMRI) studies which compare the relationship of CR-related activation patterns in young vs older adults or in groups of cognitively healthy adults vs adults with mild cognitive impairment (MCI) or AD (Anthony & Lin, 2017; Steffener et al., 2011). These studies have revealed that activation within medial temporal lobe regions and suppression of activity within the default mode network (DMN) may underlie neural reserve (Anthony & Lin, 2017) whereas increased activation of frontal lobe regions may underlie neural compensation (Anthony & Lin, 2017).

CR may also be implemented via a generic CR neural network that is unrelated to specific task demands but is actively involved in many different cognitive processes (Steffener & Stern, 2012; van Loenhoud et al., 2020). Individuals with greater ability to express this generic, or task-invariant, network are better able to maintain cognitive performance across multiple tasks despite age-related brain changes or pathology (Stern et al., 2018). Studies assessing generic CR networks aim to identify CR-related patterns of activation or connectivity that are expressed across different cognitive tasks. Such studies have supported this implementation of CR as greater expression of generic networks active across multiple cognitive tasks has been associated with better fluid reasoning and episodic memory, beyond the effects of brain structure, as measured by mean cortical thickness (Stern et al., 2018; van Loenhoud et al., 2020).

The accurate measurement of these three constructs is important from a theoretical and research perspective in order to account for the variability in cognitive decline and therefore better understand individual differences in cognitive ageing. From a clinical perspective, the accurate measurement of these constructs is crucial in order to improve the early detection or prediction of cognitive decline, particularly in cognitively healthy adults. Individuals at risk for severe cognitive decline and/or dementia could be identified despite still displaying normal cognition if they were found to have lower BR, BM, or CR. This would enable effective secondary preventative strategies to be targeted at these individuals (McDade et al., 2020). Moreover, as research suggests that BM and CR, in particular, are modifiable and can be influenced by life experiences and lifestyle, accurately measuring

these constructs is necessary to identify their life experience and lifestyle determinants. This would improve the design of interventions aimed at increasing levels of BM and CR in order to prevent or slow cognitive decline. Finally, accurate measurement of BM and CR would further enable the efficacy of such interventions to be more precisely evaluated in terms of their effects on BM and CR.

Of the three constructs, BR is the most easily measured as it requires a single variable that reflects the structural capacity of the brain. For instance, in the famous example described by Katzman et al. (1988), the individuals who possessed the neuropathological characteristics of AD, but maintained cognitive performance, were reported to have significantly heavier brains and greater numbers of neurons. As such, brain weight and neuronal counts constitute simple post-mortem measures of BR. In-vivo measures of BR can also be obtained using neuroimaging measures including total intracranial volume (TIV; Groot et al., 2018; van Loenhoud et al., 2018; Vuoksimaa et al., 2013), total or regional grey matter (GM) volumes (Laubach et al., 2018), cortical thickness (Neth et al., 2020), or measures of white matter (WM) microstructural properties (Stern et al., 2020) such as fractional anisotropy of the genu of the corpus callosum (Neth et al., 2020). Other suggested in-vivo measures include dendritic spine length and synaptic density or integrity (Stern et al., 2020; van Loenhoud et al., 2018).

The measurement of BM and CR is less straightforward than BR as they cannot be assessed solely by a single variable, such as measurements of brain structural health or cognitive function (Habeck et al., 2017). Instead, the relationship between these two variables must be assessed in order to derive satisfactory measures (Habeck et al., 2017; Stern et al., 2020). The following sections of Chapter 1 reviews the measurement of BM and CR, focusing on the potential of accurate measurement, the approaches to measurement, and the challenges in developing accurate measures.

1.3 Brain Maintenance

The accurate measurement of BM has important research and clinical potential. As BM may explain some of the variability in cognitive decline, accurate measurement of BM may improve the identification of individuals at risk for dementia and indeed the design and evaluation of interventions. Early identification of at-risk individuals could be improved by identifying individuals based on their levels of BM, instead of focusing on their levels of cognitive function, which might not yet have begun to deteriorate. This could enable interventions to be targeted towards these individuals before the onset of significant cognitive decline.

The construct of BM suggests that improving or maintaining the youthfulness or integrity of brain structure and function may be a possible strategy for maintaining cognition or slowing cognitive decline (Nyberg et al., 2012). Accurate measurement of BM would enable researchers to firmly identify factors that are associated with improved BM. These factors could then be included in lifestyle interventions targeted at BM in order to prevent or reduce cognitive decline. This is important because there have been inconsistent findings relating specific factors or life experiences to better brain maintenance. For instance, while some studies have suggested that education might contribute to better BM (Gazzina et al., 2019; Steffener et al., 2016), this has been contradicted by others (Mungas et al., 2018; Zahodne et al., 2019).

Accurate measurement of BM would also improve the evaluation of lifestyle interventions as it would enable more accurate measurement of their efficacy. Typically, these interventions are evaluated with respect to their impact on cognitive function (Bhome et al., 2018; Whitty et al., 2020). However, if the intervention is targeted at BM, it would be useful to include a measure of its impact on the targeted mechanism, in addition to measuring the impact on cognition. If a lifestyle intervention was not shown to have any impact on cognitive decline as well as BM, then there would be stronger evidence suggesting that the intervention is ineffective. This could enable researchers to avoid spending excessive amounts of time and resources on ineffective interventions. On the other hand, if a lifestyle intervention failed to reduce cognitive decline but improved BM, this could suggest that the intervention might have some utility but might require adaptations, such as improved dose or duration. This would prevent researchers from unnecessarily discounting promising interventions. It could also suggest that the cognitive outcome measures may be obscured by practice effects (Elman et al., 2018) or individual-level factors such as comprehension levels, reading ability, self-efficacy, motivation, fatigue and fluctuations in concentration (McCaffrey & Westervelt, 1995).

An optimal measure of BM would use longitudinal data, including a measure of age-related brain change or pathology and a measure of associated cognitive change, to provide an index of the relative preservation of brain structure (Stern et al., 2020). However, cross-sectional data may also be used to measure BM (Cabeza et al., 2018), which is important given the costs and difficulty of obtaining sufficient longitudinal neuroimaging and cognitive data. Moreover, a cross-sectional measure would enable BM to be measured in individuals lacking prior neuroimaging data. An intuitive way of measuring BM with cross-sectional data is a ‘residual’ approach where the relative state of an individual’s brain is compared to the state expected for that age (Stern et al., 2020). Operationally, this takes the form of a residual from a regression of age on measures of brain structure. For this residual to be

considered a valid measure of BM, it would also need to be associated with a measure of cognitive function.

The general approach to deriving a cross-sectional measure of BM has been applied more broadly to develop a potential neuroimaging biomarker of biological ageing. This was motivated by the problem that chronological age is not the most accurate marker of an individual's rate of biological ageing (Sprott, 2010), as ageing is a process with significant heterogeneity across individuals (McCrory & Kenny, 2018). Consequently, ageing biomarkers are required to obtain additional information about an individual's health status and life expectancy (Dean & Morgan, 1988). The general approach to creating neuroimaging biomarkers of ageing has been to quantify the relationship between structural MRI data and chronological age, using machine learning, in order to estimate an individual's 'brain age'. Subtracting chronological age from the estimated 'brain age' results in a brain predicted-age difference score (*brainPAD*, also referred to as brain age gap, brainAGE, Brain-Age Score; Beheshti et al., 2018; Franke et al., 2010; Schnack et al., 2016) which quantifies how a person's brain health differs from what would be expected for their chronological age. Higher brainPADs reflect older brains, or accelerated brain ageing, and are associated with earlier mortality, weaker grip strength, reduced lung function, slower walking speed, and greater allostatic load (Cole et al., 2018). As a result, brainPAD, has been considered a promising biomarker of general brain ageing.

BrainPAD may further serve as a measure of BM as various studies have identified associations with cognition. Higher brainPADs have been associated with cognitive impairment (Liem et al., 2017), impaired fluid cognitive performance (Cole et al., 2018) and have been reported in adults with AD and MCI (Franke & Gaser, 2012; Gaser et al., 2013; Löwe et al., 2016). The reported associations with cognitive impairment suggest that brainPAD may be a useful measure of BM. However, this relationship between brainPAD and cognition could be biased by three factors: the inclusion of clinical samples in studies, the failure to statistically control for age when assessing the brainPAD-cognition relationship, and the failure to control for multiple comparisons. Consequently, it is unclear if the relationship between brainPAD and cognition is reliable in cognitively healthy adults. As such, the validity of brainPAD as an operational measure of BM may be limited.

Studies relating specific cognitive functions and brainPAD have been assessed in solely clinical samples (e.g., Cole et al. (2015), traumatic brain injury), or in mixed samples of clinical groups and healthy controls (e.g., Beheshti et al. (2018); AD, MCI, and healthy controls) and not samples comprised only of healthy adults. As such, the reported associations between brainPAD and specific domains of cognitive function in such studies

may be skewed towards statistical significance by the inclusion of the clinical samples with typically higher brainPADs and lower cognitive function. Consequently, these findings may not represent the brainPAD-cognition relationship in cognitively healthy adults. For example, Le and colleagues (2018) reported a significant negative correlation between brainPAD and response inhibition and selective attention in a sample of individuals comprised of healthy controls and patients with mood or anxiety disorders, substance use disorder and/or eating disorders. However, significantly increased brainPADs have been reported in mood disorders such as major depression (Koutsouleris et al., 2014) and in substance use disorders such as alcohol dependence (Guggenmos et al., 2017). As both major depression and alcohol dependence are associated with cognitive impairments (Chanraud et al., 2007; McIntyre et al., 2013), the significant brainPAD-cognitive function correlations reported across samples including such populations could be driven by the inclusion of such clinical groups.

While some studies have reported significant associations between brainPAD and cognition in cognitively healthy adults (Cole, Underwood, et al., 2017), the reported associations were not assessed after adjusting for age. It has now been empirically demonstrated that chronological age must be controlled for when testing relationships between brainPAD and cognitive functions (Le et al., 2018; Smith et al., 2019). Failure to correct for chronological age can result in false positive findings because some cognitive variables are correlated with chronological age – but *not* brain ageing – and brainPAD is typically correlated with chronological age (Le et al., 2018). In light of this recent work, it is unclear whether the association between brainPAD and cognition is independent of chronological age in cognitively healthy adults. Moreover, researchers testing the brainPAD-cognition relationship have tended to carry out multiple statistical tests of the correlation between brainPAD and various cognitive measures. The performance of multiple statistical tests can increase the Type I error and result in false positive findings (Ranganathan et al., 2016). However, some previous studies did not control for multiple comparisons when investigating the brainPAD-cognition relationship (Beheshti et al., 2018; Cole, Underwood, et al., 2017; Gaser et al., 2013).

Although some studies have investigated the relationship between brainPAD and cognition in cognitively healthy adults, while controlling for chronological age and multiple comparisons, there are conflicting results for most cognitive domains. For example, a significant correlation between verbal fluency and brainPAD was reported by Franke and colleagues (2013) whereas Richard and colleagues (2018) found no association between verbal fluency and brainPAD. The brainPAD-cognition findings are summarised in Table 1.1. In sum, the relationship between brainPAD and cognitive function in cognitively healthy

adults is currently unclear due to a lack of studies that have tested this relationship, adjusting for age and controlling for multiple comparisons, in non-clinical samples. Furthermore, the few studies adjusting for age and correcting for multiple comparisons in cognitively healthy adults, have reported conflicting evidence for associations between brainPAD and specific cognitive domains (Franke et al., 2013; Richard et al., 2018). Consequently, while brainPAD is a potentially useful and intuitive measure of BM, its relationship with cognitive function is not entirely clear and therefore its validity as a measure of BM is unsatisfactory.

Table 1.1 Summary of findings on the relationship between brainPAD and cognitive function.

Cognitive Domain	Measure	Reference	Sample	n	Sig.	Sig. in HCs	Age adj.	MC corr.
General Cognitive Status	MOCA	(Richard et al., 2018)	HC	265	✗	✗	✓	✓
	MMSE	(Kaufmann et al., 2019)	MCI; DEM	921; 707	✓	✗	✓	✓
	MMSE	(Gaser et al., 2013)	MCI	195	✗	✗	✗	✗
	CDR				✓	✗	✗	✗
	ADAS				✓	✗	✗	✗
	MMSE	(Löwe et al., 2016)	APOE; Non APOE	219; 186	✓	✗	✗	?[1]
	CDR		APOE; Non APOE	219; 186	✓	✗	✗	?[1]
	ADAS		APOE; Non APOE	219; 186	✓	✗	✗	?[1]
	MMSE	(Beheshti et al., 2018)	AD; pMCI; sMCI; HC	147; 112; 102; 146	✓	✗	✗	✓[2]
	CDR				✓	✗	✗	✓[2]
	ADAS				✓	✗	✗	✓[2]
	Composite measure[3]	(Cole, Underwood, et al., 2017)	HIVp; HC	161; 102	✓	✓	✗[4]	✓[2]
Verbal Fluency	Composite measure[5]	(Cole, Underwood, et al., 2017)	HIVp; HC	161; 102	✗	✓	✗[4]	✗
	Composite measure[6]	(Richard et al., 2018)	HC	265	✗	✗	✓	✓
	Semantic (Category Fluency Test)	(Franke et al., 2013)	DM2; HC	98; 87	✓	✓	✓	✓
	Phonemic (Letter Fluency Test)	(Cole et al., 2015)	TBI	89	✓	✗	✓	✓
Processing Speed	Composite measure[7]	(Cole, Underwood, et al., 2017)	HIVp; HC	161; 102	✓	✓	✗[4]	✗
	Composite measure[8]	(Richard et al., 2018)	HC	265	✗	✓	✓	✓
	TMT-A	(Cole et al., 2015)	TBI	90	✓	✗	✓	✓

[1] Inconclusive information on multiple comparison correction as corrections were not specifically outlined in relation to brainPAD-cognition tests but were used elsewhere in paper.

[2] Finding not corrected for multiple comparison but likely would have survived Bonferroni correction so not affected by lack of correction.

[3] Average of average standardised t-scores (adjusted for age, sex, education) across domains of verbal fluency, processing speed, executive function, memory, attention, and motor function

[4] T-scores controlled for effect of age on cognitive scores, but relationship between brainPAD and age was not controlled for, so not fully adjusted for age.

[5] Average of standardised t-scores (adjusted for age, sex, education) from Category Fluency and Letter Fluency tests

[6] Cluster measure combining Phonological Flow and Semantic Flow measures from CABPad (Willer, Pedersen, Forchhammer, & Christensen, 2016)

[7] Average of standardised t-scores (adjusted for age, sex, education) from TMT-A, WAIS-III Digit Symbol and Symbol Search, and Stroop Colour-Word Test

[8] Cluster measure combining processing speed parameters based on the Theory of Visual Attention obtained from test battery using CABPad

	CRT		TBI	66	✓	✗	✓	✓
Executive Function	Composite measure[9]	(Cole, Underwood, et al., 2017)	HIVp; HC	161; 102	✓	✓	✗[4]	✗
	TMT-B	(Cole et al., 2015)	TBI	90	✓	✗	✓	✓
Executive Function (without Processing Speed)	TMT-B minus TMT-A	(Cole et al., 2015)	TBI	90	✓	✗	✓	✓
Response Inhibition and Selective Attention	D-KEFS CWIT Composite measure [10]	(Richard et al., 2018)	HC	265	✗	✗	✓	✓
	D-KEFS CWIT (Inhibition vs Color Naming – scaled)	(Le et al., 2018)	HC, MOOD/ANX, SUD, ED	489	✗	✗	✓	✓
	D-KEFS CWIT (Inhibition/Switching)	(Cole et al., 2015)	TBI	89	✓	✗	✓	✓
	D-KEFS CWIT (Inhibition/Switching minus Baseline Stroop performance)		TBI	89	✗	✗	✓	✓
Sustained Attention	Composite measure[9]	(Cole, Underwood, et al., 2017)	HIVp; HC	161; 102	✗	✓	✗[2]	✗
Verbal Episodic Memory	Composite measure[11] (*General)*	(Cole, Underwood, et al., 2017)	HIVp; HC	161; 102	✓	✓	✗[2]	✗
	CVLT Immediate Recall, CVLT Delayed Recall, CVLT Learning 1-5, *(all tested separately*)	(Richard et al., 2018)	HC	265	✗	✗	✓	✓
	People Test *(Immediate)*	(Cole et al., 2015)	TBI	90	✓	✗	✓	✓
Working Memory	Composite measure[12]	(Richard et al., 2018)	HC	265	✗	✗	✓	✓
	Blocked Verbal N-back Task	(Scheller et al., 2018)	HC	34	✗	✓	✗	n/a[13]
Motor Function	Composite measure[14]	(Cole, Underwood, et al., 2017)	HIVp; HC	161; 102	✗	✓	✗[2]	✗
Intelligence	WASI Similarities *(Abstract verbal reasoning)*	(Cole et al., 2015)	TBI	90	✗	✗	✓	✓
	WASI Matrix Reasoning *(Non-verbal reasoning)*		TBI	88	✗	✗	✓	✓
	Composite measure[15] *(Fluid-type intelligence)*	(Cole et al., 2018)	HC	669	✓	✓	✓	✗

[9] Average of standardised t-scores (adjusted for age, sex, education) from TMT-B and WCST (Number of total errors, perseverative errors and responses)

[10] Cluster measure combining scores from the Colour-Naming, Reading, Inhibition, and Inhibition/Switching trails of the D-KEFS CWIT

[11] Average of standardised t-scores (adjusted for age, sex, education) from Rey Auditory Verbal Learning test and WMS-IV Visual Reproduction

[12] Cluster measure combining measures from working memory test of CABPad

[13] Only one test of brainPAD-cognition relationship conducted so multiple comparison correction not necessary

[14] Average of standardised t-scores (adjusted for age, sex, education) from Grooved Pegboard and Finger Tapping tasks

[15] Index derived from a principal components analysis of WASI-III Letter-number sequencing, digit span backwards, matrix reasoning, block design, digit symbol coding, symbol search

Moray House Test *(Childhood IQ)*	✗	✓	✗	✗

Note: Sig: results were statistically significant; Sig. in HC: results were statistically significant in healthy controls; *Age adj.: results were adjusted for age; MC corr.: results were corrected for multiple comparisons. MOCA = Montreal Cognitive Assessment; MMSE = Mini Mental State Examination, MCI = Mild Cognitive Impairment, DEM = Dementia; CDR = Clinical Dementia Rating Scale, AD = Alzheimer's Disease, ADAS = Alzheimer's Disease Assessment Scale, HC = Healthy Controls, APOE = APOE e4 carrier, Non APOE = APOE e4 non-carrier, sMCI = Stable MCI, pMCI = Progressive MCI, HIVp = HIV-positive, DM2 = Diabetes Mellitus Type 2, TBI = Traumatic Brain Injury, TMT-A = Trail Making Test A (Time to complete), CRT = Choice Reaction Time Task (Median reaction time), TMT-B = Trail Making Test B (Time to complete), D-KEFS CWIT = Delis-Kaplan Executive Function System D Color-Word Interference Test, MOOD/ANX = Mood/Anxiety Disorder, SUD = Substance use Disorder, ED = Eating Disorder, CVLT = California Verbal Learning Test, WASI = Weschler Abbreviated Scale of Intelligence.*

1.4 Cognitive Reserve

As with BM, the accurate measurement of CR has important research and clinical potential. In particular, the accurate measurement of CR could provide benefits in three key areas: 1) diagnosis and detection of cognitive decline and dementia; 2) clinical trials and intervention studies; and 3) design of CR-targeted interventions.

First, an accurate measure of CR could improve the clinical diagnosis of dementia as a clinician could account for CR when evaluating the cognitive status of the patient. The clinician would then be better able to estimate the optimal level of cognitive function for that patient against which their current level of functioning can be compared (Stern, 2012). Accurate measurement of CR might also enable future neuroimaging-assisted diagnoses of cognitive decline and dementia. While the application of machine learning and deep learning techniques to neuroimaging data harbours great potential for the early detection of age-related cognitive decline and dementia (Jo et al., 2019), this approach has had limited success to-date and is not yet suitable for clinical applications (Pellegrini et al., 2018). This was starkly illustrated by the inability of 33 international teams of experts to accurately predict cognitive decline from a rich set of neuroimaging variables, including diffusion tensor imaging (DTI), structural MRI, fluorodeoxyglucose-positron emission tomography (FDG-PET), and amyloid and tau PET, using state of the art machine learning algorithms (Marinescu, Oxtoby, et al., 2020). Accounting for differences in the functional brain and cognitive processes underlying CR could therefore potentially increase the accuracy of predictions of cognitive decline.

Second, an accurate measure of CR could improve clinical trials and intervention studies. Imbalances in CR across treatment and control groups could result in differential rates of cognitive decline that are unrelated to the intervention (Stern, 2012). Effective statistical control of CR when matching participants across groups, or when assessing outcomes, would refine the measurement of intervention efficacy (Mondini et al., 2016). The measure would further provide an effective means of stratification in intervention studies (Stern, 2012). If the measure is shown to strongly predict cognitive decline and/or dementia risk, it might also enable shorter and cheaper clinical trials as it could be used as a surrogate endpoint in place of the typical – costly and protracted – endpoints such as a slowed rate of cognitive decline and a reduced risk of developing Alzheimer's disease and dementia (J. K. Harrison et al., 2016; Vellas et al., 2008).

Third, an accurate measure of CR could enable better development of interventions designed to enhance CR in order to reduce or prevent cognitive decline and dementia. The majority of pharmacological interventions for Alzheimer's disease focus on either improving cognitive function or modifying the disease process

(Cummings et al., 2014). Nearly all of these interventions have failed to-date, with a failure rate of 99.6% between 2002 and 2012 (Cummings et al., 2014). Therefore, an alternative approach may be necessary and enhancing CR may be a promising strategy for delaying and/or reducing cognitive decline (Moga et al., 2019). This strategy requires an accurate measure of CR that would facilitate the assessment of the impact of the interventions on CR. Accurate measurement would improve the identification of CR determinants which could then be targeted via behavioural and lifestyle interventions. At present, support for these determinants is largely based on epidemiological evidence and the evidence base for associations between changes in these determinants and changes in CR is limited due to difficulties in measuring CR. Additionally, an accurate neuroimaging measure could uncover specific brain networks underlying CR, which could then be targeted using neuromodulation techniques such as connectome-based neurofeedback (Scheinost et al., 2020) and brain-computer interface neurofeedback (Arvaneh et al., 2019), or neurostimulation techniques such as transmagnetic stimulation (Kim et al., 2019).

Despite the considerable clinical and research potential of CR, there are significant difficulties in directly measuring CR (Conti et al., 2021; Stern et al., 2020) to the extent that it has been stated that “there is no direct way to quantify CR” (Marques et al., 2016, p. 3311) and that "no operational measures exist for accurately estimating an individual's CR" (Ward et al., 2015, p. 579). While the most direct measures of CR are likely to be developed using neuroimaging (Stern et al., 2020), the considerable cost of MRI scanning (Sarracanie et al., 2015) limits access to such measures, particularly in lower income countries (Ogbole et al., 2018). As such, socio-behavioural variables reflecting the degree of exposure to various lifetime experiences that are thought to contribute to CR, are often used as proxies of CR (Stern et al., 2020).

Valid measurement of CR requires some putative measure of CR (e.g., a proxy or a candidate neuroimaging measure) and two further components: a measure of brain structure/pathology and a measure of cognitive function (Christensen et al., 2008; Stern et al., 2020). A complete CR model refers to a model or analysis including all three components. In a complete CR model, the candidate measure of CR can be assessed with respect to its protective effects on cognition. The evaluation of the protective effects of the CR measure on cognition has been described as the *cognitive benefit criterion* (Franzmeier, Duering, et al., 2017). The cognitive benefit criterion can be satisfied via the observation of 1) an *independent effect* in which the candidate measure is positively associated with cognitive function, independent of brain structure, or 2) a *moderation effect* in which the candidate measure moderates the relationship between brain

structure and cognitive function (Stern et al., 2020; Stern & Habeck, 2018). Demonstration of a moderation effect is considered the strongest level of evidence for a CR measure, whereas the independent effect is considered a weaker level of evidence for a CR measure (Stern et al., 2020). A valid neuroimaging measure of CR is also required to show face validity, which can be demonstrated via a positive correlation with a CR proxy (Franzmeier, Duering, et al., 2017; Stern & Habeck, 2018).

1.5 Behavioural Measurement of Cognitive Reserve

CR is most commonly measured using proxy variables (Stern & Barulli, 2019), given that they are easy and inexpensive to obtain particularly in comparison to neuroimaging data. The rationale for using socio-behavioural proxies as measures of CR is that greater exposure to certain lifetime experiences increases the adaptability of cognitive and functional brain processes, thereby enabling a greater ability to cope with brain changes or damage (Stern et al., 2020). Considerable epidemiological evidence indicates a reduced risk and/or delayed onset of dementia and cognitive decline in individuals with greater educational attainment (Dekhtyar et al., 2016; H. X. Wang et al., 2012; Xu et al., 2016); occupational complexity/status (Andel et al., 2005; Kröger et al., 2008; Potter et al., 2007); literacy and/or verbal intelligence (Cervilla et al., 2000; Kaup et al., 2013; Manly et al., 2003; Pavlik et al., 2006); engagement in activities that were cognitively stimulating (Marioni et al., 2015; H. X. Wang et al., 2002); leisure-related (Akbaraly et al., 2009; Paillard-Borg et al., 2009); physical (Bowen, 2012; Marioni et al., 2015; Ogino et al., 2019; Rovio et al., 2005); and social (Marioni et al., 2015; H. X. Wang et al., 2002; Zhou et al., 2018). Proxies also provide a single value with a simple interpretation: a higher degree of exposure reflects greater CR. Furthermore, some proxies, such as educational attainment, are routinely collected as part of most ageing studies.

Despite their advantages, the use of proxies to measure CR has been criticised. First, some proxies, such as educational attainment, are typically static measures (Malek-Ahmadi et al., 2017) meaning that they tend not to change after a certain point in time (e.g., after early adulthood). However, CR is considered to be a dynamic construct that can change over time (Bettcher et al., 2019). Second, some argue that a single proxy fails to reflect the full CR construct which is thought to be influenced by a range of experiences (Kartschmit et al., 2019; Zahodne et al., 2013). Finally, proxies may also be associated with cognitive decline via mechanisms other than CR (Jones et al., 2011). For instance, greater educational attainment is correlated with higher socioeconomic status (Sirin, 2005) which is itself associated with slower cognitive decline (Marden et al., 2017) and reduced risk and prevalence of dementia (Fischer et

al., 2009; Yaffe et al., 2013). Low socioeconomic status is associated with various other factors, including stress and access to healthcare, which could exacerbate cognitive decline (Yaffe et al., 2013). As such, the protective effect of education on cognitive decline and dementia may be via mechanisms related to socioeconomic status, rather than CR (Zahodne, Stern, et al., 2015).

The limitations of individual proxies may be mitigated by averaging (cf. transformation methods such as PCA) multiple proxies to create a composite proxy measure that still provides a single summary value with a simple interpretation (Fleck et al., 2017; Pettigrew et al., 2017, 2020; Soldan et al., 2017; Steffener et al., 2014). Composite proxies allow for a wider range of contributions to CR and enable the inclusion of dynamic proxies that can change over time, such as verbal intelligence or engagement in activities (Malek-Ahmadi et al., 2017). Furthermore, composite proxies may attenuate the issue of non-CR mechanisms of individual proxies because alternative mechanisms (e.g., socioeconomic status) might only be associated with some proxies, such as educational attainment, but not others like social engagement. Some composite-type approaches, including factor analytic and latent variable models, measure CR using inappropriate *reflective measurement models*, where the observed CR proxies are effectively considered to be reflective of (i.e., caused by) the latent CR construct (Jones et al., 2011). Composite proxies are a more appropriate formative measurement model, where the observed proxies are considered to form, or cause, CR. Moreover, this approach can reflect the unique additive contributions of individual proxies, whereas factor analytic models reflect only the shared variance across different proxies (Stern et al., 2020).

While the composite approach offers advantages over the use of single proxies, there is no agreed-upon gold-standard composite proxy (Stern & Barulli, 2019) just as there is likewise no gold-standard individual proxy. Similarly, it is unclear which proxy should be used when assessing candidate neuroimaging measures of CR, as face validity is assessed via their association with CR proxies (Franzmeier, Duering, et al., 2017; Stern & Habeck, 2018). The considerable variation (S. L. Harrison et al., 2015; Opdebeeck et al., 2016) and lack of coherence in the use of proxies means that there is poor comparability across studies, as an effect observed for one proxy (e.g., educational attainment), may not be observed to the same degree for another (e.g., occupational complexity), even though both putatively reflect CR. It also provides researchers in the field of CR with an additional “researcher degrees of freedom” (Wicherts et al., 2016) such that several different proxies could be examined but only statistically significant results are reported.

There have been limited attempts to-date to assess the effects of different CR proxies on cognitive function. A systematic review of reviews and meta-analyses have found that education, occupational complexity/status and engagement in cognitively stimulating activities are individually associated with a reduced risk of dementia (S. L. Harrison et al., 2015) and positively associated with cognitive function in cognitively healthy older adults (Opdebeeck et al., 2016). Composites of these proxies and also including verbal intelligence have shown similar effects (S. L. Harrison et al., 2015; Opdebeeck et al., 2016; Roldán-Tapia et al., 2012). Across studies, education, as a single proxy, and composite proxies had moderate associations with cognitive function, with smaller associations found for occupational complexity/status and cognitively stimulating activities (Opdebeeck et al., 2016). Verbal intelligence and social engagement were also associated with a reduced risk of dementia although both were less frequently used compared to other proxies (S. L. Harrison et al., 2015).

The evidence reported in the systematic reviews and meta-analyses described above were obtained from incomplete models of CR, where the cognitive benefit criterion could not be assessed (Chapko et al., 2018). Chapko et al. (2018) sought to rectify this problem and conducted a systematic review of studies assessing CR proxies in complete CR models. 58% of all models assessing education reported positive evidence for education as a CR proxy, although this dropped to 38% of models within cognitively healthy cohorts. Chapko et al. concluded that the evidence for occupational complexity/status was inconclusive. One reviewed study provided evidence that greater engagement in cognitively stimulating activities in mid- and late-life provided CR effects (Reed et al., 2011). Conflicting results were found for more general leisure activity measures, with one study finding a protective effect (Scarmeas et al., 2003) while another reported a null effect (Borroni et al., 2009).

Verbal intelligence was not considered as a CR proxy by Chapko et al. (2018) in their systematic review, although it has been relatively widely used as a proxy. Negash et al. (2013) reported that verbal intelligence was positively associated with cognitive function controlling for global AD neuropathology, in a mixed sample of cognitively healthy older adults and older adults with MCI and dementia. In another mixed sample, a moderation effect was observed for verbal intelligence on the relationship between cognition and inferior temporal lobe tau deposition, but not global amyloid burden (Rentz et al., 2017). However, this moderation effect on tau deposition was not significant when the analysis was restricted to cognitively healthy older adults. Other studies have reported positive evidence for verbal intelligence as a CR proxy in cognitively healthy older adults, including a positive association with cognitive function controlling for

hippocampal atrophy (Topiwala et al., 2019), and a moderation effect on the relationship between cognition and fibre bundle length, an index of WM microstructural integrity (Baker et al., 2017). In the latter study, the reported moderation effect may have been confounded by age, as age is negatively associated with both fibre bundle length (Baker et al., 2014) and cognitive function (Salthouse, 2009, 2010), yet the analysis did not control for age.

Chapko et al. (2018) did not assess physical activity or social engagement as CR proxies, presumably because studies with complete models including these proxies were not available at the time of the research. Complete CR models assessing physical activity and social engagement have since been published. Conflicting evidence has been reported for physical activity, which was positively associated with cognition in the presence of neuropathology (Buchman et al., 2019) but not hippocampal atrophy (Topiwala et al., 2019). Positive evidence has also been reported for social engagement, which moderated the relationship between amyloid-beta deposition and cognitive decline (Biddle et al., 2019).

Mixed evidence for CR effects of composite proxies has also been published (note, composites were not assessed by Chapko et al., (2018)). The composite of verbal intelligence and education has been reported to moderate the relationship of subcortical GM volume and cortical thickness with fluid reasoning but not memory or processing speed and attention (Steffener et al., 2014). This composite has also been associated with memory controlling for GM volume (Kwak et al., 2020) and global cognition controlling for a composite AD-biomarker (Soldan et al., 2017). Aside from the composite of verbal intelligence and education, there is very little empirical evidence regarding the effects of different CR composites within complete models.

Composite proxies have also been used to measure CR via standardised questionnaires, such as the Cognitive Reserve Index Questionnaire (CRIq; Nucci, Mapelli, & Mondini, 2012), the Lifetime Experiences Questionnaire (LEQ; Valenzuela & Sachdev, 2007), and the Cognitive Reserve Scale (CRS; León, García-García, & Roldán-Tapia, 2014), among others (for a systematic review of CR questionnaires, see Kartschmit et al., 2019). However, the methodological rigour of the various CR questionnaires is not yet conclusive, with limited evaluation of the psychometric properties of most CR questionnaires (Kartschmit et al., 2019). In particular, most CR questionnaires have limited evidence of construct validity as they have not been evaluated in complete CR models which enable the assessment of the cognitive benefit criterion (Kartschmit et al., 2019; Malek-Ahmadi et al., 2017). Furthermore, while a

questionnaire can theoretically account for all of the proposed CR indicators, to-date no questionnaire accounts for the full set of the most common CR proxies: educational attainment, occupational complexity, verbal intelligence, engagement in leisure-, social-, physical- and cognitively stimulating-activities.

In conclusion, while proxies are routinely used to measure CR, there is a lack of consistent empirical evidence demonstrating the validity of individual and composite proxies as measures of CR, particularly for the measurement of CR in cognitively healthy older adults. Consequently, while proxies are the most common measure of CR, researchers and ultimately clinicians, may inadvertently use proxy variables that are not valid measures of CR. Similarly, it is unclear which proxy should be used to assess the face validity of neuroimaging measures of CR.

1.6 Neuroimaging Measurement of Cognitive Reserve

1.6.1 Structural Measurement of Cognitive Reserve

One approach to measuring CR with neuroimaging is the *CR residual,* where CR is operationally defined as the unexplained variance in cognitive function after accounting for the variance explained by brain structure and demographic factors (Reed et al., 2010). The CR residual was first developed using latent variable models (Reed et al., 2010; Zahodne et al., 2013) but more straightforward methods, such as multiple regression, have since been used. With the regression method, a measure of cognitive function, typically episodic memory (Franzmeier, Göttler, et al., 2017; Franzmeier, Hartmann, et al., 2017; Habeck et al., 2017; Reed et al., 2010; Zahodne et al., 2013), is used as the dependent variable with independent variables including a measure of brain structure and demographic factors such as age, gender, and sometimes education (Franzmeier, Göttler, et al., 2017; Franzmeier, Hartmann, et al., 2017; D. H. Lee et al., 2019; Zahodne, Manly, et al., 2015). CR is then indexed by the residuals from this linear regression, where positive residuals reflect greater CR as cognitive performance is better than expected given the individual's brain structure (Reed et al., 2010).

The CR residual is an intuitive and relatively easily computed single scalar index of CR. It does not require the use of socio-behavioural proxies and indeed is a more direct measure of CR than proxies (Stern et al., 2020). The measure may be dynamic (Stern et al., 2020) and therefore could be used to track changes in CR over time, enabling the measurement of the efficacy of interventions aimed at increasing CR (Zahodne, Manly, et al., 2015). This measure has been empirically supported as higher values (i.e., more positive residuals) have been associated with a reduced risk of conversion to dementia (Reed et al., 2010). Moderation effects on the relationship between atrophy and cognitive decline have also been observed using the CR residual

such that atrophy was more strongly related to cognitive decline in individuals with lower (i.e., more negative) CR residuals (Reed et al., 2010). Similar moderation effects have been demonstrated for the relationship between brain structure and executive function (Reed et al., 2010) and language ability (Zahodne et al., 2013). Face validity has also been established for the CR residual as it has been associated with common CR proxies including educational attainment, occupational complexity, and verbal intelligence (Habeck et al., 2017; D. H. Lee et al., 2019).

Despite the various advantages of the CR residual, it does possess some important limitations. First, it relies on structural neuroimaging data, and therefore does not directly measure the functional processes underlying reserve. Second, while it may be a more direct measure than CR proxies, it is still an indirect measure which necessarily will contain a significant proportion of measurement error, by definition (Ewers, 2020). Third, different combinations of independent (i.e., brain structure) and dependent (i.e., cognitive function) variables have been used across studies. This introduces variability to the measure across different studies (Stern et al., 2020) and reduces the comparability across studies (Ewers, 2020). This issue extends to the choice of demographic variables used in the CR residual which have sometimes been inconsistent with CR theory. While education has been included as a demographic predictor variable in some CR residuals (Reed et al., 2010; Zahodne et al., 2013; Zahodne, Manly, et al., 2015), education is regarded as a key indicator of CR (Stern, 2002) and is the most commonly used CR proxy (Opdebeeck et al., 2016). The inclusion of education as an independent variable in a regression model creates a CR residual that explicitly excludes the variance in cognition that is attributed to education. Consequently, this CR measure will not reflect any information from a presumed key contributor to CR. Finally, the CR residual does not provide any spatial information about CR and is uninformative about the underlying neural processes.

An alternative to the CR residual, which uses a similar approach, but can provide spatial information about CR is the ‘W-score’ measure (van Loenhoud et al., 2017). The W-score is essentially a reverse of the CR residual, as a measure of brain structure is regressed on cognitive function and demographic variables. However, unlike the CR residual which is computed for global summary measures of brain structure, the W-score is computed at the voxel level. This provides spatial information about CR as demonstrated by van Loenhoud et al. (2017) who reported a strong association between educational attainment and mean W-scores in a temporoparietal region of interest. Like the residual measure, the W-score has been associated with cognitive decline (van Loenhoud et al., 2019) and progression to more advanced stages of dementia (van

Loenhoud et al., 2017). Face validity has also been established for the W-score as lower scores – reflecting higher CR – were associated with higher levels of education (van Loenhoud et al., 2017). However, like the CR residual, the W-score measure will inevitably contain a large proportion of measurement error and fails to directly assess the functional brain processes underlying CR as it is solely based on structural data (Stern et al., 2020).

1.6.2 Functional Measurement of Cognitive Reserve: EEG

The functional brain processes and networks underlying CR can be more directly assessed using functional neuroimaging (Stern et al., 2020). Two functional neuroimaging methods, electroencephalography (EEG) and magnetoencephalography (MEG) enable the measurement of functional brain processes and networks by measuring the electrical and magnetic activity of the brain, respectively (van Straaten & Stam, 2013). However, neither method has been widely used to study CR, a systematic review identified only eleven EEG and five MEG valid studies of CR (Balart-Sánchez et al., 2021). Nonetheless, they may offer practical benefits over fMRI measures, as EEG, in particular, is better tolerated (Fleck et al., 2017) and more widely accessible (Cassani et al., 2018; Farina et al., 2020). Indeed, EEG may be a promising method for measuring CR as various functional connectivity metrics have been shown to correlate with the severity of cognitive decline (Briels et al., 2020).

Promising EEG targets for developing a measure of CR may include coherence during resting-state and an event-related potential, the P300, which reflects neural efficiency (van Dinteren et al., 2014). Age-related decreases in coherence, a functional connectivity metric, were observed in a low CR group whereas an age-related increase was observed in a high CR group (Fleck et al., 2017). Positive associations between CR and imaginary coherence were observed in the theta band over a right frontocentral region, with negative associations in the theta band over a right parietotemporal region, and negative associations in the alpha band over an occipitoparietal region (Moezzi et al., 2019). CR was also positively associated with greater neural efficiency, measured by the P300 (Gu et al., 2018; Speer & Soldan, 2015). In cognitively healthy older adults, this increased efficiency was associated with better task performance (Gu et al., 2018). Other EEG indices, namely measures of relative power in each frequency band derived from spectral analysis, have not shown associations with cognitive reserve (Amodio et al., 2017).

Promising MEG targets include measures of brain oscillations that have been positively related to CR in cognitively healthy older adults, specifically higher resting-state gamma power in the right temporal region and higher beta intensity during an n-

back task in the parietal and occipital regions (Yang & Lin, 2020). However, in adults with MCI, beta power was not correlated with CR (López et al., 2016). Overall, while there are some promising initial findings relating EEG and MEG measures to CR, no candidate measures have been assessed in relation to the cognitive benefit criterion or in terms of their face validity. Furthermore, no individual-level measures have yet been developed using EEG or MEG.

1.6.3 Functional Measurement of Cognitive Reserve: task-based fMRI

Compared to EEG and MEG, fMRI has been more widely used in research investigating possible measures of CR. Task-based fMRI has been widely used to investigate potential patterns of task-related activations that are associated with CR. A systematic review of fMRI studies concluded that increased activation of frontal lobe regions and decreased activation of medial temporal lobe regions may index CR as these activation patterns were associated with higher CR in cognitively healthy older adults, whereas they were associated with lower CR in younger adults (Anthony & Lin, 2017). CR may also be reflected by suppression of activation within the DMN as reduced activity was associated with higher CR in multiple studies (Anthony & Lin, 2017). Another possible marker of CR, identified using task-based fMRI, is global functional connectivity of a region of the left frontal cortex, Brodmann area 6/44. Increased global functional connectivity in this region, a key hub of the frontoparietal network (FPN), was associated with both higher education and higher CR as measured with a CR residual (Franzmeier, Hartmann, et al., 2017). Connectivity of this region was also shown to mediate the relationship between increased efficiency of the DMN and dorsal attention network with memory performance (Franzmeier et al., 2018).

Although task-based fMRI studies have identified possible patterns of activation and functional connectivity that are related to CR, these patterns have rarely been tested in regards to their protective effects on cognition (Belleville et al., 2021). Consequently, the validity of such activation patterns as candidate measures of CR is largely unclear. The protective effects of a pattern of activation specific to the right inferior temporal gyrus was tested by Belleville et al. (2021). Increased activation in this region was observed during an associative memory task and was positively associated with a composite proxy measure of CR. A moderation effect of this pattern of activation was subsequently observed on the relationship between hippocampal volume and associative memory performance. This effect indicated that individuals with reduced hippocampal volumes may sustain cognitive performance via greater activation of the right inferior temporal gyrus. Nevertheless, focusing on specific regions of activation or connectivity may not be an optimal approach to deriving a measure of CR, as CR may also be influenced by

variations in connectivity that are globally distributed across the entire brain rather than just variations in connectivity limited to specific candidate regions. Indeed, Zhao et al. (2021) reported that individual differences in cognition were more accurately predicted by global patterns of task-related activations than by activations specific to localised brain regions.

Globally distributed activations and connectivity may be particularly relevant to identify generalised neural networks that may underlie CR (Steffener et al., 2011; Steffener & Stern, 2012). These networks have been described as generic and task-invariant because the network is expressed across different tasks and is unrelated to specific task demands (Stern et al., 2018). Two notable studies have been attempted to identify and measure these networks using task-based fMRI. The first study identified a covariance pattern of activation across twelve different fMRI tasks (Stern et al., 2018). This covariance pattern was shown to display face validity as pattern expression was positively correlated with a CR proxy, verbal intelligence. Furthermore, protective effects were observed for this covariance pattern, as it was positively associated with fluid reasoning, after controlling for cortical thickness and it moderated the relationship between cortical thickness and fluid reasoning. The loadings of the identified covariance pattern suggested that individuals with higher CR had stronger task-related activity in clusters within the cerebellum, medial frontal gyrus, and the anterior portion of the bilateral superior temporal gyrus, but lower activity in clusters within the bilateral inferior parietal lobe, bilateral middle frontal gyrus, and bilateral inferior frontal gyrus (Stern et al., 2018).

The second study, using the same dataset as Stern et al. (2018), attempted to measure a generic CR network using a measure of task potency (van Loenhoud et al., 2020). Task potency was calculated as the change in connectivity, from a resting-state baseline, in response to a task. Positive task potency values represented greater connectivity, or synchronisation of activity, between different brain nodes whereas negative values reflected a possible decoupling of activity in different nodes. A network was identified where task potency was significantly associated with verbal intelligence, a CR proxy, and which was positively associated with episodic memory and fluid reasoning, independently of cortical thickness (van Loenhoud et al., 2020). This network was relatively sparse with 57 identified connections (i.e., edges) mostly located within the DMN followed by the FPN and the salience network. The most highly connected network, the DMN, was solely comprised of connections that were negatively associated with verbal intelligence.

To-date, task-based fMRI studies have identified specific regions where activation exerts protective effects prescribed to CR, and have been used to derive potential generic, or task-invariant, CR networks. Despite these promising findings, developing a measure of CR using task-based fMRI may not be the optimal approach to measuring CR. One problem with this approach is that task-related activations could be confounded by various individual-level factors which affect task performance and engagement, including task difficulty (Stern, 2005), motivation, concentration, and fatigue (McCaffrey & Westervelt, 1995). Moreover, task-fMRI can be difficult for clinical populations, including individuals with cognitive impairment, and therefore may be less useful to measure CR in clinical populations (Franzmeier, Caballero, et al., 2017). Additionally, whereas resting-state fMRI data are acquired in a relatively standard manner across different sites (Woodward & Cascio, 2015), task-based data may be less standardised due to differences in experimental designs and stimulus definitions of the task (Mennes et al., 2013). As a result, a task-based fMRI measure of CR may not be as easily shared across research groups or sites in comparison to a measure that can be generated using resting-state data.

1.6.4 Functional Measurement of Cognitive Reserve: resting-state fMRI

Resting-state fMRI has various advantages for the measurement of CR in comparison to task-based fMRI. A measure derived from resting-state fMRI would less be affected by individual-level factors that may confound task-related activations. Indeed, resting-state fMRI measures have high test-retest reliability (Shehzad et al., 2009) and can have better signal-to-noise ratios than task-related activations (Fox & Greicius, 2010). Resting-state fMRI also has better clinical utility as no task-related demands are placed on participants and therefore can be applied to a wider population, including individuals with significant cognitive impairment (Fox & Greicius, 2010). Resting-state data can be more easily shared and aggregated with data from other sites as part of large data-sharing initiatives, thereby enabling greater use of any derived CR measures (Mennes et al., 2013; Woodward & Cascio, 2015).

As a proof of concept, Stern and Habeck (2018) demonstrated that resting-state fMRI may be a suitable modality for measuring CR. They generated a relatively simple metric, the intraindividual variability of resting-state functional connectivity, and assessed its validity as a candidate neuroimaging measure of CR. The intraindividual variability of resting-state functional connectivity measures regional variation in connectivity with lower values reflecting relatively greater uniformity in whole-brain functional connectivity. This single summary value was obtained for each participant by calculating the standard deviation across all functional connections within the brain. This

metric demonstrated face validity as lower values were associated with higher verbal intelligence, a CR proxy. The metric was also reported to satisfy the cognitive benefit criterion, as lower values were positively associated with cognition, as measured by vocabulary performance, independently of cortical volume and cortical thickness (Stern & Habeck, 2018). However, this measure was not validated on unseen data so the generalisability of this metric to novel data is unclear. Nonetheless, while more sensitive functional connectivity measures, accounting for the differential associations between various functional connections and CR, are likely to be more informative, the results reported for this blunt metric demonstrated the viability of resting-state fMRI for measuring CR.

Resting-state fMRI has also provided evidence suggesting that connectivity of specific functional networks, including the DMN, FPN and salience network, may underlie CR or form part of a broader generic CR network. These three networks have been separately associated with slower global cognitive decline, controlling for age, education and GM volume (Buckley et al., 2017). Moreover, connectivity of these three networks moderated the relationship between amyloid burden and cognitive decline, such that individuals with stronger connectivity and high amyloid burdens were less vulnerable to cognitive decline than those with weaker connectivity and high amyloid burdens (Buckley et al., 2017). The FPN may be particularly pertinent as increased connectivity of the FPN was related to higher CR, as measured by years of education, in individuals with MCI (Franzmeier, Caballero, et al., 2017; Serra et al., 2016). Global connectivity of the FPN has also been found to moderate the impact of WM lesions on executive function in cognitively healthy older adults (Benson et al., 2018). Connectivity of a specific hub of the FPN, the left frontal cortex (Brodmann area 6/44), has been consistently associated with CR in adults with MCI. Stronger negative connectivity of this hub to the DMN and stronger positive connectivity to the dorsal attention network have been associated with higher education and higher CR, as measured by a CR residual, in two separate cohorts of adults with MCI (Franzmeier, Göttler, et al., 2017). Global connectivity of this hub was further shown to mitigate the negative effect of glucose hypometabolism in the precuneus on memory ability in adults with MCI (Franzmeier, Duering, et al., 2017).

Compared to the FPN, there is less consistent evidence relating resting-state connectivity of the DMN and salience network to CR. The protective effects of global FPN connectivity identified by Benson et al. (2018) were not observed for the DMN or the salience network as a whole. However, connectivity of a specific node of the salience network, the anterior cingulate cortex to the medial frontal cortex was reported to

moderate the relationship between WM lesions and executive function (Benson et al., 2018). Connectivity of this node with regions including the right hippocampus, right posterior cingulate cortex/gyrus, left inferior frontal lobe, and left angular gyrus has also been positively associated with a CR proxy, years of education (Arenaza-Urquijo et al., 2013). Higher connectivity of these connections were further associated with verbal fluency, but as this association did not adjust for brain structure, the cognitive benefit criterion was not satisfied. Increased connectivity of the anterior cingulate cortex has also been associated with higher CR in adults with MCI and in cognitively healthy adults (Serra et al., 2016). These findings suggest that functional connectivity of the salience network, especially the anterior cingulate cortex, may be a functional correlate of CR. Despite evidence from task-activation studies implicating the DMN in CR, evidence for a role of resting-state connectivity of the DMN in CR, appears to be limited to a single study where DMN connectivity demonstrated a protective effect on cognition (Buckley et al., 2017).

In addition to functional networks, resting-state functional connectivity of specific neuroanatomical landmarks have also been linked to CR. D. H. Lee et al. (2019) reported that functional connectivity of the right middle temporal pole to the left amygdala and superior temporal pole was positively associated with a CR residual. In a subgroup analysis restricted to cognitively healthy older adults, stronger connectivity of the right precentral gyrus to the bilateral cuneus, bilateral supplementary motor area, and left post-central gyrus was related to higher CR. The right middle temporal pole and right precentral gyrus were also identified as key nodes in a widely distributed resting-state network where greater connectivity was associated with higher education in cognitively healthy older adults (Marques et al., 2015). Degree strength and betweenness centrality of the inferior temporal gyrus were positively associated with a CR residual in cognitively healthy adults (Marques et al., 2016), supporting the protective effects on cognition demonstrated for task-related activation of this region (Belleville et al., 2021). This CR residual was also associated with local efficiency and clustering of the middle occipital cortex as well as the cuneus, another key node of the widely-distributed resting-state network described by Marques et al. (2015). In conclusion, while connectivity of functional networks, including the FPN, DMN, and salience network, may underlie a generic CR network, connectivity of specific regions including the temporal pole, precentral gyrus, inferior temporal gyrus, and the cuneus may further contribute to such a network.

In sum, there has been considerable growth in the number of studies using neuroimaging to investigate and attempt to measure CR. EEG and MEG have been less

frequently used to study CR than MRI, and to-date no individual-level measures using these two modalities have been described that display face validity or protective effects on cognition. The CR residual has been consistently shown to display face validity and demonstrate protective effects on cognition, but it is suboptimal as it is based on structural neuroimaging. Task-based fMRI measures have similarly displayed face validity and demonstrated protective effects (Stern et al., 2018; van Loenhoud et al., 2020) but cannot be easily applied to data collected from the more clinically applicable resting-state fMRI. Resting-state fMRI has been used to identify key networks and regions where functional connectivity may underlie CR but its potential for deriving a measure of CR is mostly unexplored.

1.7 Specific aims of the research

Based on the research reviewed to this point in Chapter 1, there is a clear need for the development and validation of objective measures of BM and CR in cognitively healthy older adults. While a potential measure of BM, brainPAD, has been previously developed, the validity of this measure is uncertain as the relationship between brainPAD and cognitive function has been obscured by a lack of studies in cognitively healthy older adults that statistically controlled for the effects of age and corrected for multiple comparisons. Moreover, although socio-behavioural proxies are the most commonly used measures of CR, the validity of many proxies is unclear as there is a lack of empirical evidence from complete CR models tested in cognitively healthy older adults. Finally, while functional neuroimaging should provide the most direct measure of CR, sensitive and generalisable measures that can be applied to resting-state fMRI data have not yet been described.

Chapter 2 examined the validity of brainPAD as a measure of BM. A penalised regression with cross-validation was applied to structural MRI data, collated from open-access datasets, in order to predict chronological age. Chronological ages were then subtracted from the predicted ages to create brainPAD scores. This model was then applied to three independent datasets, which contained measures of cognitive function. Across these three datasets, the association between brainPAD and specific domains of cognitive function were assessed to determine the validity of brainPAD as an operational measure of BM in cognitively healthy adults.

Chapter 3 established the validity of different socio-behavioural variables as proxy measures of CR. Complete CR models, containing a CR proxy, a measure of brain structure, and a measure of cognition were created in two datasets of cognitively healthy adults. CR proxies included five standard CR proxies: educational attainment, occupational complexity, verbal intelligence, engagement in leisure activities, and engagement in physical activity. All possible combinations of composite measures of these CR proxies were also included. The validity and robustness of these different proxy variables were assessed using hierarchical moderated linear regressions. The analysis framework enabled the identification of the CR proxy with the largest independent associations with cognition. These results established data-driven recommendations supporting the selection of specific CR proxies when measuring CR and when assessing the face validity of candidate neuroimaging measures of CR.

Chapter 4 developed and validated a novel functional neuroimaging measure of CR. Connectome-based predictive modelling was applied to task-based functional connectivity in order to predict a CR residual measure in one dataset. The measures

derived from this model were assessed in terms of their face validity and their ability to satisfy the cognitive benefit criterion (i.e., by demonstrating protective effects on cognition). This model was then applied to an independent dataset in order to assess the generalisability and validity of these measures when generated using the more clinically applicable and widely usable resting-state fMRI data.

2 Chapter 2: Validation of the brain-predicted age difference as a measure of brain maintenance

2.1 Introduction

BrainPAD may be useful and intuitive measure of BM but, as outlined in Chapter 1 (see section 1.3 and Table 1.1), the relationship between brainPAD and cognition is unclear. To date, the relationship between brainPAD and specific cognitive functions has not been systematically examined using appropriate statistical methods in cognitively healthy adults. As a result, the validity of brainPAD as an operational measure of BM remains to be determined.

The first step in generating a brainPAD score is creating a feature set of neuroimaging data which is correlated with chronological age. Neuroimaging data have high dimensionality, which can result in overfitting and overoptimistic predictions (Whelan & Garavan, 2014). Brain age prediction models thus rely on feature engineering techniques such as principal component analysis (PCA; Franke et al., 2010; Gutierrez Becker, Klein, & Wachinger, 2018) or even dot products of different features (e.g., vectors of GM and WM voxels as in Cole et al., 2015; Cole, Ritchie, et al., 2018; Cole, Underwood, et al., 2017) in order to reduce the dimensionality of the data (Mwangi et al., 2014). These techniques map the original variables onto a feature space (in effect, creating 'new' variables) typically using linear transformations in the case of dot products (Snyder et al., 2013), although non-linear transformation may also be used for kernel methods (Honeine & Richard, 2009; Kwok & Tsang, 2004). While these models create generalisable and accurate predictions, this may come at the cost of reduced interpretability of the contributions of the features (Bunea et al., 2011; Mateos-Pérez et al., 2018), which is important for assessing the neurobiological validity of the model (Woo et al., 2017) and to identify specific brain areas for further investigation (Scheinost et al., 2019).

Due to the importance of interpretability in neuroimaging, unlike with other data (e.g., credit card transactions for fraud detection), the application of machine learning to MRI does not necessarily involve the goal of achieving the highest accuracy (Mateos-Pérez et al., 2018). While methods do exist for projecting the 'new' variables back from the feature space to the input space (Honeine & Richard, 2009; Kwok & Tsang, 2004; Snyder et al., 2013), thus enabling interpretability of models employing dot products, PCA or kernel methods, these methods are not always implemented and/or reported in brain-age papers (Cole et al., 2015, 2018; Cole, Poudel, et al., 2017; Gaser et al., 2013; Gutierrez Becker et al., 2018; Nenadić et al., 2017). In contrast, penalised regression

methods (e.g., the Elastic Net; Zou & Hastie, 2005) do not require the back-projection of coefficients from feature space to input space and therefore have good interpretability, particularly when less complex feature sets are used (Luo et al., 2019).

GM data is particularly well-suited for age prediction as GM volume linearly declines with age (but cf. Fjell et al., 2013) whereas WM volume has a less straightforward relationship with age, as it doesn't decline significantly until middle age (Farokhian et al., 2017; Ge et al., 2002). The Elastic Net is a machine learning model well-suited to the high dimensionality and multicollinearity inherent in neuroimaging data as shown by the finding that it produced the most consistent predictions as compared to various other models over datasets with varying sample-, feature set-, and effect-sizes (Jollans et al., 2019).

A final challenge in the development of neuroimaging biomarkers, or neuromarkers, is ensuring the generalisability of the neuromarker to new data. For practical reasons, cross-validation, where a dataset is split into a training set and a test set (Varoquaux et al., 2017), is often used as an estimate of model accuracy for new data (Jollans & Whelan, 2018; Scheinost et al., 2019). However, cross-validation accuracy estimates are often optimistically biased and can vary considerably (Varoquaux et al., 2017), particularly when preprocessing and feature selection are carried out on the entire dataset before splitting it into training and test sets (Dwyer et al., 2018; Woo et al., 2017). As such, the gold-standard for assessing the external validity and generalisability of a neuromarker is by testing how the model performs on a completely independent held-out dataset (Jollans & Whelan, 2018).

Several brainPAD studies have externally validated their models (Beheshti et al., 2018; Cole et al., 2015, 2018; Cole, Underwood, et al., 2017; Franke et al., 2010; Gutierrez Becker et al., 2018; Lancaster et al., 2018; Liem et al., 2017; Madan & Kensinger, 2018; Varikuti et al., 2018), but only a few studies have reported model performance in terms of accuracy (i.e., correlation or mean absolute error between brain-predicted age and chronological age) on the external validation dataset (Cole et al., 2015; Lancaster et al., 2018; Liem et al., 2017; Madan & Kensinger, 2018). This does not necessarily cast doubt on the validity of the models whose accuracy is reported in terms of internal cross-validation performance. However, not reporting the external validation performance limits the interpretation of the accuracy and generalisability of various brainPAD models as typically performance will be lower in the external validation dataset.

In order to clarify the unclear relationship between brainPAD and specific domains of cognitive function, an interpretable brainPAD model was created by using a cross-validated Elastic Net regression to predict chronological age from GM voxel-wise data in 1,359 T1 weighted MRI scans. To externally validate this model, it was then applied to MRI data from three independent datasets, Dokuz Eylül University (DEU; n=175), the Cognitive Reserve/Reference Ability Neural Network study (CR/RANN; n=380), and The Irish Longitudinal Study on Ageing (TILDA; n=487). To determine the validity of brainPAD as an operational measure of BM and to establish the specific domains of cognitive function that are reliably correlated with brainPAD across different datasets, the correlation between brainPAD scores and several cognitive measures across the three datasets were subsequently assessed.

2.2 Methods

2.2.1 Study Design

The present study used data from open-access neuroimaging repositories to form a training set in which a machine learning model was developed. Data from three separate datasets, DEU, CR/RANN, and TILDA, were then used to form three external validation sets in which the machine learning model was validated and the relationship between brainPAD and cognitive function was investigated. In all cases, the data was collected prior to conception and design of the present study. The target population were cognitively healthy adults.

2.2.2 Participants

Training Set

The data were comprised of MRI scans from 1,359 cognitively healthy adults (mean age 40.04 years, SD = 17.78 years, range = 18.00 - 88.36 years; 855 females) drawn from various open-access data repositories (see Table S1 in 7.1.1 Supplemental Methods). Inclusion criteria for the training cohort were: over 18 years old, age and gender data available, and not diagnosed with any neurological, psychiatric or major medical conditions.

Test Set 1 – DEU

The first test set was comprised of 175 community-dwelling adults (mean age = 68.95 years, SD = 8.59 years; range = 47.56 – 93.51 years; 104 females) recruited as part of a study conducted at Dokuz Eylül University, Izmir, Turkey. Exclusion criteria included history of neurological or psychiatric diseases, use of psychotropic drugs including cholinesterase inhibitors, traumatic brain injury, history of stroke, drug and/or alcohol addiction and uncontrolled systemic diseases.

Test Set 2 – CR/RANN

The second test set was comprised of 380 community-dwelling adults (mean age = 52.41 years, SD = 17.09 years; range = 19 – 80 years; 210 females) who participated in the CR/RANN studies (Stern et al., 2014, 2018). These participants were screened for MRI contraindications, hearing and visual impairments, medical or psychiatric conditions, and dementia and MCI. Further inclusion criteria were a score of over 135 on the Mattis Dementia Rating Scale (Jurica et al., 2001), a reading level at least equivalent to the US 4th grade, and minimal complaints of functional impairment.

Test Set 3 –TILDA

The third test set was comprised of an MRI subset of a nationally representative longitudinal cohort study of community-dwelling adults in Ireland (Kearney et al., 2011; B. J. Whelan & Savva, 2013). This data was collected during Wave 3 of the TILDA study (Donoghue et al., 2018). All participants were screened for MRI contraindications. From an initial subset of 553 participants, participants were excluded due to motion artefacts (n = 32), presence of lesions, (n = 18), motion artefacts and presence of lesions (n = 1) missing a portion of the cerebellum (n = 2), a history of Parkinson's disease, stroke, or transient ischemic attack (n = 11) and no cognitive data (n = 2). The final test set was comprised of MRI data from 487 participants (mean age = 68.6 years, SD = 7.21 years; range = 50 – 88 years; 260 females).

2.2.3 MRI data acquisition

Training Set

A range of T1-weighted MRI scans from different scanners and using different protocols were used as the training set (see Table S1 in 7.1.1 Supplemental Methods).

Test Set 1 – DEU

DEU participants underwent a 10 minute T1 scan in a 1.5 T Philips Achieva scanner as part of a larger 20 minute MRI battery. Two separate protocols were used for scans included here. The Alzheimer's Disease Neuroimaging Initiative T1 protocol was followed for 126 scans using the turbo field echo sequence with the following parameters: number of slices = 166, FOV = 240mm^3, matrix size = 256×256, slice thickness = 1 mm, slice gap = 0 mm, TR = 9 ms, TE = 4 ms. For 49 scans, a local protocol using a gradient echo sequence was followed with the following parameters: FOV = 230mm^3, matrix size = 400×512, slice thickness = 1 mm, slice gap = 0 mm, TR = 25 ms, TE = 6 ms.

Test Set 2 – CR/RANN

CR/RANN participants underwent a 5 minute T1-weighted 3D magnetization-prepared rapid gradient echo (MPRAGE) scan in a 3T Philips Achieva scanner as part of a larger 2-hr imaging battery. The following parameters were used: FOV = 256×256×180 mm, matrix size = 256×256, slice thickness = 1 mm, slice gap = 0 mm, TR = 6.5 ms, TE = 3 ms.

Test Set 3 – TILDA

TILDA participants underwent a 5 minute 24 seconds T1 MPRAGE scan in a 3T Philips Achieva scanner as part of a larger 45 minute MRI battery. The following parameters were used: FOV = 240×240×162mm^3, matrix size = 288×288, slice thickness = 0.9 mm, slice gap = 0 mm, TR = 6.7 ms, TE = 3.1 ms.

2.2.4 MRI pre-processing

All images were preprocessed using SPM12 (University College London, London, UK). Prior to processing, all scans were automatically approximately reoriented to a canonical SPM template, the MNI single subject T1 image, using a custom MATLAB function, auto_reorient.m – based on the same-named function created by Carlton Chu. All scans were then visually inspected for good orientation and gross artefacts before preprocessing. In the training set, badly oriented scans (n = 632), or scans with gross artefacts (n = 42), were excluded from further analysis (see Table S1 in 7.1.1 Supplemental Methods for information on exclusions from each open-access dataset). It should be noted that the majority of the excluded scans were due to poor re-orientation and although this reduced the training set sample size, this eliminated the need for time-consuming manual re-orientation of individual images. To prevent data loss in the test set, any badly oriented scans were manually re-oriented before preprocessing but scans with artefacts were still excluded. In both training and test sets, each individual dataset was preprocessed in a separate batch. Bias correction was applied to image which were then segmented into GM, WM, and cerebrospinal fluid (CSF). Segmented GM images were non-linearly registered to a custom template, using SPM's *DARTEL*. Images were then affine registered to MNI space (1 mm^3) and resampled with modulation to preserve the total amount of signal from each voxel. Images were smoothed with a 4 mm full-width at half maximum Gaussian kernel. Finally, images were visually inspected for accurate segmentation. The code used to auto-reorient and preprocess the MRI data is available at https://github.com/rorytboyle/brainPAD.

2.2.5 Machine learning

Data preparation

As a simple data reduction step, GM images were resized to 2 mm^3 voxels and individual voxel values were extracted from each image. A threshold was applied to exclude voxels with a low probability of reflecting GM such that a voxel was retained if it had GM density > 0.2 in that voxel across all 1,359 training set images. 2 mm^3 voxels were selected based on a balance between computational efficiency and predictive accuracy (see 7.1.1 Supplemental Methods: Choice of voxel size and Table S2) and on

its use in previous studies (Daniels et al., 2015; Hanssen et al., 2018; Seubert et al., 2013; Sowman et al., 2017). Although a lower threshold can result in greater accuracy, it also greatly increases the computational expense of the model as well as the probability of including non-GM information in the model (see 7.1.1 Supplemental Methods: Choice of GM threshold and Table S3). After thresholding, the training data consisted of 1,359 images, each with 54,869 voxels.

Machine learning model

The goal of the training phase was to construct a generalisable model that could predict chronological age from GM data. The Elastic Net was applied in the present study as it is particularly well-suited for data with a much larger number of predictors than observations, such as neuroimaging data. The Elastic Net combines the Least Absolute Shrinkage and Selection Operator (LASSO) regression, where regression weights are penalised for increasing model complexity based on their absolute size and can be set to zero, and ridge regression, where regression weights are penalised for increasing model complexity based on their squared values and as such cannot be set to zero (Zou & Hastie, 2005).

Each nested training set was divided into 10 cross-validation (CV) folds, each consisting of 10% of the subsampled training set. Nine folds were then used to create the regression model and the model's prediction were then tested on the one left-out fold. This entire procedure was repeated 10 times, with each CV fold being left-out once. Furthermore, within each fold, nested cross-validation with 10 partitions was then used for optimisation of the two Elastic Net model parameters: alpha (*α)*, which is the weight of the lasso vs. ridge regularization and lambda (*λ),* which is the regularization coefficient*.* Thirty values of each parameter were used, with *α* parameters ranging from 1e-25 to 1 and *λ* ranging from 1 to 1e-04. The most frequently occurring parameter values across nested CV folds were used to create the final prediction model for each CV fold.

In order to increase generalizability of the model, a data resampling ensemble approach was used. That is, 500 participants, with a 50:50 gender ratio, were randomly sampled without replacement from the training data to form a nested training set. This process was repeated 25 times, creating 25 nested training sets. Each nested training set (500 participants x 54,869 voxels), was used as the input to a regularised linear regression model (Elastic Net), with 10-fold cross-validation (CV), to predict the chronological age of each participant. The data resampling ensemble approach controlled for the effect of sex and reduced any possible individual model effects.

The performance of the model was quantified using the mean of each of the 25 nested models' Pearson's correlation between chronological age and predicted age (*r*), total variance explained (R^2), mean absolute error (MAE), and the weighted MAE. The weighted MAE is equal to the MAE divided by the age range of the sample tested and is a more suitable metric for comparing the MAE of brainPAD models across studies as it accounts for the impact of a sample's age range on prediction accuracy (Cole et al., 2019). A lower weighted MAE reflects greater accuracy.

Application to independent test sets

First, the average coefficient value for each voxel across all folds in all 25 training models was calculated, resulting in a vector of length 54,869. For each independent test set, the mean coefficient values were multiplied by the voxels' GM density values and the product was summed to create a brain-age prediction for each participant. To correct for the proportional bias in the model, the prediction was added to the intercept of the training set, and the result was then divided by the slope of the training set. This correction does not affect the relationship between brainPAD and outcome measures but scales the data correctly so that brainPAD scores can be interpreted in units of years proportional to a person's chronological age. Similar corrections have been applied in other brainPAD models (Cole et al., 2018). BrainPAD was calculated by subtracting chronological age from the corrected predicted age, hence, a positive brainPAD value indicates a brain-predicted age that exceeds the participant's chronological age, suggesting accelerated brain ageing. The code used to make brain-age predictions and calculated brainPAD scores for independent test sets is available at https://github.com/rorytboyle/brainPAD.

2.2.6 Cognitive function measures

Each of the three datasets contained a wide range of cognitive measures. However, as the three datasets were completely independent of one another, and all data collection was completed prior to conception and design of the present study, different cognitive measures were used across the datasets. For the purposes of the present study, a cognitive measure was selected for analysis if it assayed a cognitive domain that was assessed in at least one other dataset. For example, the Psychomotor Vigilance Test (PVT) and Sustained Attention to Response Task (SART) assessed sustained attention in CR/RANN and TILDA respectively so both measures were selected for analysis and considered as 'comparable' measures. The cognitive domains assessed by each measure were decided with reference to the literature. Across all three datasets, 17 common cognitive domains were identified (see Table 2.1 for list of cognitive domains and cognitive measures across datasets).

General Cognitive Status: Total scores on the Mini-Mental State Examination (MMSE; Folstein, Folstein and McHugh, 1975) in DEU and TILDA, and on the Mattis Dementia Rating Scale-2 (DRS; Jurica, Leitten, & Mattis, 2001) in CR/RANN were used to assess general cognitive status. Both scales assess cognitive functioning across domains which are typically affected by Alzheimer's disease and cognitive impairment (Monsch et al., 1995; Tombaugh & McIntyre, 1992).

Verbal Intelligence: Raw scores on the American National Adult Reading Test (AMNART; Grober and Sliwinski, 1991) in CR/RANN and on the National Adult Reading Test (NART; Nelson and Willinson, 1982) in TILDA were used to assess verbal intelligence. In TILDA, 60.57% of those with NART scores completed the full NART whereas 39.22% completed the first half of the NART only. Participants only proceeded to the second half of the NART if they scored over 20 on the first half. This is both a time-saving measure and serves to reduce distress and anxiety in people with poor reading skills (Strauss et al., 2006). Scores of 0-11 were used as full scores but scores of 12-20 were corrected using a conversion table outlined by Beardsall and Brayne (1990). There was no comparable measure of verbal intelligence in DEU.

Phonemic Verbal Fluency: Phonemic verbal fluency was assessed using the total score on a Turkish language version of the FAS test in DEU, the KAS test (Tumac, 1997), and on the CFL test in CR/RANN. These tests measure the ability to spontaneously produce words beginning with specific letters (i.e., 'K' in KAS or 'C' in CFL; Strauss, Sherman and Spreen, 2006). There was no comparable measure of phonemic verbal fluency in TILDA.

Semantic Verbal Fluency: Semantic verbal fluency was assessed using the total score on the Animals test in all three datasets. A Turkish language version of this task was used in DEU (Tumac, 1997). This test measures the ability to spontaneously produce the name of animals (Strauss et al., 2006).

Processing Speed: Cognitive processing speed was assessed using time to complete the Trail Making Test A (TMT; Reitan, 1955) in DEU and CR/RANN, and the Colour Trails Task 1 (CTT; D'Elia et al., 1996) in TILDA. The CTT is considered a cross-culturally valid form of the TMT (Strauss et al., 2006).

Executive Function: Executive function was assessed using time to complete the TMT B in DEU and CR/RANN and the CTT 2 in TILDA. These measures both also involve processing speed (Strauss et al., 2006).

Executive Function (without Processing Speed): A purer measure of executive function was obtained by subtracting the simpler TMT A and CTT 1 from the more complex TMT B and CTT 2, respectively. This difference score was calculated in all 3 datasets and controls for general processing speed (Strauss et al., 2006).

Cognitive Flexibility: Cognitive flexibility was measured by the percentage of perseverative errors on the Wisconsin Card Sorting Task (WCST; Heaton, Chelune, Talley, Kay, & Curtiss, 1993) in DEU and CR/RANN. Cognitive flexibility is considered a specific process underlying executive function (Logue & Gould, 2014) and is not associated with processing speed (Liozidou et al., 2012), unlike the TMT B or CTT 2.

Response Inhibition and Selective Attention: The Stroop test in DEU and CR/RANN was used to assess response inhibition and selective attention. A Turkish version of the Stroop test, the Stroop Test Çapa Version, was used in DEU (Emek-Savaş et al., 2020) and the measure used was the resistance to interference in seconds as calculated by subtracting the time taken to read the colour names from the time taken to name the colour ink of written colour names. The Golden version of the Stroop test (Golden, 1978) was used in CR/RANN and number of words completed in 45 seconds on the Color-Word page was used as the measure. There was no comparable measure of response inhibition and selective attention in TILDA.

Sustained Attention: Sustained attention was assessed with the PVT (Dorrian, Rogers and Dinges, 2005) in CR/RANN using the number of false alarms (i.e., errors of commission) and the median reaction time across trials with an inter-trial interval of two to four seconds. It was assessed with the SART (Robertson et al., 1997) in TILDA using the number of errors of commission and the coefficient of variation in reaction time as measures. There was no comparable measure of sustained attention in DEU.

Verbal Episodic Memory (Immediate): Immediate verbal episodic memory was assessed in DEU with the immediate recall score from the Öktem Verbal Memory Processes Test (OVMPT; Öktem, 1992) which is a validated Turkish version of the Rey Auditory Verbal Learning Test (RAVLT; Bosgelmez et al., 2015) and measures the number of words immediately recalled from a 15-item word list. The CR/RANN measure was the total recall score on the Selective Reminding Test (SRT; Buschke & Fuld, 1974), which measures the total number of words recalled from 6 trials of a 12-item word list (Strauss et al., 2006). This was assessed in TILDA using the average number of words immediately recalled from 2 trials of a 10-item word list as used originally in the Health and Retirement Study (Wallace & Herzog, 1995).

Verbal Episodic Memory (Delayed): Delayed verbal episodic memory was assessed in DEU using the delayed recall score from the OVMPT which consisted of the number of words recalled from the 15-word list after a 40 minute delay, and in CR/RANN using the delayed recall score from the SRT which consisted of the number of words recalled from the 12-item word list after an approximate 15 minute delay. It was assessed in TILDA using the number of words recalled after an approximate 20-25 minute delay from a 10-item word list (depending on length of time it took participants to complete intervening items). The TILDA measure was taken as the average score over two trials.

Verbal Episodic Memory (Learning): Verbal episodic memory learning was assessed in DEU using the OVMPT total learning score which was the total number of words recalled in each trial and in CR/RANN using the consistent long-term retrieval score on the SRT which was the number of words consistently recalled on all subsequent trials (Strauss et al., 2006). There was no comparable measure of verbal episodic memory learning in TILDA.

Working Memory: Working memory was assessed in DEU using both the Digit Span Forward and Digit Span Backward tests from the Wechsler Memory Scale – Revised Edition (WMS-R; Wechsler, 1987). In CR/RANN, the Letter-Number Sequencing test from the Wechsler Adult Intelligence Scale – Third Edition (WAIS-III; Wechsler, 1997) was used. There was no comparable measure of working memory in TILDA.

Visuospatial Ability: Visuospatial ability was assessed in DEU using the Judgement of Line Orientation Test (BLOT; Benton, Varney, & Hamsher, 1978) which measures participants' capacity to discriminate the direction of lines. In CR/RANN, the Block Design test from the WAIS-III, which measures ability of participants to replicate models or pictures presented to them using blocks (Strauss et al., 2006), was used. There was no comparable measure of visuospatial ability in TILDA.

Table 2.1 Cognitive measures available across each dataset in comparable cognitive domains.

Cognitive Domain(s)	DEU Measure *(N)*	CR/RANN Measure *(N)*	TILDA Measure *(N)*
General Cognitive Status	MMSE *(172)*	DRS *(370)*	MMSE *(485)*
Verbal Intelligence	n/a	AMNART *(362)*	NART *(486)*
Phonemic Verbal Fluency	KAS Test *(137)*	CFL Test *(360)*	n/a
Semantic Verbal Fluency	Animals Test *(175)*	Animals Test *(361)*	Animals Test *(487)*
Processing Speed	TMT A *(93)*	TMT A *(361)*	CTT 1 *(487)*
Executive Function	TMT B *(84)*	TMT B *(357)*	CTT 2 *(482)*
Executive Function (without Processing Speed)	TMT B minus TMT A *(84)*	TMT B minus TMT A *(357)*	CTT 2 minus CTT 1 *(482)*
Cognitive Flexibility	WCST Perseverative Errors *(50)*	WCST Perseverative Errors *(327)*	n/a
Response Inhibition, Selective Attention	Stroop[16] Interference Score - Time *(150)*	Stroop[17] Interference Score - Words *(359)*	n/a
Sustained Attention (Errors of Commission)	n/a	PVT False Alarms *(176)*	SART Errors of Commission *(482)*
Sustained Attention (RT)	n/a	PVT Median RT *(176)*	SART CV RT *(479)*
Verbal Memory (Immediate)	OVMPT Immediate Recall *(175)*	SRT Total Score *(360)*	Immediate Recall *(487)*
Verbal Memory (Delayed)	OVMPT Delayed Recall *(175)*	SRT Delayed Recall *(360)*	Delayed Recall *(487)*
Verbal Episodic Memory (Learning)	OVMPT Total Learning Score *(175)*	SRT Consistent Long Term Retrieval *(360)*	n/a
Working Memory	WMS-R Digit Span Forward *(171)*	WAIS-III Letter Number Sequencing *(360)*	n/a
	WMS-R Digit Span Backward *(170)*		
Visuospatial Ability	BLOT *(80)*	WAIS-III Block Design *(356)*	n/a

Note: MMSE = Mini-mental state examination (Folstein et al., 1975)*; DRS Total Score = Mattis Dementia Rating Scale-2 – Total Score* (Jurica et al., 2001)*; NART = National Adult Reading Test* (Nelson & Willinson, 1982)*; AMNART = American National Adult Reading Test* (Grober & Sliwinski, 1991)*; CTT = Colour Trails Test* (D'Elia et al., 1996)*; TMT = Trail Making Test* (Reitan, 1955)*; WCST = Wisconsin Card Sorting Test* (Heaton et al., 1993)*; SART = Sustained Attention to Response Test* (Robertson et al., 1997)*; PVT Median RT = Median Reaction Time on Psychomotor Vigilance Task* (Dorrian et al., 2005)*; SART CV RT = SART Coefficient of Variation of Reaction Time OVMPT = Öktem Verbal Memory Processes Test* (Öktem, 1992)*; SRT = Selective Reminding Test* (Buschke & Fuld, 1974)*; WMS-R = Wechsler Memory Scale* (Wechsler, 1987)*; WAIS-III = Wechsler Adult Intelligence Scale – Third Edition* (Wechsler, 1997)*; BLOT = Benton's Judgement of Line Orientation Test* (Benton et al., 1978)

[16] Turkish Capa version of Stroop test (Emek-Savaş et al., 2020)
[17] Golden version of Stroop test (Golden, 1978)

2.2.7 Statistical Analysis

The statistical analysis was conducted using the following procedure:

1. *Correlate.* Within each independent test set, partial Spearman's rank order correlations were conducted between brainPAD scores and cognitive measures, controlling for chronological age and sex. Sex was adjusted for to account for a significant sex difference in brainPADs ($p < 0.0001$).

2. *Replicate.* For findings replicated in multiple datasets, the probability of obtaining statistically significant p-values across multiple datasets by chance was calculated by random-label permutation (Good, 1994). Within each dataset, brainPAD scores were randomly shuffled using the *randperm* function in MATLAB. Spearman's partial correlations were then conducted between randomly shuffled brainPAD scores and the cognitive dependent variables, controlling for age and sex. This process was repeated one million times. The number of iterations in which the random p-values were more extreme (i.e., smaller) than the actual p-values across datasets was summed and divided by one million to obtain the probability of the finding replicating across multiple datasets by chance. Replicated findings were deemed significant if this probability was less than .05. For example, if a finding was replicated across three datasets, the replication p-value obtained by the random-label permutation test represented the probability of observing correlations larger than the true correlations across all three datasets.

3. *Correct for multiple comparisons.* All other correlations were then corrected for multiple comparisons, or the familywise error rate within individual datasets, while allowing for correlations among dependent cognitive variables, using a maximum statistic approach. In each test set, brainPAD scores were randomly shuffled using the *randperm* function in MATLAB. Spearman's partial correlations were conducted between the randomly shuffled brainPAD scores and the cognitive dependent variables, controlling for age and sex. This process was repeated one million times and the maximum rho value was stored each time. Correlations between actual brainPAD scores and cognitive variables were deemed significant if they exceeded the 95th percentile of the maximum rho values. This approach was a variant of max T adjustments (Dudoit et al., 2003) and was used because Bonferroni and Šidák adjustments can be overly conservative when there are correlated dependent measures (Conneely & Boehnke, 2007; Dudoit et al., 2003).These analyses were

conducted in MATLAB and the code used to perform steps 2 and 3 of the above procedure is available here: https://github.com/rorytboyle/brainPAD_dataAnalysis.

2.3 Results

2.3.1 Brain age prediction

Training set

The model accurately predicted chronological age (r = 0.85, R^2 = 67.24%, MAE = 7.28 years, weighted MAE = 0.10, p < 0.0001) in the training set. As with other brainPAD models (e.g., Cole et al., 2018), a proportional bias was observed in this model where chronological age correlated with prediction error (r = - 0.4452, p = 1.1036e-10).

Independent test sets

Application of the coefficients to all three test sets combined resulted in a statistically significant correlation between brain-predicted age and chronological age of r = 0.73 (p < 0.0001), which explained 52.79% of the variance (R^2). The combined test set had a mean chronological age of 62.75 (SD = 14.3) years at the time of scanning and mean brain-predicted age of 62.94 (SD = 13.3) years. Mean brainPAD was +0.18 (SD = 10.25) years. MAE was 8.33 years and weighted MAE was 0.112 years. Application of the coefficient to each individual test set similarly resulted in statistically significant correlations between brain-predicted age and chronological age (see Table 2.2).

Table 2.2 Results of application of trained model parameters to independent test sets.

Test Set	r	Mean brainPAD (SD)	MAE	Weighted MAE
Test Set 1 – DEU	0.78*	+6.60 (6.44)	7.60	0.17
Test Set 2 – CR/RANN	0.87*	+6.39 (8.57)	8.56	0.14
Test Set 3 - TILDA	0.65*	-6.97 (7.52)	8.42	0.22

*Note: * = p < 10-37. R = Pearson's r between brain age and chronological age, Weighted MAE = MAE divided by age range.*

2.3.2 Sex differences in brainPAD

Mean brainPAD differed significantly by sex in all datasets, Welch's t(1009.55) = - 5.79, p < .0001. Males (M = -1.81, SD = 9.92) had significantly lower brainPADs than females (M = 1.81, SD = 10.23; see Fig. 2.1). Within individual test sets, males had significantly lower brainPADs, compared to females, in CR/RANN (p < .0001) and TILDA (p < .0001) but not in DEU (p = 0.148; see Fig. 2.2).

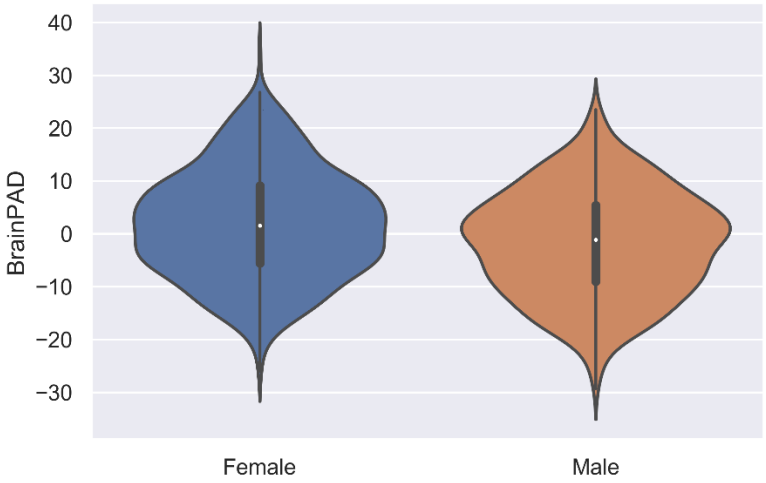


Figure 2.1. Violin plot comparing distributions of brainPADs between sexes across all datasets.

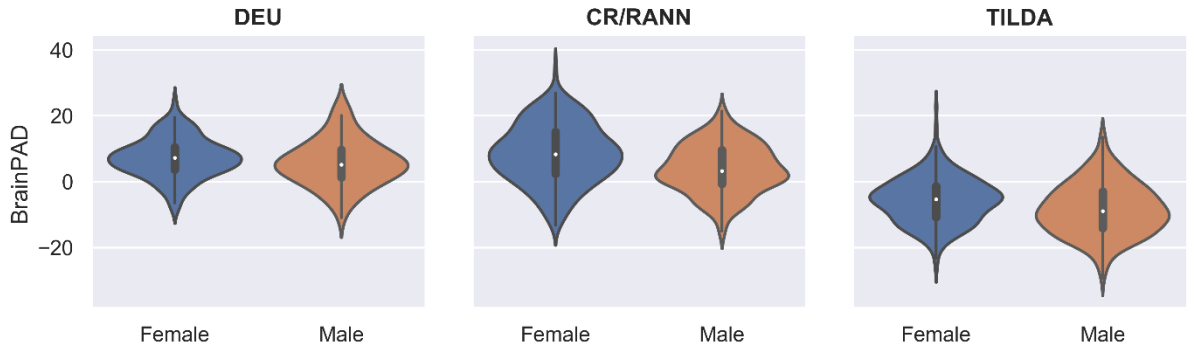


Figure 2.2. Violin plots comparing distributions of brainPADs between sexes within datasets.

2.3.3 Brain regions involved in brain age prediction

The voxel-wise method used here to predict brain age resulted in individual coefficient values for each voxel, where voxels with positive values contributed to older brain age predictions and voxels with negative values contributed to younger brain age predictions (see Fig. 2.3). As it is difficult to visualise a 3D object containing many small voxels, a .nii file of the regression coefficients can be downloaded here: https://osf.io/5n6t8/ and used as an overlay in a viewer such as mricroGL to obtain a better visualization. Moreover, an .xlsx file containing the absolute value and sign direction (i.e., positive/negative) of each regression coefficient, the coefficient rank in terms of absolute values, as well as the MNI coordinates and anatomical labels of the coefficients, is available here: https://osf.io/dkz67/.

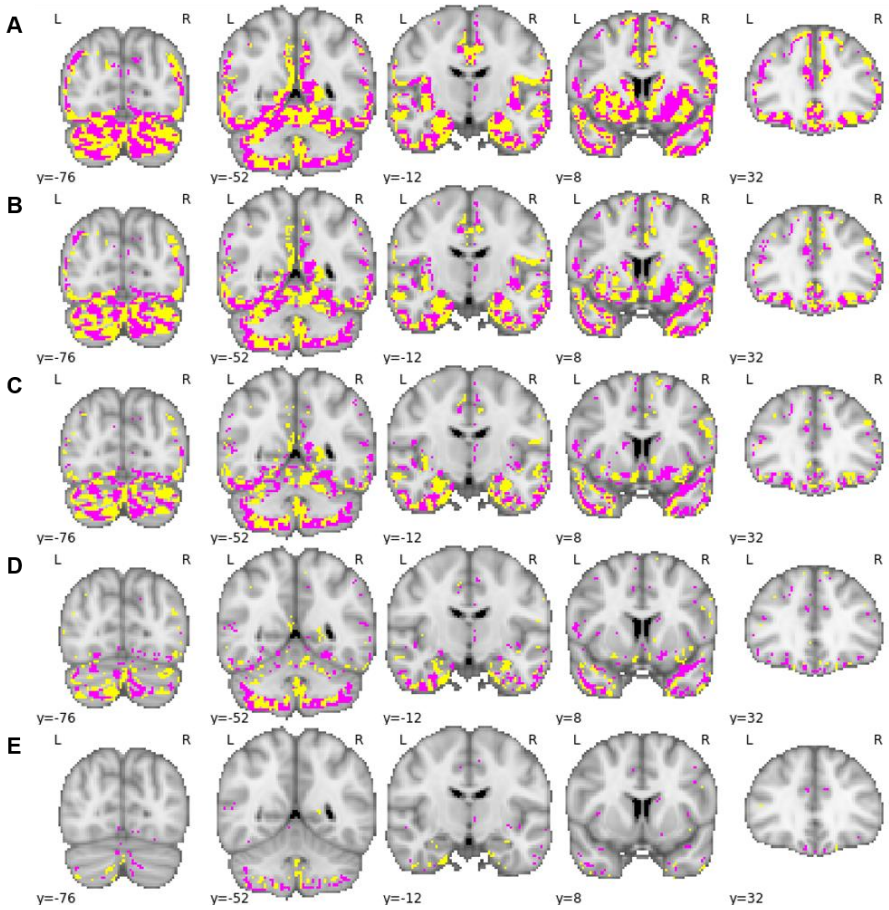


Figure 2.3. Binarised regression coefficients overlaid on 5 coronal slices of the brain. *Positive coefficients shown in pink, negative coefficients shown in yellow. A: No threshold applied; B: thresholded at 25th percentile of absolute value of regression coefficients; C: thresholded at 50th percentile of absolute value of regression coefficients;*

D: thresholded at 75th percentile of absolute value of regression coefficients; E: thresholded at 95th percentile of absolute value of regression coefficients.

2.3.4 BrainPAD and Cognitive Function

BrainPAD was negatively correlated with performance in specific cognitive domains across multiple datasets (see Fig. 2.4 and Table 2.3). The replication of associations between brainPAD and cognition was statistically significant in four cognitive domains: general cognitive status (DEU and CR/RANN, replication p-value = 0.000002), semantic verbal fluency (DEU and CR/RANN, $p < 0.000001$), executive function (DEU, CR/RANN, TILDA, $p = 0.000054$), and executive function without processing speed (DEU and CR/RANN, $p = 0.000966$).

BrainPAD was negatively associated with performance, in a single dataset, in four further cognitive domains: phonemic verbal fluency (DEU, rho = -0.326, $p = 0.0001$), verbal intelligence (CR/RANN, rho = -0.2322, $p < 0.00001$), verbal episodic memory – learning (DEU, rho = -0.3196, $p < 0.00001$), and visuospatial ability (CR/RANN, rho = -0.1824, $p = 0.0006$). These associations were not replicated but survived correction for multiple comparisons within a single dataset.

BrainPAD was not significantly associated with performance in any dataset, in seven cognitive domains: processing speed, response inhibition and selective attention, working memory, verbal episodic memory – immediate, verbal episodic memory – delayed, sustained attention, and cognitive flexibility (see Fig. S1 in 7.1.2 Supplemental Results for scatterplots of the brainPAD-cognition associations in each dataset that were either non-replicated or non-significant).

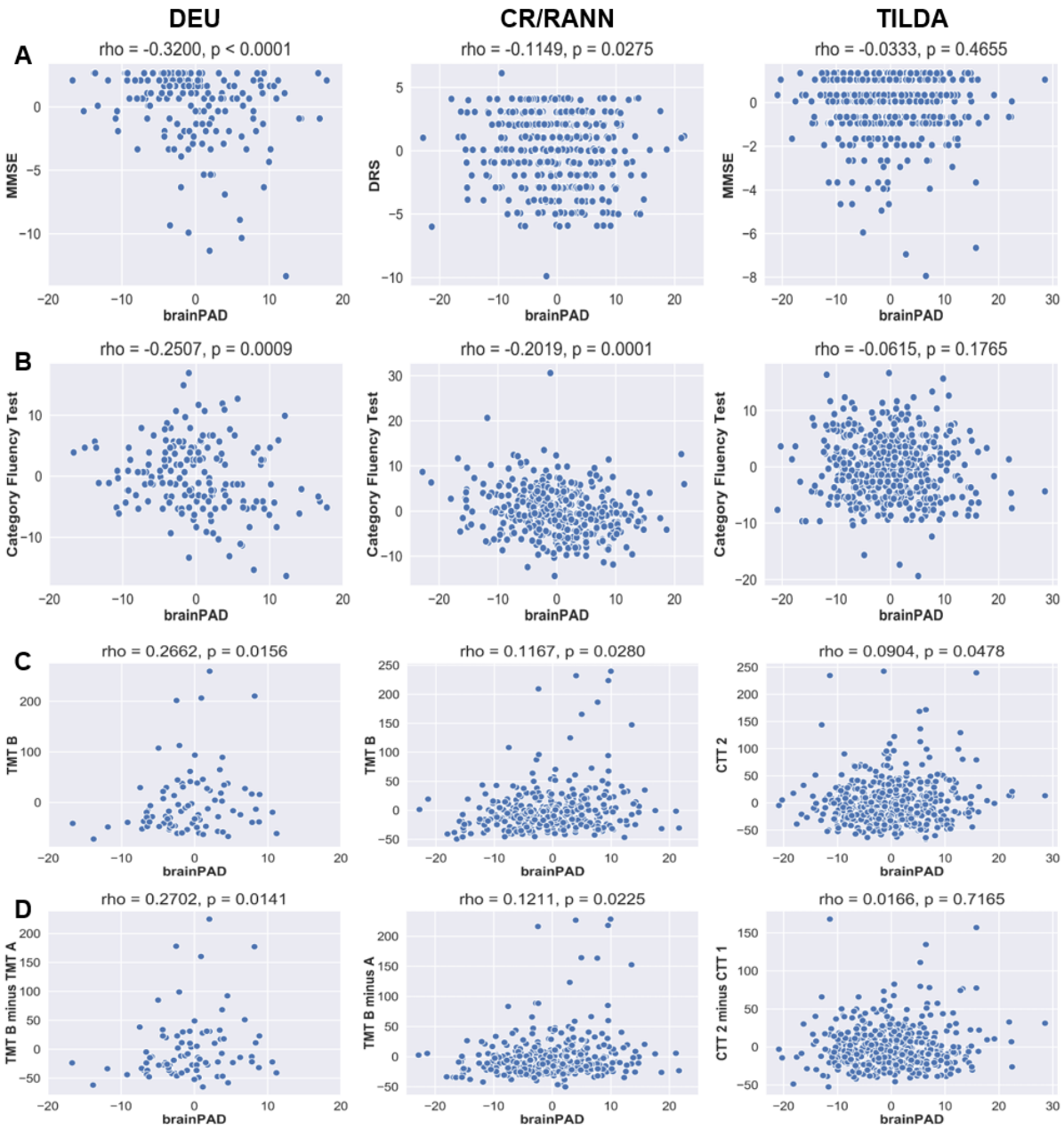


Figure 2.4. Scatterplots of replicated correlations between the residuals of brainPAD and cognitive measures after regressing brainPAD on age and sex, and each cognitive measure on age and sex. *A: General cognitive status; B: Semantic verbal fluency; C: Executive Function; D: Executive Function (without Processing Speed).*

Table 2.3 Results of Spearman's partial correlations between brainPAD and 17 cognitive domains, adjusting for age and sex.

Cognitive Domain	DEU			CR/RANN			TILDA			Replication
	rho	*df*	*p*	*rho*	*df*	*p*	*rho*	*df*	*p*	*p*
General Cognitive Status	-0.3199	168	<0.0001	-0.1449	366	0.0275	-0.0333	481	0.4655	<0.00001
Verbal Intelligence		n/a		-0.2322	358	<0.0001*	0.0485	482	0.2873	n/a
Phonemic Verbal Fluency	-0.3259	134	0.0001*	-0.0771	356	0.1454		n/a		n/a
Semantic Verbal Fluency	-0.2507	171	0.0009	-0.2019	357	0.0001	-0.0615	483	0.1765	<0.00001
Processing Speed	0.1232	89	0.2448	0.0595	357	0.2610	0.1208	483	0.0077	n/a
Executive Function	0.2662	80	0.0156	0.1167	353	0.0279	0.0904	478	0.0478	0.00005
Executive Function (without Processing Speed)	0.2702	80	0.0141	0.1211	353	0.0225	0.0166	478	0.7165	0.00097
Cognitive Flexibility	0.0722	46	0.6258	0.0429	323	0.4411		n/a		n/a
Response Inhibition, Selective Attention	0.0854	146	0.3019	-0.1755	355	0.0009		n/a		n/a
Sustained Attention (Errors of Commission)		n/a		0.0203	172	0.7902	0.0499	478	0.2752	n/a
Sustained Attention (Reaction Time)		n/a		-0.0212	172	0.7813	0.0436	475	0.3425	n/a
Verbal Episodic Memory (Immediate)	0.2194	171	0.0037	-0.0407	356	0.4428	-0.0347	483	0.4114	n/a
Verbal Episodic Memory (Delayed)	0.2797	171	0.0002	0.0343	356	0.5173	0.0122	483	0.7887	n/a
Verbal Episodic Memory (Learning)	-0.3196	171	<0.0001*	0.0657	356	0.2151		n/a		n/a
Working Memory	-0.1310	167	0.0895[a]	-0.0469	360	0.3759		n/a		n/a
	-0.2974	166	0.0001[b]							
Visuospatial Ability	-0.0809	76	0.4815	-0.1824	352	0.0006*		n/a		n/a

*Note: Replication p-value reflects the probability of association between brainPAD and cognition replicating across datasets by chance. * = indicates associations that were not replicated across datasets but were statistically significant in a single dataset after correcting for multiple comparisons by applying a maximum statistic correction; [a] Digit Span Forwards; [b] Digit Span Backward.*

2.4 Discussion

A penalised regression approach produced accurate brain-age predictions from T1 MRI data in three independent datasets. In non-demented adults, brainPAD, calculated by subtracting these brain-predicted ages from chronological age, was negatively correlated with general cognitive status, semantic verbal fluency, executive function; and executive function (without processing speed) in multiple datasets. These robust associations between brainPAD and cognition demonstrated that brainPAD is a valid operational measure of BM. However, while brainPAD was significantly correlated with phonemic verbal fluency, verbal intelligence, verbal episodic memory (learning score), and visuospatial ability in single datasets after controlling for multiple comparisons; these correlations were not replicated in another dataset. As such, there is not strong evidence here in support of these relationships. BrainPAD was not significantly correlated with processing speed, cognitive flexibility, response inhibition and selective attention, sustained attention, verbal episodic memory (immediate recall or delayed recall), or working memory in any dataset. These non-robust and/or non-significant associations demonstrate that the relationship between brainPAD and cognition is specific to particular cognitive domains.

The replicated associations between lower brainPADs and better performance on measures of general cognitive status, semantic verbal fluency, executive function, and executive function (without processing speed), are supported by previous associations with these measures. BrainPAD has been previously associated with general cognitive status in non-cognitively healthy samples (Beheshti et al., 2018; Kaufmann et al., 2019) and in studies without statistically controlling for multiple comparisons or the effects of age (Beheshti et al., 2018; Cole, Underwood, et al., 2017). BrainPAD also been previously associated with semantic verbal fluency (Franke et al., 2013). Furthermore, brainPAD was previously associated with executive function (with and without the influence processing speed, i.e., TMT B and TMT B minus A) in non-cognitively healthy samples (Cole et al., 2015) and in studies without correcting for multiple comparisons (Cole, Underwood, et al., 2017). These findings clarify the literature in relation to the association of brainPAD and specific domains of cognitive function (see Chapter 5, section 5.2 for further discussion of these findings in relation to the literature).

Together, the replicated associations between brainPAD and general cognitive status, semantic verbal fluency, executive function, and executive function (without processing speed) demonstrate that brainPAD is a valid operational measure of BM, as relatively better preservation of brain integrity, or structural health, is associated with

better cognition. However, this association between brainPAD and cognition is not global as robust associations were not observed for several other cognitive domains, including cognitive flexibility, phonemic verbal fluency, processing speed, response inhibition and selective attention, sustained attention, verbal episodic memory, verbal intelligence, and visuospatial ability.

It is notable that several significant brainPAD-cognition relationships were replicated across the DEU and CR/RANN datasets, but not in TILDA. Nonetheless, given the statistically significant replication across two of the three datasets, there is reliable evidence in support of the correlation between brainPAD and these specific cognitive domains in healthy older adults. Tentative suggestions for this pattern of results include differences in confounds, age range, and range of cognitive performance across datasets. Confounding factors obscuring the brainPAD-general cognitive status relationship may have been uniquely present in TILDA. Whereas the DEU and CR/RANN cohorts were part of neuroimaging research studies, which have typically strict inclusion criteria, the TILDA MRI sample were a subset of a large nationally representative longitudinal study encompassing health, economic and social research (B. J. Whelan & Savva, 2013). TILDA therefore had few inclusion criteria: being at least 50 years old, having a residential address, and absence of dementia at baseline (Kearney et al., 2011; Savva et al., 2013). TILDA's MRI sample were screened for MRI contraindications and were on average healthier than the full sample, but it is likely that the TILDA sample included participants who might normally be excluded from neuroimaging research studies (e.g., those using psychotropic or cardiovascular medication). Moreover, the range of some cognitive measures in TILDA was also smaller than DEU and CR/RANN in some cases (see Table S4 in 7.1.2 Supplemental Results) notably for general cognitive status, and executive function (without processing speed), where the brainPAD-cognition correlations were not replicated within TILDA. Restricted range of scores on these measures in TILDA may have contributed to smaller correlation coefficients (Bland & Altman, 2011; Mendoza & Mumford, 1987). Additionally, the age range within TILDA was smaller than both DEU and CR/RANN which may have reduced the statistical power of the brainPAD-cognition correlations within TILDA as range restriction on covariates has also been shown to reduce power (Miciak et al., 2016) and decrease the magnitude of correlation coefficients (Sackett & Yang, 2000).

The smaller age range within TILDA (38 years) as compared to DEU (45.95 years) and CR/RANN (61 years) might also have contributed to the weaker correlation between chronological age and brain-predicted age in TILDA (see Table S4. in 7.1.2 Supplemental Results), as *range restriction* will reduce the size of correlation coefficients

(Goodwin & Leech, 2006). Moreover, a negative mean brainPAD was reported in TILDA (-6.97 years) whereas both DEU and CR/RANN had positive mean brainPADs, +6.6 and +6.39 years respectively. Various factors, including intelligence, educational attainment, and environmental factors, have been proposed to affect brain ageing (Irimia et al., 2015). TILDA had significantly higher levels of education versus both CR/RANN and DEU (see 7.1.2 Supplemental Results: Differences in Education across samples). Steffener and colleagues (2016) reported that brainPAD was significantly related to education, with higher education associated with younger brains (or smaller/more negative brainPADs). This association with education could be one reason why lower mean brainPADs were observed for the TILDA dataset. As the cohorts are each from different countries, there could be various other environmental factors that could further explain this relationship.

An interesting finding was that there were significantly higher mean brainPADs in females in two of the three datasets in this study (TILDA and CR/RANN). There is mixed evidence in relation to sex differences in other brain age prediction studies, with some studies reporting significantly higher mean brainPADs in males (Cole et al., 2018; Franke et al., 2013; Luders et al., 2016), some reporting no significant sex differences (Azor et al., 2019; Cruz-Almeida et al., 2019; Franke et al., 2014; Han et al., 2019), and another study, with a notably large sample size of 19,000, reporting higher mean brainPADs in females (Smith et al., 2019). Even studies using the same training sets have contrasting results in terms of sex effects. For example, one training set reported significantly higher male brainPADs in two studies (Franke et al., 2013; Luders et al., 2016) but no sex differences in another study (Franke et al., 2014). However, this divergence could have arisen due to the likely mean centering of both brainPADs in both sexes in the latter study (i.e., male and female groups had mean brainPADs of 0 years). This was also the case in another training set used in multiple studies, with one study reporting significantly higher brainPADs in males (Cole et al., 2018) but another reporting no significant differences (Azor et al., 2019). As such, it is likely that sex differences in brainPAD reflect the characteristics of the test sample. This is apparent in the present study with two out of the three datasets showing higher mean brainPADs in females but one dataset showing no significant differences.

Importantly, the brain age prediction model used here to evaluate the associations between brainPAD and cognition was accurate and generalisable. The model was evaluated based on its predictive accuracy in three independent test sets, as proposed by Madan and Kensinger (2018). While internal cross-validation is a valuable and widely used technique that can attenuate overfitting (Arlot & Celisse, 2010); the use

of cross-validation in certain situations and when it is not implemented correctly, can result in overestimated prediction accuracy and overfitting (Saeb et al., 2016; Skocik et al., 2016; Varoquaux et al., 2017). For brainPAD to be considered for clinical use, it must perform accurately with MRIs acquired in different scanners and under different protocols. However, in most instances of cross-validation, while the test set is split and held completely independent from the training set, factors common to both sets, such as scanner and protocol, could influence model performance. The evaluation of model performance on external test sets is a stronger test of generalisation (Gillan & Whelan, 2017). As such, the gold-standard evaluation for brainPAD should be accurate performance on independent external datasets. The significant correlations between chronological age and brain-predicted age in all three external datasets shows that the model is accurate and generalisable (0.65, 0.78, and 0.87 for external datasets). Although the magnitude of these correlations is lower than correlations reported elsewhere, ranging from 0.91 to 0.94 (Cole et al., 2015; Cole, Poudel, et al., 2017; Franke et al., 2010; Lancaster et al., 2018; Liem et al., 2017), it exceeds other externally validated brain-predicted age studies, ranging from 0.65 to 0.85 (Beheshti et al., 2018; Madan & Kensinger, 2018; Varikuti et al., 2018).

With respect to MAE, the model did not perform as well as other externally validated studies, ranging from 4.28 to 7.5 years (Beheshti et al., 2018; Cole et al., 2018; Franke et al., 2010; Lancaster et al., 2018; Madan & Kensinger, 2018). As a result, it could be possible that the model may have lost some precision by not integrating WM information as input, as was done by Cole et al. (2018), for example. Another potential reason is that other studies centred the age predictions using the mean of the ages from the test set. Although this correction is typically not explicitly described in method sections, Madan and Kensinger (2018) note that this is a standard correction in brain age prediction. Moreover, some studies also match the variance in predicted age in the test set with the variance of the training data (Madan & Kensinger, 2018). Both corrections are principled and acceptable methods of correcting for the regression to the mean artefact in brain age predictions but they result in biased age predictions in the test set. These corrections also limit the use of brainPAD to make single subject predictions, as both the test set mean and variance are used in the prediction. The method in the present study used only training set information and therefore produced slightly less accurate but less biased predictions. Finally, the model may also appear to be less precise in terms of MAE as an artefact of the greater age range of the samples assessed here in comparison to most brainPAD studies. An alternative metric, the weighted MAE (calculated by dividing the MAE by the age range of the sample), may

enable better comparisons across studies with different age ranges (Cole et al., 2019). While the weighted MAE is higher than some studies, ranging from 0.072 to 0.087 (Lancaster et al., 2018; Liem et al., 2017), the lowest weighted MAE in the present study (0.14 in CR/RANN) outperformed this metric when calculated for other studies, 0.178 (Beheshti et al., 2018), and 0.18 (Varikuti et al., 2018) and is comparable to 0.139 (Franke et al., 2010, 'Test 4' external test set). As such, the predictive accuracy of the model is comparable to the rest of the literature and is arguably less biased as only training set information is used.

In addition to the good accuracy and generalisability observed for the brain age prediction model here, the model was interpretable. The interpretability of machine learning models is an important and widely discussed problem (Doshi-Velez & Kim, 2017), and although it is poorly defined (Lipton, 2018) it has been described as "*the ability to explain or to present in understandable terms to a human*" (Doshi-Velez & Kim, 2017, p. 2) and elsewhere as the ability to "*understand the contribution of individual features in the model*" (Lou et al., 2012, p. 1). Additionally, Lipton (2018) argued that for a model to be considered truly interpretable, it should possess the following three properties: *algorithmic transparency* (i.e., it should be possible to understand the mechanism by which the model works), *decomposability* (each part of the model, such as the model input and parameters, should have an intuitive explanation), and *simulatability* (a person should be able to consider the entire model at once). The present model possessed these three properties as well as conforming to the definitions proposed above. First, the model possessed algorithmic transparency in that the Elastic Net is a penalised linear regression. Second, the model possessed decomposability. The inputs to the model were GM voxel density values and the parameters, or beta coefficient values, weighted the contribution of each individual value to the model output, which is brain predicted age. Third, the model possessed simulatability as the entire model can be considered as follows: summing the multiplication of GM voxel density values by the average contribution of these voxels to the prediction of chronological age in the training set (i.e., the beta coefficient values) resulted in a prediction of a new individual's brain age.

While the model of brain age demonstrated good interpretability from a statistical and machine learning perspective, the biological interpretability of the model was suboptimal. The statistical model of brain age described here contained many adjacent voxels with opposite signs. The negative weights represent those areas with less volume (associated with older age). The positive weights displayed areas that have more volume associated with older age, which may seem counterintuitive: this may be because GM

in these areas represents a shift away from the cortex or periventricular regions (i.e., younger participants would have WM or CSF in those regions). An example of a similar result can be seen in an AD classification study (Dubois et al., 2014), which also used penalised regression. The biological interpretability of the model could be further improved by forcing sparsity to limit the number of voxels making significant contributions to brain age predictions. Modified Elastic Net algorithms, such as Enet-BETA (Liu & Li, 2017), can obtain sparser models which would reduce the number of predictive voxels, thereby further improving interpretability. However, as the Elastic Net's prediction accuracy can increase with feature set size (Jollans et al., 2019), further limiting the feature set size could reduce model accuracy. As such, it might be difficult to achieve the right balance between interpretability and accuracy. An alternative approach could be to incorporate a penalty such as Total Variation within the Elastic Net in order to take into account the spatial structure of MRI data and produce weight maps that show the predictive voxels clustered in regions rather than dispersed across the brain (Dubois et al., 2014). These algorithms have been shown to produce models with greater biological interpretability (i.e., spatially organised weight maps) and comparable predictive accuracy to regular Elastic Net models for classification problems (Dubois et al., 2014). However, the technical implementation of such algorithms can be difficult and computationally expensive, although solutions such as early stopping and feature screening, have been proposed (Dohmatob et al., 2015).

In conclusion, the brain age model presented here is accurate and generalizable as it significantly predicts chronological age in three independent datasets. Furthermore, this model is interpretable. Finally, brainPAD scores, calculated using this model, are associated with reduced cognitive performance within the domains of general cognitive status; semantic verbal fluency; executive function; and executive function (without processing speed). The replication of these correlations across multiple datasets demonstrates that the relationship between brainPAD and these domains of cognitive function is robust to cultural- and site/scanner effects. Therefore, these robust associations between brainPAD and specific domains of cognition demonstrate that brainPAD is a valid operational measure of BM, as better preserved structural brain health is associated with better cognitive function. Furthermore, given that brainPAD is not limited by task effects which can hinder neuropsychological assessment, these findings provide support for the use of brainPAD as an additional or supplementary objective measure of general cognitive function with applications as a general measure of brain health and cognitive performance in the clinic and as a summary outcome measure for intervention studies in research settings.

3 Chapter 3: A systematic comparison of cognitive reserve proxy variables across two older adult cohorts

3.1 Introduction

Neuropathology and measures of brain structure do not fully explain cognitive decline (Boyle et al., 2013) nor age-related variation in cognitive function (Hedden et al., 2014). This well-established gap between brain and cognition may be explained by CR, which is most commonly measured using socio-behavioural proxy variables. As outlined in Chapter 1 (see Section 1.5), CR proxies are easy to collect, have a straightforward interpretation, and are widely associated with a reduced risk of dementia and cognitive decline in epidemiological studies.

Despite the widespread use of CR proxies, the specific proxies used vary across studies and have rarely been assessed in complete CR models (i.e., alongside both a measure of cognitive outcome and a measure of brain structure). Complete models can test independent associations between proxies and cognitive function in addition to the moderation effect of proxies on the brain-cognition relationship. This enables the CR proxy to be assessed with respect to whether it satisfies the cognitive benefit criterion, which holds that a valid measure of CR should demonstrate a protective effect on cognition (Franzmeier, Duering, et al., 2017; Stern et al., 2020). Because of the variation in proxies used across studies and the lack of complete CR models, there is insufficient empirical evidence guiding the choice of proxy measures of CR and poor comparability across studies. In effect, this means researchers, particularly those lacking neuroimaging data and therefore unable to assess complete CR models, may inadvertently measure CR with proxies that are not valid measures of CR. Similarly, clinicians attempting to adjust cognitive screening measures for CR levels, may use invalid measures of CR. Furthermore, as the face validity of candidate neuroimaging measures of CR should be assessed via their correlations with CR proxies, variation in the proxies used may also impair the comparability of different candidate neuroimaging measures.

A methodology to establish the validity, and compare the robustness, of different CR proxies is to use hierarchical linear moderated regressions to systematically assess standard CR proxies and their composites in complete CR models. This approach can establish whether a CR proxy satisfies the cognitive benefit criterion, via the examination of both moderation and independent effects within the same analysis framework. This is important because, although moderation effects should ideally be observed to validate a CR proxy or measure (Stern et al., 2020), they are typically small in real-world data

(Morris et al., 1986), explaining 1-3% of the variance in the outcome (Champoux & Peters, 1987). Consequently, large sample sizes are required to detect the typically small moderation effects (Whisman & McClelland, 2005). This issue is further exacerbated when measurement error is present in either variable in the interaction term (e.g., the CR proxy and measure of brain structure) used to assess the moderation effect (Dunlap & Kemery, 1988) or when either variable in the interaction term is associated with the outcome variable (e.g., cognitive function; Whisman & McClelland, 2005). Given these difficulties in identifying moderation effects, it is important to also consider the independent effect when assessing the validity of CR proxies.

Hierarchical moderated linear regressions allow the robustness (i.e., frequency of effects using different measures of brain structure and cognitive function) and magnitude of both moderation and independent effects of different proxies to be compared. To establish their validity and robustness, five common CR proxies – education, occupational complexity, verbal intelligence, leisure activities, and exercise – were assessed in complete CR models in two separate community-dwelling older adult cohorts: TILDA (n = 313) and CR/RANN (n = 234). In addition to the five individual proxies, composite proxies were created using all possible combinations of those proxies. Fifteen complete CR models were created with three brain structure variables (GM volume, hippocampal volume, and mean cortical thickness) and five cognitive variables (verbal fluency, processing speed, executive function, episodic memory, and global cognition). The robustness and validity of the individual and composite proxies were assessed across the two datasets using hierarchical moderated linear regressions.

3.2 Method

3.2.1 Participants

The first dataset consisted of 313 community-dwelling adults (mean age = 68.90 years, SD = 6.75 years, range = 54-88 years; 50.48% female) from the TILDA dataset (as described in Section 2.2.2). Study-specific inclusion criteria (in addition to the criteria outlined in Section 2.2.2) included: no history of neurological conditions and available data for CR proxies and cognitive function.

The second dataset consisted of 234 community-dwelling adults (mean age = 64.49 years, SD = 7.42 years, range = 50-80 years; 51.28% female) selected from the CR/RANN dataset (as described in Section 2.2.2). Study-specific inclusion criteria (in addition to the criteria outlined in Section 2.2.2) included: aged 50 years or older and available data for CR proxies and cognitive function.

3.2.2 Measures: CR Proxies

Data was available for 5 socio-behavioural proxies in both datasets: *Educational attainment, Occupational complexity*, *Verbal intelligence*, *Leisure activities*, and *Physical activity*. In TILDA, further data was available for the proxies: *Cognitively stimulating activities* and *Social engagement*.

Educational attainment was measured using years of formal education in both datasets. In TILDA, participants were asked to indicate the age at which they first left continuous full-time education. This information was missing for 4 participants in the final sample (1.28%), so it was imputed using educational qualification, father's education, age, sex, and rural residence during childhood as previously described (Henretta & McCrory, 2016).

Occupational complexity was measured using the complexity of work in the dimensions of data, people, and things (Smart et al., 2014) using ratings obtained from an online catalogue of the Dictionary of Occupational Tiles (DOT: www.occupationalinfo.org). Ratings for each dimension were reversed (such that higher scores reflected greater complexity) and then summed to create a total occupational complexity score, with scores ranging from 0 (minimal complexity) to 21 (maximal complexity). This was obtained for each participant's current occupation or last occupation before retirement in TILDA and for participant's occupation of longest duration of their lifetime in CR/RANN.

Verbal intelligence was measured using the total number of correctly pronounced words on the National Adult Reading Test (NART; Nelson & Willinson, 1982) in TILDA and the American National Adult Reading Test (AMNART; Grober & Sliwinski,

1991) in CR/RANN. In TILDA, a stress/anxiety-preventative and time-saving measure (Strauss et al., 2006) was employed such that participants only completed the second half of the NART if they scored greater than 20 on the first half. A correction procedure was employed whereby scores of 0-11 were retained as full scores, but scores of 12-20 in participants who did not complete the second half were corrected using a conversion table outlined by Beardsall and Brayne (1990). Possible scores on the NART, in TILDA, ranged from 0 to 50 and on the AMNART, in CR/RANN, from 0 to 45. While the NART is often used to provide a measure of premorbid intelligence, NART scores are labelled here as verbal intelligence in line with previous CR studies (Fleck et al., 2017; Oosterman et al., 2020).

Leisure activities were assessed in TILDA by participants rating their current frequency of engagement on an 8-point Likert scale (0 = Never to 7 = Daily/Almost Daily) in 9 activities: *watching television, going to films/plays/concerts, travel, listening to music/radio, going to the pub, eating out, sports/exercise, visiting/talking on phone*, and *volunteering*. In CR/RANN, participants rated their frequency of engagement over the preceding 6 months on a 3-point Likert scale (1 = Never to 3 = Often) in 17 activities: *television/radio, cards/games, reading, lectures/concerts, theatre/movies, travel, walks/rides, crafts/hobbies, music, visiting, sports/dancing/exercise, cooking, group membership, collecting, religious activities*, and *volunteering*. For both datasets, total scores were created by summing individual responses and possible scores ranged from 0 to 63 in TILDA and 17 to 51 in CR/RANN.

Physical activity was assessed in TILDA by calculating the total metabolic minutes arising from self-reported physical activity over the last week using the International Physical Activity Questionnaire-Short Form (Craig et al., 2003; Lee et al., 2011). This questionnaire assessed the time spent in 3 categories: vigorous, moderate, and walking. Responses were converted to metabolic equivalent minutes (Craig et al., 2003) and summed. In CR/RANN, physical activity was calculated using total metabolic hours arising from physical activity in an average week. The Godin Leisure Time Exercise Questionnaire (Godin & Shephard, 1985) assessed the frequency of activity sessions lasting > 15 mins in 3 categories: strenuous, moderate, and mild exercise. Responses were then weighted by the average estimated duration of activity in each category (0.5, 0.75, 1 hr respectively) and their metabolic equivalent values (9, 5, 3; Ogino et al., 2019; Scarmeas et al., 2009).

Cognitively stimulating activities were assessed in TILDA with a questionnaire where participants rated their frequency of engagement on an 8-point Likert scale

(0=Never to 7=Daily/Almost Daily) in 5 activities: *attending classes and lectures*, *working in the garden/home or on a car, reading books/magazines, spending time on hobbies/creative activities*, and *playing cards/bingo/games*. Total scores were created by summing individual responses and possible scores ranged from 0 to 35.

Social engagement was measured in TILDA using the Social Network Index (Berkman & Syme, 1979) which provides a total score, ranging from 0 to 4, reflecting an individual's degree of social connection (McCrory et al., 2016).

Composite proxies were created by first standardising (z-scoring) individual proxies. Next, every unique combination of proxies was generated and the composite proxy was the average of those proxies. For TILDA, this produced 120 unique composite proxies. For CR/RANN, this resulted in 26 composite proxies.

To summarise, for TILDA there were 127 proxies in total (individual and composite) and 31 in total for CR/RANN. To attenuate possible effects of outliers, all proxies were Winsorized using a robust technique based on the median absolute deviation (Leys et al., 2013). Outliers were identified as values greater than a threshold of 3 median absolute deviations from the median. Identified outliers were replaced by the median +/- 3 median absolute deviations.

3.2.3 Measures: Cognitive Function

Verbal Fluency was assessed using the Semantic Verbal Fluency measures as described in Section 2.2.6. Whereas Chapter 2 had both a measure of semantic and phonemic verbal fluency, only semantic verbal fluency was assessed in the present study and so the term verbal fluency is used for brevity. These variables – and all cognitive function measures – were standardised and Winsorized, as described above for the CR proxies.

Processing Speed was measured using the Processing Speed variables as described in Section 2.2.6. Scores were reversed coded, such that higher scores reflected greater cognitive performance.

Executive Function was assessed using the Executive Function variables[18] as described in Section 2.2.6. Scores were reverse coded such that higher scores reflected greater cognitive performance.

Episodic Memory was measured in both datasets with a composite measure created from the Verbal Episodic Memory (Immediate) and Verbal Episodic Memory

[18] For clarity, Executive Function variables refers to CTT 2 (TILDA) and TMT B (CR/RANN), not the Executive Function (without Processing Speed) variables.

(Delayed) variables as described in Section 2.2.6. The composite measure was created using the average of the standardised and Winsorized immediate and delayed recall variables.

Global Cognition was measured using a composite measure of all five cognitive variables in each dataset: verbal fluency, processing speed, executive function, verbal episodic memory (immediate) and verbal episodic memory (delayed). All variables were Winsorized and standardised prior to creation of the composite. The composite variable was then Winsorized and standardised itself.

3.2.4 Measures: Brain Structure

T1 MRI scans were acquired in both datasets as described in section 2.2.3. T1 MRIs were inspected and processed in TILDA and CR/RANN using FreeSurfer v6.0 and v5.1 (Fischl, 2012), respectively, as described previously (Carey et al., 2019; Habeck et al., 2016). Total GM volume and hippocampal volume were obtained from Freesurfer and both were divided by Freesurfer's estimated total intracranial volume to adjust for head size. Brain images were parcellated using the Desikan Killiany atlas, with 34 cortical regions of interest (ROIs) per hemisphere (Desikan et al., 2006). The mean cortical thickness of each cortical ROI was calculated. Mean cortical thickness was calculated as the mean over cortical ROIs. All variables were standardised and Winsorized.

3.2.5 Statistical Analysis

Fifteen individual brain structure-cognitive function models were created for each combination of brain structure and cognitive function variable, where one brain structure variable was selected as an independent variable and one cognitive function variable was selected as an outcome variable (Fig. 3.1). A moderated hierarchical regression was conducted within each brain structure-cognitive function model (n = 15) for each unique proxy (TILDA = 127; CR/RANN = 31). In Step 1, a cognitive measure was regressed on age, sex, and a measure of brain structure. In Step 2, a proxy variable was included as an independent variable. In Step 3, the interaction term for brain structure and the proxy was added.

To protect against violations of linear regression assumptions, the analysis was repeated using a robust regression, specifically an iteratively reweighted least squares regression with Tukey's biweight function and median absolute deviation scaling. Significant effects within each dataset were only considered significant if they were statistically significant in both the linear regression and robust regression. To control for multiple comparisons and to ensure generalisability of findings, effects were only

considered significant if they were statistically significant across both datasets. The analysis was conducted with customised Python code (available here: https://github.com/rorytboyle/hierarchical_regression) which used the *statsmodels* module (Seabold & Perktold, 2010). The change in R^2 (i.e., amount of variance explained) from Step 1 to Step 2, and from Step 2 to Step 3 in linear regression models were used to assess the size of the independent and moderation effects of CR proxies, respectively. Where significant effects were observed, the mean R^2 change across both datasets was calculated to assess the average additional variance explained by the proxy and its interaction with brain structure.

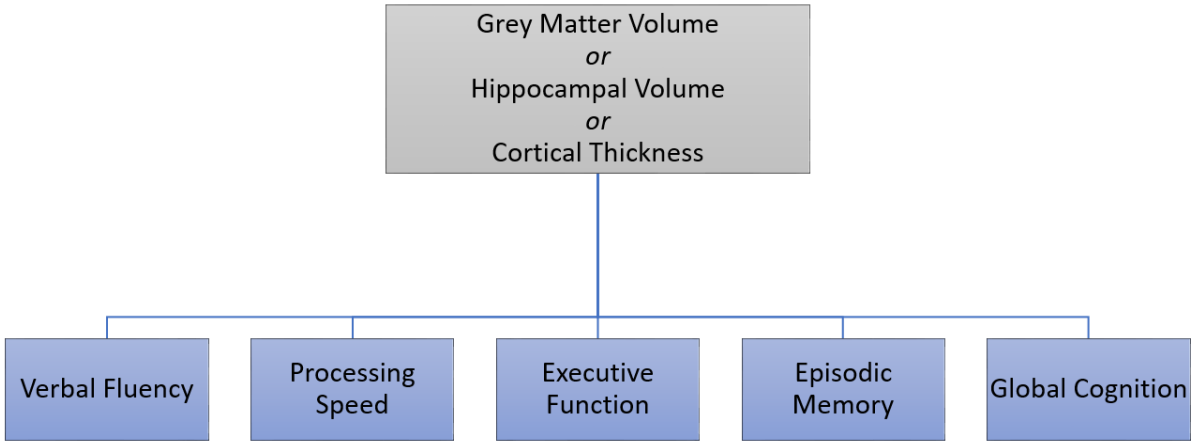


Figure 3.1. Schematic of basic brain structure-cognitive function models created for analysis.

3.3 Results

3.3.1 Demographics

In TILDA, some data were missing for mean cortical thickness (n = 34) and CTT 2 and Global Cognition (n = 2). In CR/RANN, the same n was used (n = 234) in all models. Consequently, different sample sizes were used across models within TILDA (see Table 3.1).

Table 3.1 Demographics for complete CR models assessed in hierarchical regressions.

Dataset	Brain Structure	Cognition	n	Mean Age (SD, Range)	Sex (M/F)
TILDA	Grey Matter Volume, Hippocampal Volume	Verb Flu, Proc Speed, Epi Mem	313	68.90 (6.75, 54 – 88)	155/158
	Grey Matter Volume, Hippocampal Volume	Exec Func, Glob Cog	311	68.91 (6.77, 54 – 88)	154/157
	Mean Cortical Thickness	Verb Flu, Proc Speed, Epi Mem	279	69.16 (6.64, 54 – 88)	137/142
	Mean Cortical Thickness	Exec Func, Glob Cog	277	69.18 (6.66, 54 – 88)	136/141
CR/RANN	All	All	234	64.49 (7.42, 50 – 80)	114/120

Note: Verb Flu = Verbal Fluency, Proc Speed = Processing Speed, Epi Mem = Episodic Memory, Exec Func = Executive Function, Glob Cog = Global Cognition.

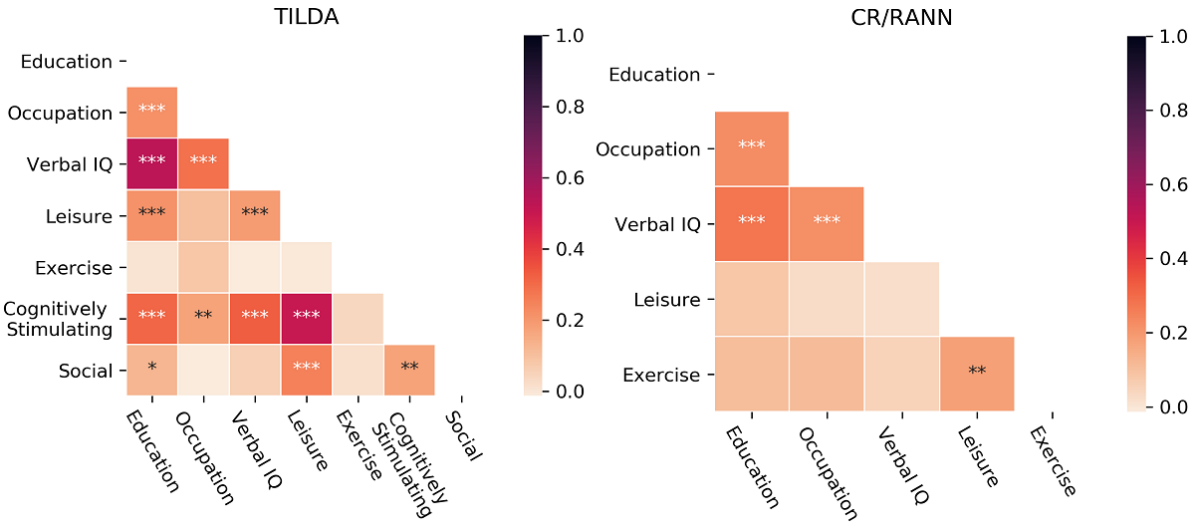


Figure 3.2. Heatmaps showing Pearson's correlations between individual proxies in each dataset. ** = p < .05, ** = p < .01, *** = p <.001.*

3.3.2 Step 1: Brain-Cognition Relationships

Models in Step 1 of the hierarchical regression (i.e., containing a brain structure measure, sex, and age) were significantly associated with cognitive measures across both datasets (see Tables 3.2 and 3.3), except for two models in CR/RANN (hippocampal volume-executive function, and hippocampal volume-episodic memory). Sex was independently associated with cognitive function in 40% and 20% of brain-cognition models in TILDA and CR/RANN, respectively. These associations were such that in these models in TILDA, females had higher cognitive function than males, on average, with other variables (i.e., brain structure and age) being equal. In these models in CR/RANN, females had lower cognitive function than males, on average, with other variables being equal. Age was negatively associated with cognitive function, independent of brain structure and sex, in 100% and 40% of models in TILDA and CR/RANN, respectively.

In TILDA, only one brain structure variable, mean cortical thickness, was independently positively associated with cognitive function (processing speed). In CR/RANN, GM volume was independently positively associated with all cognitive measures and cortical thickness was independently positively associated with all cognitive measures except for processing speed. Hippocampal volume was not independently associated with any measure of cognition in either dataset.

Table 3.2 Results from step 1 of hierarchical regressions in TILDA.

Cognition	Model Statistics			Brain Structure		Sex	Age
	n	R^2	*f*	Variable	*β*	*β*	*β*
Verb Flu	313	.043	4.597**	Grey Matter Volume	.042	-.030	-.205**
Proc Speed	313	.129	15.320****		.041	.084	-.360****
Exec Func	311	.143	17.070****		.048	.052	-.383****
Epi Mem	313	.079	8.780****		.021	.352**	-.207**
Glob Cog	311	.159	19.400****		.048	.217*	-.373****
Verb Flu	313	.042	4.475**	Hippocampal Volume	-.005	-.004	-.229**
Proc Speed	313	.129	15.226****		-.025	.120	-.394****
Exec Func	311	.143	17.010****		-.041	.101	-.428****
Epi Mem	313	.080	8.902****		.044	.341**	-.195**
Glob Cog	311	.158	19.171****		.002	.243*	-.396****
Verb Flu	279	.051	4.898**	Mean Cortical Thickness	.103	.002	-.192**
Proc Speed	279	.173	19.217****		.122*	.042	-.370****
Exec Func	277	.195	22.040****		.090	.065	-.428****
Epi Mem	279	.091	9.202****		-.036	.414**	-.216***
Glob Cog	277	.195	22.105****		.065	.251*	-.391****

*Note: * = p < .05, ** = p < .01, *** = p < .001, **** = p < .0001. Verb Flu = Verbal Fluency, Proc Speed = Processing Speed, Exec Func = Executive Function, Epi Mem = Episodic Memory, Glob Cog = Global Cognition.*

Table 3.3 Results from step 1 of hierarchical regression models in CR/RANN.

Cognition	Model Statistics			Brain Structure		Sex	Age
	n	R^2	*f*	Variable	*β*	*β*	*β*
Verb Flu	234	.087	7.320***	Grey Matter Volume	.258***	-.073	-.062
Proc Speed	234	.087	7.344***		.218**	-.296*	-.120
Exec Func	234	.047	3.762*		.175*	-.247*	-.063
Epi Mem	234	.061	4.998**		.221**	.070	-.072
Glob Cog	234	.130	11.498****		.330****	-.148	-.117*
Verb Flu	234	.043	3.449*	Hippocampal Volume	.078	<-.001	-.111*
Proc Speed	234	.061	5.014**		.034	-.225	-.173**
Exec Func	234	.030	2.339		.026	-.190	-.107
Epi Mem	234	.033	2.608		.032	.142	-.127*
Glob Cog	234	.069	5.671***		.061	-.044	-.195**
Verb Flu	234	.065	5.303**	Mean Cortical Thickness	.166**	-.024	-.098
Proc Speed	234	.073	6.063***		.129	-.252*	-.152*
Exec Func	234	.048	3.834*		.153*	-.226	-.077
Epi Mem	234	.053	4.281**		.159*	.106	-.098
Glob Cog	234	.109	9.401****		.231***	-.092	-.158**

*Note: * = p < .05, ** = p < .01, *** = p < .001, **** = p < .0001. Verb Flu = Verbal Fluency, Proc Speed = Processing Speed, Exec Func = Executive Function, Epi Mem = Episodic Memory, Glob Cog = Global Cognition.*

3.3.3 Step 2a: Independent Effects

Significant positive independent effects were observed for 18 proxies, including 2 individual proxies and 16 composites, across the 15 models in both datasets. As there was a large number of individual significant independent effects across both datasets, a spreadsheet containing detailed data on these effects is available here: https://tinyurl.com/sigAcrossBoth. Similar spreadsheets are available for effects within TILDA: https://tinyurl.com/sigInTILDA and within CR/RANN: https://tinyurl.com/sigInCRRANN.

The proxy with the largest average independent effect on cognition was verbal intelligence (mean R^2 change = 0.097; see Fig. 3.3). Verbal intelligence was the most robust proxy: independent effects were replicated across both datasets in 100% of models. The largest average independent effects were observed for verbal intelligence on global cognition where it explained a mean additional 16.80% (hippocampal volume), 15.87% (GM volume), and 14.66% (mean cortical thickness) of the variance after accounting for age, sex,

and brain structure (see Fig. S2 in 7.2.1 Supplemental Results for scatter plots of proxies with 10 largest average independent effects). Education was the only other individual proxy with reproducible independent effects on cognition (mean R^2 change = 0.033), which were observed in 20% of models, all of which contained executive function.

The most robust composite proxy was comprised of occupational complexity and verbal intelligence (mean R^2 change = 0.064) which was replicated in 86.67% of models. The composite proxy with the largest average independent effect on cognition was educational attainment and verbal intelligence (mean R^2 change = 0.08) which was replicated in 80% of models. Only one composite with reproducible independent effects on cognition – occupational complexity and physical activity – did *not* include verbal intelligence. This was the least robust composite as it was replicated in a single model and had the smallest average effect (mean R^2 change = 0.002).

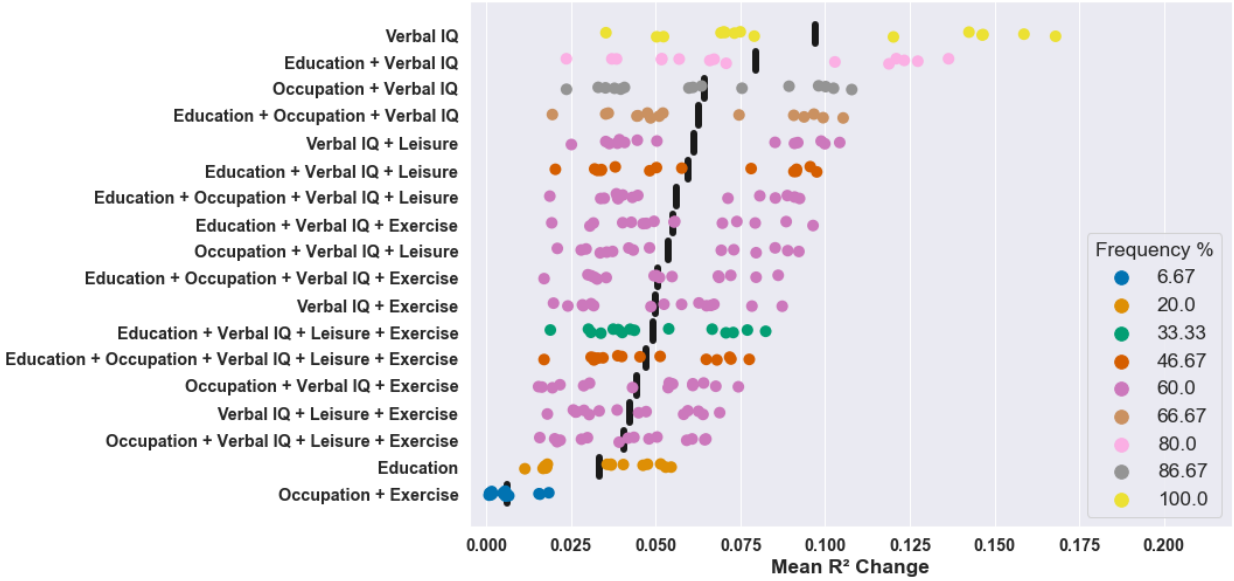


Figure 3.3. Mean R^2 change across datasets in all models for proxies with significant effects. *+ indicate composite proxies (e.g., Education + Verbal IQ = composite of educational attainment and verbal intelligence). Black vertical bars represent the mean R^2 change across all models for that proxy in both datasets. All models were adjusted for brain structure, age, and sex.*

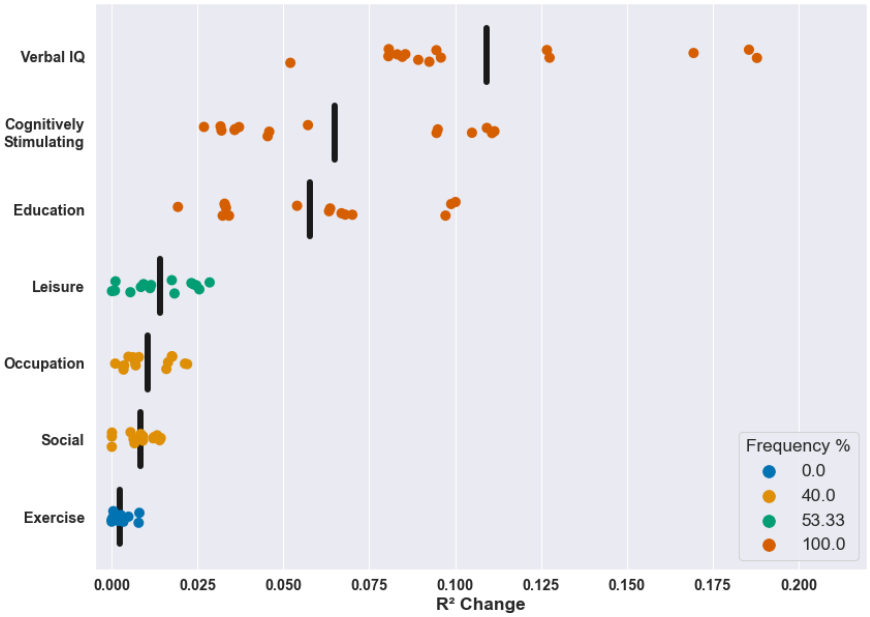


Figure 3.4. R^2 change in all TILDA models for individual proxies. *Black vertical bars represent the mean R^2 change across all models for that proxy. All models were adjusted for brain structure, age, and sex.*

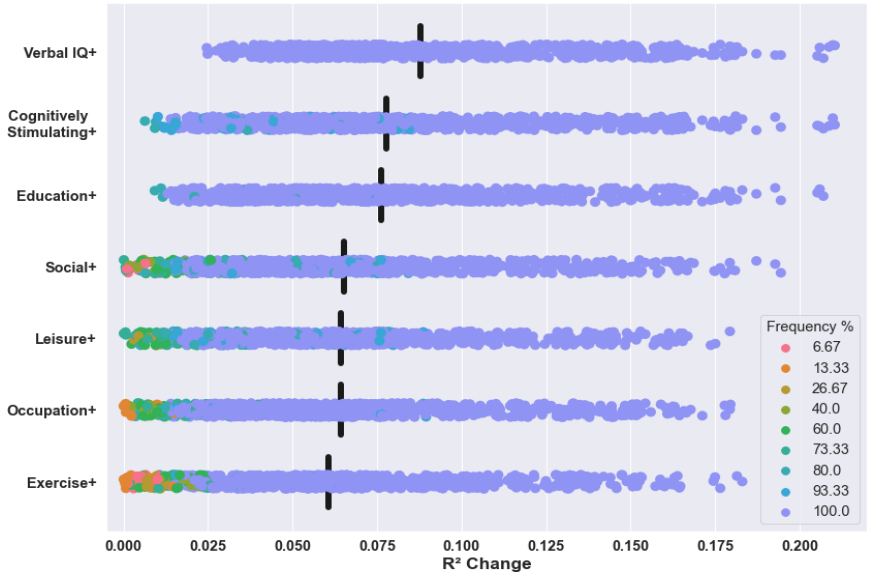


Figure 3.5. R^2 change in all TILDA models for composite proxies. *Each row refers to all composites including that proxy (e.g., Verbal IQ+ refers to all composites including verbal intelligence). Black vertical bars represent the mean R^2 change across all models for all composites containing that proxy. All models were adjusted for brain structure, age, and sex.*

3.3.4 Step 2b: Additional Independent Effects

Data was only available for cognitively stimulating activities and social engagement in TILDA. Consequently, these effects could not be assessed in terms of their reproducibility. However, within TILDA, positive independent effects of cognitively stimulating activities on cognition were observed in 100% of models and this proxy had the second largest average independent effect of all individual proxies (mean R^2 change = 0.065, see Fig. 3.4). In contrast, positive independent effects of social activities on cognition were observed in only 40% of models and this proxy had the second smallest average independent effect of all individual proxies (mean R^2 change = 0.008). The only individual proxy with smaller effects than social engagement was physical activity which did not have significant effects on cognition in any model.

Composite proxies including verbal intelligence had the largest average effects on cognition, followed by cognitively stimulating activities, and then education (see Fig. 3.5). Composites including verbal intelligence had significant effects on cognition in all models in TILDA. The composite with the largest effect in TILDA was verbal intelligence and cognitively stimulating activities (mean R^2 change = 0.13).

3.3.5 Step 3: Moderation Effects

There were no significant moderation effects in either dataset for any proxy. There were however non-replicated moderation effects that did not survive correction for multiple comparisons. Negative moderation effects are consistent with the CR hypothesis as they reflect weaker associations between brain structure and cognition in individuals with higher CR, suggesting that individuals with higher CR are less reliant on brain structure to sustain cognitive function. 31 non-replicated negative moderation effects were observed in TILDA (see Table S5 in 7.2.1 Supplemental Results), but none survived correction for multiple comparisons (Bonferroni-adjusted alpha = 0.0004: alpha [0.05] / comparisons per model [127]). 61.29% of these effects were observed for composite proxies including cognitively stimulating activities, which was not available in CR/RANN. No negative moderation effects were observed in CR/RANN.

Positive moderation effects contradict the CR hypothesis as they reflect stronger associations between brain structure and cognition in individuals with higher CR, suggesting that individuals with higher CR are more reliant on brain structure to sustain cognitive function. Non-replicated positive moderation effects were observed in both datasets (see Table S6 in Section 7.2.1 Supplemental Results) but none survived correction for multiple

comparisons. Eight effects were observed in TILDA (Bonferroni-adjusted alpha = 0.0004) and seven effects were observed in CR/RANN (Bonferroni-adjusted alpha = 0.0016: alpha [0.05] / comparisons per model [31]). The Bonferroni corrections for multiple comparisons applied here are liberal as they correct for number of proxies compared per brain-cognition model (TILDA: 127, CR/RANN: 31) rather than number of total comparisons across all proxies and all brain-cognition models (TILDA: 1905; CR/RANN: 465).

3.4 Discussion

The reproducibility and magnitude of moderation and independent effects of 33 CR proxies, comprised of 5 standard individual proxies and all their unique combinations, were assessed across 2 datasets to investigate their validity as measures of CR. No moderation effects of CR proxies on the association between brain structure, as measured by GM volume, hippocampal volume, or mean cortical thickness, and cognition were observed across both datasets. Replicated independent effects – positive associations with cognitive function, independent of brain structure – were observed for 2 individual proxies (verbal intelligence and educational attainment) and 16 composites. The most robust and largest independent effects on cognition were found for verbal intelligence, which satisfied the independent effect criterion in all 15 brain-cognition models across both datasets. Educational attainment satisfied the independent effect criterion in 3 brain-cognition models. No composite proxy had larger or more robust independent effects on cognition than verbal intelligence alone. These results suggest that when proxies are used to measure or adjust for CR in cognitively healthy older adults in cross-sectional studies, verbal intelligence should be used.

Verbal intelligence had the largest and most robust independent effects on cognition. Unlike previous studies, due to the availability of two large neuroimaging datasets, it was demonstrated that independent effects of verbal intelligence on cognition were present in several brain-cognition models and were replicable. This validation of verbal intelligence as a CR proxy supports previous, narrower, associations between verbal intelligence and cognitive function in the presence of hippocampal atrophy (Topiwala et al., 2019), a neuropathological 'residual' measure of CR (Negash et al., 2013), a functional connectivity measure of CR based on task potency (van Loenhoud et al., 2020), and a possible neuromarker of CR, locus coeruleus signal intensity (Clewett et al., 2016).

Aside from verbal intelligence, the only other individual proxy with replicable independent effects on cognition was educational attainment. These effects were only observed in brain-cognition models where executive function was the cognitive outcome variable. While education has been previously positively associated with executive function, without accounting for brain structure, in cognitively healthy older adults (Laubach et al., 2018) and in a systematic review (Opdebeeck et al., 2016), the present results show that this association is independent of GM volume, hippocampal volume, and mean cortical thickness. Notably, the effects of education on cognition were less robust than verbal

intelligence, as positive associations were not seen across both datasets for verbal fluency, processing speed, episodic memory and global cognition. As such, these results suggest that educational attainment is not a reliable individual proxy of CR in cognitively healthy older adults (see Chapter 5, section 5.2.2 for further discussion of the evidence favouring the use of verbal intelligence over educational attainment). This conclusion is supported by previous findings including a systematic review which found positive evidence for education in only 38% of complete models with cognitively healthy samples (Chapko et al., 2018) and a non-significant association between education (when considered separately from other possible CR proxies) and a neuropathological residual measure of CR (Reed et al., 2011). Based on their findings using ex-vivo neuropathological measures, Reed et al. (2011) concluded that the observed effects of education on cognition should not be simply considered as reserve effects. These results further show that this conclusion is valid when using in-vivo neuroimaging measures of brain structure.

Significant positive independent effects on cognition were observed for 16 different composite proxies across both datasets. Three composites had significant independent effects on cognition in at least two thirds of the brain-cognition models assessed: occupational complexity and verbal intelligence (86.67% of models); education and verbal intelligence (80% of models); and education, occupational complexity, and verbal intelligence (66.67% of models). This is a novel finding as the most robust composite – occupational complexity and verbal intelligence – has never (to the best of my knowledge) been used previously as a CR proxy, likely due to the predominant use of education both as an individual proxy and in composites. The next most robust composite of education and verbal intelligence has been widely used (Fleck et al., 2017; Kwak et al., 2020; Oosterman et al., 2020; Pettigrew et al., 2013, 2017; Soldan et al., 2017; Steffener et al., 2014) and the present results support a previous positive association between this composite and episodic memory, controlling for GM volume (Kwak et al., 2020). A speculative explanation for the greater robustness of occupational complexity and verbal intelligence as a composite may be that occupational complexity and verbal intelligence are less strongly correlated with each other than educational attainment and verbal intelligence (see Fig. 3.2).

While composite proxies purportedly provide advantages over individual proxies, these results show that their independent effects on cognition are less robust (i.e., less frequently observed across brain-cognition models) and smaller in magnitude than those found for verbal intelligence alone. This may be explained by the large individual effects of

verbal intelligence and its strong correlation with other proxies (see Fig. 3.2) considering that all composite proxies with replicated effects contained verbal intelligence, except for the composite with the least robust effects, occupational complexity and physical activity. While adding another proxy to verbal intelligence to form a composite should have an additive effect, this could also add noise to an already strong proxy measure as well as shared variance in situations where the proxies are correlated. Consequently, the overall effect of the composite may then be smaller than verbal intelligence alone. Alternative methods to creating composites, such as PCA, could potentially mitigate this issue but may not be theoretically appropriate (Jones et al., 2011) and incorporating this method within the analysis framework used here would have significantly increased the complexity of the analysis. Of all composites considered here, these results especially support the use of education and verbal intelligence as well as occupational complexity and verbal intelligence as composite proxies in cross-sectional studies using cognitively healthy older adults, where multiple proxies are available. However, using composites may lead to more Type II errors than using verbal intelligence alone, given the more robust and larger effects of verbal intelligence. As such, based on these results, it is recommended that researchers should use, or at least repeat analyses using, verbal intelligence alone.

Robust independent effects on cognition, independent of GM volume, hippocampal volume, or mean cortical thickness, were not found for 3 individual proxies across both datasets. Occupational complexity was not positively associated with any domain of cognitive function, adjusting for brain structure. This suggests that the small positive associations between this proxy and cognition, as reported in a meta-analysis (Opdebeeck et al., 2016), may not be independent of GM volume, hippocampal volume, or mean cortical thickness. Unlike the detailed nature of the occupational complexity measure used here, occupational complexity has been typically measured using government classification codes that are effectively a socioeconomic classification of occupations (e.g., the UK's Office Of Population Statistic classification as in Staff et al., 2004). As such, previously reported effects for occupational complexity may have in fact reflected the effect of socioeconomic status, which can support cognitive health via greater access to resources and healthcare, among many other mechanisms (Jones et al., 2011). While Chapko et al. (2018) concluded that the evidence for this proxy in complete CR models using cognitively healthy samples was inconclusive, based on the present results, it is concluded that occupational complexity should not be used as an individual CR proxy in cross-sectional studies of cognitively healthy older adults.

As with occupational complexity, there was no robust evidence to support the use of leisure activities as an individual CR proxy in cognitively healthy older adults. Although it has been associated with a reduced risk of dementia and AD (Crowe et al., 2003; but cf. Sommerlad et al., 2020), few studies have rigorously tested this proxy in a complete CR model. One study found a moderation effect for midlife leisure activities but in line with the present findings, they did not find evidence of either a moderation or independent effect for later life leisure activities (Chan et al., 2018). Future research is warranted to clarify which specific leisure activities should be included in measures for this proxy given that only a few activities have been associated with cognition in mid-/old-age samples, albeit without adjusting for brain structure (Anatürk, Suri, Smith, et al., 2020; Park et al., 2019). Considering these results, it is suggested that later life leisure activities should not be used as an individual CR proxy in cross-sectional studies of cognitively healthy older adults.

Finally, the present results do not support the use of physical activity as an individual CR proxy. While this proxy has been previously associated with cognitive function in older adults without controlling for brain structure (Lipnicki et al., 2019; Tsai & Chang, 2019), the present results show that these associations are not independent of GM volume, hippocampal volume, or mean cortical thickness. This supports previous findings of non-significant associations from the few complete CR models assessing this proxy adjusting for brain structure using GM volume and hippocampal atrophy (Casaletto et al., 2020; Topiwala et al., 2019). The disparity in the observed associations when brain structure is accounted for could be because the protective effects of exercise may be exerted via improved brain maintenance, i.e., the relative preservation of brain structural health (Nilsson & Lövdén, 2018; Stern et al., 2020), rather than improved CR (Cheng, 2016). This is supported by the finding that the protective effects of exercise on cognition were mediated by increases in prefrontal cortex volume (Weinstein et al., 2012) and also by associations of greater physical activity with lower brain-predicted age difference scores (Steffener et al., 2016), which reflects better brain maintenance (Habeck et al., 2017), and greater cortical thickness (Stern et al., 2019) and regional GM volumes (Arenaza-Urquijo et al., 2017; Erickson et al., 2014). Setting aside a possible contribution of physical activity to brain maintenance, these results suggests that it does not contribute to greater CR and therefore should not be used as an individual CR proxy in cross-sectional studies of cognitively healthy older adults.

Robust moderation effects were not identified for any proxy here across datasets. This lack of evidence is in line with previously reported non-significant moderation effects

on the relationship between episodic memory and GM volume (Kwak et al., 2020) and right hippocampal volume (Resende et al., 2018) but conflicts with previous evidence of significant moderation effects reported for CR proxies in similar brain-cognition models (Chan et al., 2018; Steffener et al., 2014; Vuoksimaa et al., 2013). However, the evidence for moderation is largely inconsistent as highlighted by the finding of moderation effects reported on 1 measure, but not on 2 other measures, of episodic memory within the same study (Vuoksimaa et al., 2013) and even findings of a positive moderation effect, which contradicts the CR hypothesis, on the relationship between left hippocampal volume and episodic memory (Resende et al., 2018). It is likely that the non-significant moderation effects observed here further reflect the general difficulties in detecting CR moderation effects.

The ability to detect a moderation effect here may have been impaired because the participants were cognitively and neurologically healthy and therefore had a relatively restricted range of cognitive function and brain atrophy in comparison to cognitively and/or neurologically impaired individuals. The relatively restricted range of the predictor variable of brain structure restricts the range of the interaction term (McClelland & Judd, 1993) which can substantially reduce statistical power to detect a moderation effect (Aguinis & Stone-Romero, 1997). This is exacerbated by the fact that neuroimaging variables explain a relatively small amount (20%) of variance in healthy older adults cognition (Hedden et al., 2014), which effectively constrains the size of the moderation effect (Whisman & McClelland, 2005). While the present study was designed using pre-existing data from two cognitively and neurologically healthy cohorts, an experimental approach where individuals with extremely low or high scores on measures of cognitive function and brain structure are oversampled may be better able to detect the existence of a moderation effect for these CR proxies (McClelland & Judd, 1993).

It was not possible to assess the reproducibility of effects for cognitively stimulating activities and social engagement across datasets as only the TILDA dataset had sufficient data for these proxies. Within TILDA, cognitively stimulating activities was highly robust as it was significant in all brain-cognition models, and had the largest average independent effect on cognition after verbal intelligence. This finding supports associations between this proxy and neuropathological ‘residual’ measures of CR (Negash et al., 2013; Reed et al., 2011) and suggests that previously reported consistent positive associations (S. L. Harrison et al., 2015; Opdebeeck et al., 2016) can be observed with various cognitive domains when

controlling for GM volume, hippocampal volume, and mean cortical thickness. Social engagement was less robust as it was significant in only 40% of brain-cognition models and had the second smallest average independent effect on cognition of all individual proxies. This inconsistent evidence emphasises a need for further study of social engagement in complete CR models. While mixed evidence of moderation effects have been reported to-date for this proxy controlling for neuropathology (Bennett, Schneider, Tang, et al., 2006; Biddle et al., 2019), this is the first attempt to assess it in a complete CR model including neuroimaging variables. As the focus of the present study was on replication across datasets rather than single dataset findings, requiring correction for multiple comparisons, and because these proxies were only available in a single dataset, these findings remain speculative until they can be replicated. With this in mind, it is tentatively suggested that cognitively stimulating activities, but not social engagement, may be a reasonable choice of CR proxy in cross-sectional studies of cognitively healthy older adults where verbal intelligence is not available.

The present study established empirically supported recommendations in the use of proxies to measure CR cross-sectionally. Nonetheless, there were some limitations which, if addressed in future research, could further strengthen these recommendations and provide additional insights. As discussed above, it was not possible to assess the effects of cognitively stimulating activities and social engagement across both datasets. This limited the ability to make definitive conclusions about these proxies and may have further limited the ability to detect robust moderation effects as ~60% of proxies with moderation effects within TILDA contained cognitively stimulating activities but could not be assessed across both datasets.

The present study did not assess all possible CR proxies. In particular, one potential CR proxy that was not assessed here was bilingualism, which has been associated with a delayed onset of dementia (Mendez et al., 2019; but cf. Yeung et al., 2014; Zahodne et al., 2014). However, bilingualism is not a common CR proxy. It has only been included as a CR proxy in a single questionnaire measure of CR (Kartschmit et al., 2019) and in less than 10% of reviewed studies (Hannigan et al., 2015). While assessing this proxy could have potentially provided some clarity on the validity of bilingualism as a CR proxy, data on bilingualism was not available.

Some CR proxies, namely leisure activities and physical activity, were measured differently in both datasets. Differences in these measures or in the specific activities

included in each measure may have contributed to differing effects across both datasets. This may be particularly pertinent for leisure activities as its relationship with cognitive function can vary based on the specific leisure activities assessed (Anatürk, Suri, Smith, et al., 2020). For example, the measure of leisure activities used in TILDA did not contain information on leisure-time computer use, which was the only leisure activity in a large UK Biobank study found to be positively associated with trail making performance and prospective memory (Anatürk, Suri, Smith, et al., 2020). However, this variability across the two datasets reflects the variability in the measurement of CR with proxies in the wider literature.

Despite the discussed limitations, the present findings are informative for researchers using proxies as measures of CR, particularly in cross-sectional studies of cognitively healthy older adults. Building on previous meta-analyses and systematic reviews of CR proxies, a wider set of standard proxies, including their composites, were assessed and their effects were evaluated across complete and theoretically consistent models of CR and in multiple brain-cognition relationships. The analysis framework enabled the comparison of the robustness and magnitude of effects. Furthermore, the reported findings are stringent, robust and replicable, as they were only considered statistically significant if they were replicated in a robust regression and across two datasets.

In conclusion, the present study is the first systematic investigation of the validity of different proxies, and their composites, in complete CR models. Verbal intelligence was associated with better cognitive function in all variables assessed, controlling for mean cortical thickness, GM volume, and hippocampal volume. The independent effects on cognition of education and composite proxies, including verbal intelligence and occupational complexity as well as verbal intelligence and education, were smaller and less robust. These results provide firm, data-driven, recommendations for the use of verbal intelligence as a CR proxy in cross-sectional studies focused on cognitively healthy older adults, over other proxies including education, occupational complexity, leisure activities, exercise, and composites including all possible combinations of these proxies. While no robust moderation effects were found for any proxy here, this may be due to the considerable statistical difficulties in detecting such effects in normal healthy ageing samples. In sum, the finding of robust independent effects on cognition across all brain-cognitive domains assessed provides strong evidence for the use of verbal intelligence as a CR proxy in cross-sectional studies of cognitively healthy older adults.

4 Chapter 4: Measuring cognitive reserve using connectome-based predictive modeling

4.1 Introduction

Chapter 3 demonstrated that verbal intelligence is a robust and valid cross-sectional proxy measure of CR in cognitively healthy older adults. However, across 15 different complete CR models, verbal intelligence explained at most 16.8% of the variance in cognition, after accounting for age, sex and brain structure. As such, other measures of CR are likely necessary to obtain optimal accuracy when measuring CR. Using functional neuroimaging to identify neural networks, whose strength or expression differs as a function of CR, may provide a more direct measure of the CR construct (Stern et al., 2020; Stern & Barulli, 2019). There are several advantages to this approach, if validated, compared to the use of proxies for measuring CR. A neuroimaging measure would provide an objective measure of CR, whereas proxy variables typically rely on self-reported information, which can be inaccurate, especially among individuals with poorer memory and for proxies such as cognitively stimulating activities (Everson-Rose, 2003). Moreover, a neuroimaging measure could be influenced by exposure to various lifetime experiences (Stern & Barulli, 2019) without directly reflecting the change in exposure itself. This would enable the evaluation of interventions designed to increase CR. For instance, imagine an intervention consisting of engagement in cognitively stimulating activities. Assessing this intervention using engagement in cognitively stimulating activities as a proxy measure would not provide any information about the effects of the intervention, as the proxy measure reflects exposure to a specific factor, and not the outcome of that exposure. In contrast, assessing the effects of the exposure on a neuroimaging measure of CR may provide information about the outcome, given that interventions comprised of cognitively stimulating activities have been found to increase functional connectivity within the posterior DMN (De Marco et al., 2016). Importantly, a brain-based approach could provide important insights into the mechanisms of CR and may enable more focused research into the underlying neural processes of CR.

A novel approach to developing a functional neuroimaging measure of CR is to use connectome-based predictive modeling (CPM; Shen et al., 2017). CPM is a data-driven method that enables the prediction of cognitive or behavioural phenotypes from functional connectivity data. In short, CPM summarises the most relevant connections – also referred to as edges – for the phenotype, across the whole brain. Within cross-validation frameworks, these edges are summed to create three measures – positive, negative, and combined

network strength – which summarise the connectivity strength of edges that are related to the phenotype of interest.

CPM is a promising strategy for measuring CR for a number of reasons. First, it provides a single scalar value – combined network strength – (or three values if desired by also using positive and negative network strength) that summarises the connectivity strength of CR-related networks. Second, it has been successfully applied to predict specific cognitive domains – fluid intelligence (S. Gao et al., 2019; Greene et al., 2018), attention (Fountain-Zaragoza et al., 2019; Rosenberg et al., 2016), and executive function (Henneghan et al., 2020) – which themselves have been directly associated with proxy measures of CR (Chan et al., 2018; Lavrencic et al., 2018). Third, CPM is not constrained by *a priori* hypotheses as whole-brain functional connectivity is used in a data-driven manner. This approach means that the search for CR-related networks is not constrained to hypothesised regions or networks of interest. This is important given that individual differences in cognition have been more accurately predicted using whole-brain data rather than data obtained from specific ROIs (Zhao et al., 2021). Fourth, CPM has been widely shown to create measures of cognitive and behavioural phenotypes that generalise across datasets (M. Gao et al., 2020; Rosenberg et al., 2016; Yip et al., 2019; Yoo et al., 2018). This means that CPM has successfully developed measures of a phenotype in one dataset that have then been applied to accurately measure that phenotype in an independent dataset. This could be very beneficial for clinical use and for researchers without access to large neuroimaging datasets as they would be able to generate a CR measure using a model previously developed on a large dataset.

CPM can capitalise on recent developments in measuring CR using neuroimaging by using the CR residual as the outcome – or target – variable to be predicted from the functional connectivity data. The CR residual has face validity (Habeck et al., 2017; D. H. Lee et al., 2019), satisfies the cognitive benefit criterion (Reed et al., 2010; Zahodne et al., 2013) and provides a more direct measure of CR than proxies that have been used as target variables in previous attempts to measure CR with fMRI (Stern et al., 2018; van Loenhoud et al., 2020). Furthermore, these previous attempts to measure CR relied on task-based fMRI data, and it is not clear if they could be applied to measure CR from resting-state data. In contrast, CPM has been used to develop accurate measures of cognitive phenotypes using task-based fMRI data which have been then successfully applied to resting-state data from independent datasets (Rosenberg et al., 2016). This is critical if the measure is to be

widely used in research and clinical settings as resting-state fMRI is easier to collect and less affected by task-related confounds (Fox & Greicius, 2010; Mennes et al., 2013; Stern, 2005). Nevertheless, compared to resting-state data, task-based fMRI data may augment individual differences in neural processes or networks underlying a phenotype (Greene et al., 2018; Yoo et al., 2018). As such, measures initially derived from task-based data may be more accurate. Indeed, it has been shown that task-based data enables more accurate predictions of fluid intelligence using CPM than resting-state data (Greene et al., 2018). Therefore, CPM may provide the best of both worlds as it enables a measure to be developed using task-based data, potentially increasing the accuracy of the measure, which can then be applied to resting-state data, thereby increasing the usability and clinical potential of the measure.

The present study aimed to develop a functional neuroimaging measure of CR by applying CPM to task-based fMRI data and to externally validate the measure on resting-state fMRI data in an independent dataset. Furthermore, the present study aimed to assess the validity of the functional neuroimaging measures as measures of CR in both datasets by establishing whether they display face validity and satisfy the cognitive benefit criterion.

4.2 Methods

4.2.1 Participants

Training Set (CR/RANN)

The training set consisted of data from 220 participants of the CR/RANN studies (as described in Section 2.2.2). From an initial dataset of 384[19] participants, participants were excluded according to the following criteria in the present study: (1) task fMRI data not available in session 1 or 2 of imaging data collection (3 participants excluded); (2) missing volumes in task fMRI scan (2 participants excluded); (3) presence of possible lesions or severe atrophy (5 participants excluded); (4) data quality issues including scanner artefacts, motion artefacts, and signal dropout (91 participants excluded); (5) excessive head motion defined as mean framewise displacement (FWD) > 0.4 mm or frame to frame movements > 97.5th percentile of head movements during resting-state fMRI scan (41 participants excluded); (6) missing data for CR residual (22 participants excluded).

Test Set (TILDA)

The test set consisted of 294 participants from the TILDA dataset (as described in Section 2.2.2). From an initial dataset of 561[20] participants, participants were excluded according to the following criteria: (1) structural and fMRI data available (5 participants excluded); (2) history of Parkinson's disease, stroke, or transient ischaemic attack (11 participants excluded); (3) GM or WM lesion evident in structural MRI (19 participants excluded); (4) data quality issues including scanner artefacts, motion artefacts, and signal dropout (22 participants excluded); (5) excessive head motion defined as mean FWD > 0.4 mm or frame to frame movements > 97.5th percentile of head movements during resting-state fMRI scan (154 participants excluded); (6) missing data for CR residual (56 participants excluded)**.** Demographic information for both datasets is presented in Table 4.1.

[19] In Chapter 2, the initial CR/RANN dataset contained 380 participants. After completion of Chapters 2 and 3, data for 4 further participants was made available for analysis in Chapter 4. As a result, the initial CR/RANN dataset in Chapter 4 contained 384 participants.

[20] In Chapter 2, the initial TILDA dataset contained 553 participants. Due to a technical error, 8 participants were not included in this initial dataset, but their data was made available for analysis in Chapter 4. As a result the initial TILDA dataset in Chapter 4 contained 561 participants.

4.2.2 Image acquisition

Training Set (CR/RANN)

CR/RANN imaging data were obtained using a 3 T Philips Achieva scanner over the course of 2 separate 2-hour imaging sessions in order to accommodate 12 fMRI scans and additional imaging modalities. Here, a single fMRI scan session was used, which was collected during completion of the Paper Folding task (Ekstrom et al., 1976). A screen at the foot of the MRI bed, which was viewed by participants using a mirror system within the head coil, displayed the task stimuli. Stimuli were back-projected onto the screen using an LCD projector. Participants responded via a LUMItouch response system (Photon Control Company). Participants who required their vision corrected to normal were provided with MRI compatible glasses (manufactured by SafeVision, LLC. Webster Groves, MO). The task was administered using EPrime (v2.08).

The Paper Folding task requires participants to select a pattern of holes that would be created following a sequence of folds in a sheet of paper, in which a hole is punched. Participants were shown the sequence of folds at the top of the screen and were presented with 5 different patterns in a row below. Participants were instructed to respond by pressing 1 of 5 buttons corresponding to the number of the correct pattern. Participants completed a short, 4-6 trial, practice test which provided feedback and explanations for each trial. In order to maintain participant engagement during the task, a timing protocol was designed which enabled participants to respond at their own rate (Stern et al., 2014). This protocol consisted of a variable number of trials, depending on participant's performance. Trials consisted of a 24 second fixation-cross followed by presentation of the stimulus. The stimulus was terminated immediately after response except if a response was made within 11 seconds, where the stimulus was terminated at 11 seconds. Where no responses were made, the stimulus terminated after 85 seconds. There was a 35 second inter stimulus interval.

The fMRI data were acquired using a 14 minute 26 second echo-planar imaging (EPI) pulse sequence (flip angle = 72°, slice thickness = 3 mm, slice gap = 0 mm, slices = 33, TR = 2000 ms, TE = 2 ms). In addition to 430 volumes, 3 dummy volumes were acquired at the start of the task fMRI scan and automatically discarded. Structural MRI data were acquired as described in Section 2.2.3.

Test Set (TILDA)

TILDA imaging data were obtained using a 3 T Philips Achieva scanner during a 45-min MRI battery. Resting-state fMRI data were acquired using a 6 minute 51.9 second gradient EPI sequence (flip angle = 90°, slice thickness = 3.2 mm, slice gap = 0.3 mm, slices = 38, TR = 2000 ms, TE = 28 ms). In addition to 200 volumes, 4 dummy volumes were acquired at the start of the resting-state fMRI scan and automatically discarded. Structural MRI data were acquired as described in Section 2.2.3.

4.2.3 Image Preprocessing

The same preprocessing pipeline was applied to the training and test sets but each dataset was preprocessed separately. Prior to preprocessing, functional and structural images were manually reoriented to ensure that images were in roughly the same orientation in MNI space. Images were then visually inspected for artefacts, data quality issues, possible lesions, and severe atrophy. Images were preprocessed using SPM12 and fMRI images were corrected for slice-timing and head motion. Nuisance regressors consisted of 6 motion estimates, mean WM signal, mean CSF signal, and mean global signal. For each of these 9 parameters, their derivatives, quadratic terms, and squares of derivatives were included (i.e., the '36 Parameter model': Ciric et al., 2017; Satterthwaite et al., 2013). After preprocessing, the presence of registration, normalisation and other data quality issues were assessed by visual inspection of normalised functional images. Artefacts and data quality issues relating to excessive motion were assessed via visual inspection of variance images. Exclusions relating to the issues flagged during visual inspection before and after preprocessing are noted for each dataset in the Participants subsection of the Methods. Finally, data were temporally smoothed with a zero-mean unit-variance Gaussian filter (approximate cut-off frequency of 9.37 Hz) using BioImageSuite (Joshi et al., 2011).

4.2.4 Functional connectivity network construction

The Shen 268-node (Shen et al., 2013) functional atlas was used to parcellate the fMRI data in both datasets into 268 nodes, in line with previous studies applying CPM (Finn et al., 2015; M. Gao et al., 2020; Greene et al., 2018; Horien et al., 2019). First, the fully preprocessed functional volumes, already in MNI space, were resliced to the Shen functional parcellation image using spm_reslice. Second, using BioImageSuite, the mean time series for each node was calculated as the average time series across all voxels within each node, for each participant. Third, as there was incomplete coverage of the cerebellum for a large proportion of the training set (n = 125; 56.82% of final sample), nodes within the

cerebellum and brainstem were removed from all participants in each dataset. This approach was chosen to avoid excluding a large number of participants from the training set. Nodes within the cerebellum and brainstem were identified based on a previously reported anatomical labelling of the Shen parcellation atlas where the functional parcellations were assigned an anatomical label using the Tailarach atlas (see Table S1 in Salehi et al., 2020). This resulted in the removal of 63 nodes in total, leaving a 430 (volume/time point) * 205 (node) and 200 (volume/time point) * 205 (node) time series for each participant in CR/RANN and TILDA, respectively. Finally, using BioImageSuite, functional connectivity between each pair of nodes was calculated by correlating the average time course between each pair of nodes. The Pearson correlation coefficients were normalised by a Fisher z-transformation. This resulted in a 205 * 205 connectivity matrix for each participant in both datasets.

4.2.5 Measures

CR residual

A residual measure of CR (CR residual) was generated using a linear regression with global cognition as the dependent variable, and age, gender, grey matter volume, hippocampal volume, and mean cortical thickness as the independent variables (see Fig. 4.1). Global cognition, total GM volume, hippocampal volume, and mean cortical thickness were measured in both datasets as described in Sections 3.2.3 and 3.2.4. The linear regressions were conducted separately for the training set and test set. The residuals from these regressions represented the CR residual.

CR proxy

Based on the results of Chapter 3, verbal intelligence was selected as the CR proxy to use in the assessment of face validity of the neuroimaging measure of CR. Verbal intelligence was measured as described in Section 3.2.2.

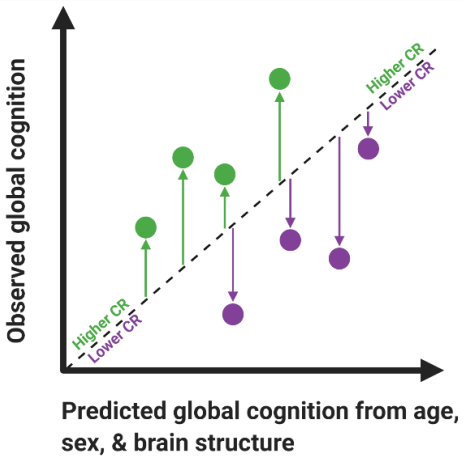


Figure 4.1. Illustration of CR residuals from the regression of global cognition on age, sex, and brain structure. *Positive residuals (green arrows) reflect better cognitive performance than expected given age, sex, and brain structure. Negative residuals (purple arrows) reflect poorer cognitive performance than expected. Higher/more positive residual values reflect higher CR. Image created with BioRender.com and adapted from Fig. 1 (Franzmeier, Hartmann, et al., 2017).*

Table 4.1 Descriptive statistics and statistical comparison of the demographics and studied variables in both datasets.

	CR/RANN (Training Set) n = 220		TILDA (Test Set) n = 294		Group Comparison
	Mean	***SD***	***Mean***	***SD***	***t***
Age (Years)	51.905	17.043	68.301	7.177	-14.837*
Sex (Female/Male)	115/105		152/142		χ^2 = 0.016
Education (Years)	16.209	2.380	17.629†	3.163	-5.537*
NART Score	32.859	9.011	30.432	10.596	5.710*‡
Global Cognition	0.076	0.711	0.094	0.650	-0.300
Grey Matter Volume	0.426	0.028	0.410	0.037	5.453*
Hippocampal Volume	0.005	0.001	0.005	0.001	1.511
Mean Cortical Thickness	2.610	0.114	2.410	0.075	23.892*
CR Residual	0.000	0.593	0.000	0.585	-0.664
Mean FWD (mm)	0.199	0.074	0.250	0.072	-7.745*

*Note: SD = Standard deviation; FWD = framewise displacement; t = t-statistic from independent samples t-test; χ^2 = chi-square statistic; * = p < .001; † = 14 participants missing years of education in TILDA; ‡ = Mean NART scores were normalised separately in each dataset using min-max normalisation before conducting an independent samples t-test as different versions of the NART with different possible maximum scores were used in each dataset.*

4.2.6 Connectome-based prediction of cognitive reserve

To develop a measure of CR using functional connectivity, CPM with a leave-one-out cross-validation (LOOCV) framework was applied to the training set data (Shen et al., 2017). This was conducted in MATLAB (code available here: https://github.com/rorytboyle/flexible_cpm). The CPM method applied here was comprised of the following steps: edge selection, network strength calculation, model fitting, model application, model evaluation. This method iterated through the entire training set n times, with 1 participant left aside in each iteration for the model application step, and n-1 participants retained in each iteration for the feature selection, network strength calculation, and model fitting steps (see Fig. 4.2).

Edge selection: In each iteration, in the n-1 participants, each edge in the 205*205 connectivity matrix was correlated with the CR residual, using a Pearson's correlation. Edges with p-values below an optimised threshold of $p < .0009$ were selected (see *Optimising edge selection threshold)*. Thresholded edges were then separated into positive and negative edges, where positive edges were positively related to the CR residuals, and negative edges were negatively related to the CR residuals.

Network strength calculation: Positive and negative network strength values were then calculated for positive and negative edges separately. Positive Pearson's r values were summed and the sum was divided by 2 to account for matrix symmetry (i.e., the fact that each edge was represented twice in the symmetrical matrix). This was repeated for negative Pearson's r values. A combined network strength value was then calculated by subtracting negative network strength from positive network strength.

Model fitting: 3 separate linear regressions were fitted where positive/negative/combined network strength was the independent variable and the CR residual was the dependent variable. The model parameters (i.e., model intercept and regression coefficient/slope for network strength) were extracted from each regression.

Model application: Network strength values were then calculated for the left out participant. The edge strengths of the selected edges in the n-1 participants in the *Edge selection* step were summed for the left out participant. Summed edge strengths were then added to the fitted linear regression equations using the model parameters from the *Model fitting* step, in order to calculate 3 network strength predicted CR values (*positive network strength predicted CR, negative network strength predicted CR, combined network strength predicted CR)* for the left out participant.

For example:

PosNet CR = PosNet fitted intercept + (PosNet fitted slope * PosNet strength)

Where: PosNet = positive network; PosNet CR = positive network strength predicted CR; PosNet strength = summed edge strength of left out participant's positive network.

Model evaluation: In each iteration, the network strength predicted CR values were stored, as were the selected edges and the fitted model parameters. After n iterations, such that each participant was left out once, all participants had 3 network strength predicted CR values *(positive network strength predicted CR, negative network strength predicted CR, combined network strength predicted CR)*. Model performance, or the accuracy of each prediction, was evaluated for each network strength predicted CR value using 3 metrics: Pearson's correlation between network strength predicted CR values and the CR residual, coefficient of determination (R^2) from a linear regression with network strength predicted CR values as the independent variable and the CR residual as the dependent variable, and the mean absolute error (MAE) between the network strength predicted CR values and the CR residual.

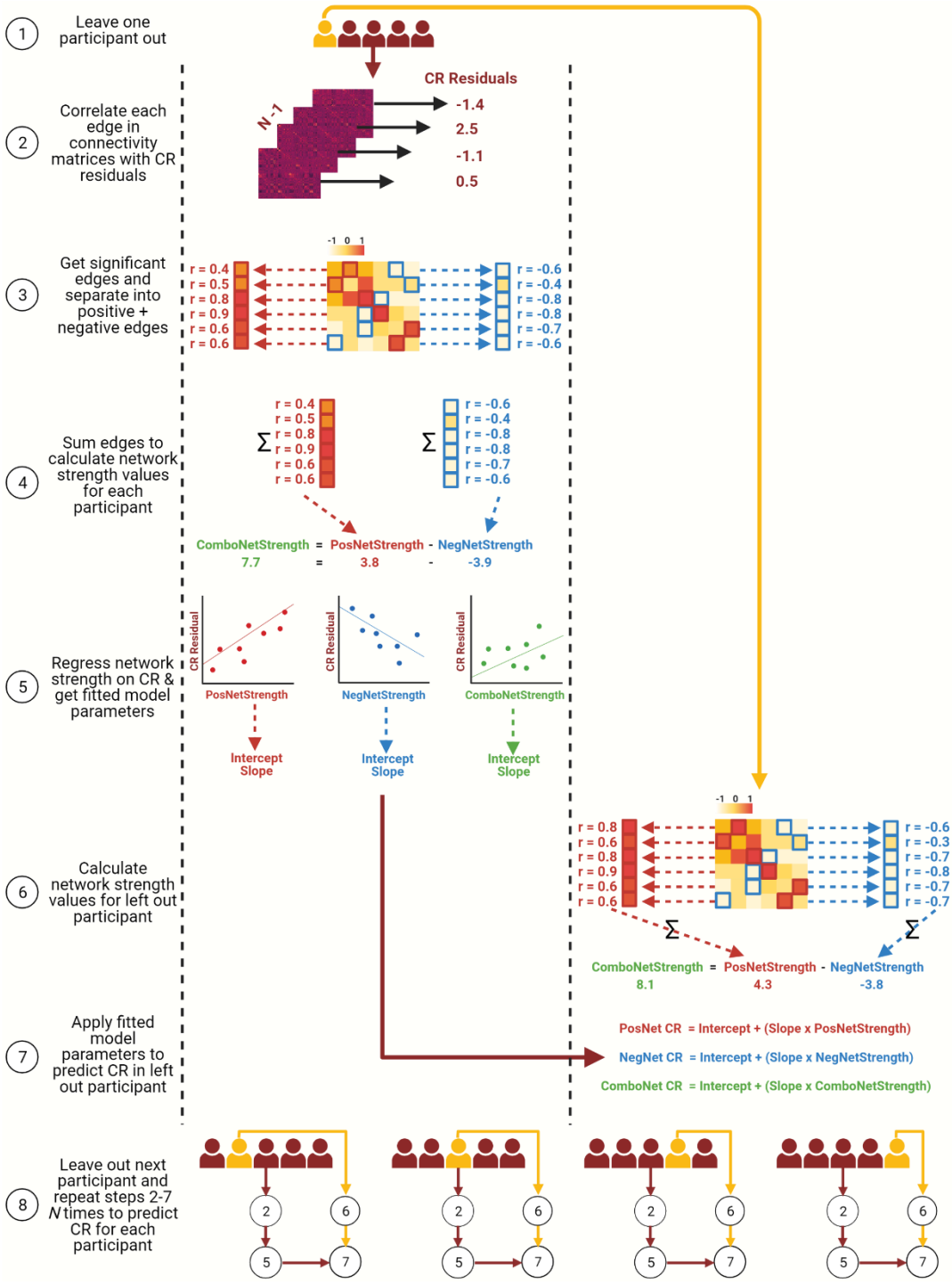


Figure 4.2. Schematic of CPM with LOOCV to predict CR residuals in the training set. *PosNetStrength = positive network strength; NegNetStrength = negative network strength; ComboNetStrength = combined network strength; PosNet CR = positive network strength predicted CR; NegNet CR = negative network strength predicted CR; ComboNet CR = combined network strength predicted CR. Image created with BioRender.com*

4.2.7 Optimising edge selection threshold

As edge-selection p-value thresholds are arbitrary (Greene et al., 2018), the following method was applied in order to select an optimal threshold in a data-driven manner. Starting with a p-value of 0.0001, CPM with LOOCV was repeated in the training set with 100 different p-value thresholds, increasing the p-value by 0.0001 at each step, to a maximum p-value of 0.01. The p-value threshold resulting in the largest Pearson's r between combined network strength predicted CR and the CR residual was selected as the optimal p-value threshold. This p-value threshold was 0.0009 with r = 0.2896 (see Table S7 in 7.3.1 Supplemental Methods).

4.2.8 Assessing validity of network strength predicted CR

To assess the validity of the network strength predicted CR measures as objective neuroimaging measures of CR, two further tests were carried out for each measure. First, the face validity of the network strength predicted CR measures was assessed by establishing their association with a CR proxy, verbal intelligence, using a Pearson's correlation. Network strength predicted CR values were considered to display face validity if they were positively associated with verbal intelligence. Second, the cognitive benefit of the measures were assessed by establishing whether they a) moderated the relationship between mean cortical thickness and global cognition (i.e., displayed a moderation effect), or b) were positively associated with global cognition, independent of mean cortical thickness (i.e., displayed an independent effect). The moderation and independent effects of each measure were assessed using moderated hierarchical regressions where global cognition was regressed on age, sex, and mean cortical thickness in Step 1, with network strength predicted CR added as an independent variable in step 2, and the interaction term for mean cortical thickness and network strength predicted CR included as an independent variable in Step 3. The change in R^2 (i.e., amount of variance explained) from Step 1 to Step 2, and from Step 2 to Step 3 in linear regression models were used to assess the size of the independent and moderation effects of CR proxies, respectively. This analysis was conducted in Python (code available here: https://github.com/rorytboyle/hierarchical_regression). Global cognition and mean cortical thickness were chosen to represent cognitive function and brain structure, respectively, in these regressions as this model accounted for the largest amount of variance explained in both TILDA and CR/RANN in Chapter 3.

4.2.9 External generalisability of connectome-based prediction

To evaluate if network strength predicted CR generalised to independent data, network strength predicted CR values were created in TILDA. First, network strength values were calculated in TILDA. Positive and negative network strength was computed by summing the positive and negative edges selected in each iteration of the LOOCV in the training set. As in the training set, these values were divided by two to account for the symmetrical matrix. Combined network strength was computed as positive network strength minus negative network strength. Second, the regression parameters generated in the training set were averaged across all iterations of the LOOCV and applied to their respective network strength values to calculate network strength predicted CR values. Third, these values were evaluated with respect to their predictive accuracy of the CR residual, using Pearson's correlation, R^2, and MAE. Finally, as described for the training set in *4.2.8 Assessing validity of network strength predicted* CR, the validity of the network strength predicted CR values as measures of CR was assessed for the test set predictions.

4.2.10 Possible confounds in the relationship between connectivity and CR

The '36 Parameter model' used in preprocessing includes global signal regression which has been shown to attenuate motion-related artifacts and noise in the data (Power et al., 2014; Yan et al., 2013). Further steps were also taken to control for this source of noise, given the noted effect of motion on functional connectivity (Power et al., 2012). First, images were visually inspected before and after preprocessing and excluded if motion-related artifacts were present. Second, participants with mean FWD > 0.5mm were excluded after preprocessing. Third, remaining participants who had individual head movements during the scan > 97.5th percentile of individual head movements across all participants were excluded. To assess whether functional connectivity was then related to head motion, the correlation between mean FWD and global functional connectivity (the average of functional connectivity in the upper triangle of the connectivity matrix) was assessed in each dataset. While the correlation in CR/RANN was not significant ($r = .03$, $p = 0.62$), global functional connectivity in TILDA was significantly related to mean FWD ($r = 0.23$, $p < .001$). As such, a final FWD threshold was applied such that participants with mean FWD > 0.4mm ($n = 52$) were excluded. After removal of these participants, the correlation between global functional connectivity and mean FWD did not change ($r = 0.23$, $p < .001$). Although the mean FWD > 0.4mm threshold did not reduce the correlation between mean FWD and global functional connectivity, there were fewer edges correlated to mean FWD at this threshold (18% fewer correlated edges). This threshold was chosen as it removed the remaining participants with

the highest average head motion in TILDA (i.e., n = 52 with FWD <= 0.5 mm and > 0.4 mm) but retained a large final N. While a more conservative threshold of FWD > 0.2 mm has been previously used (M. Gao et al., 2020), this would have resulted in the exclusion of a further 95 and 239 participants from the final CR/RANN and TILDA samples respectively.

To ensure that the network strength predicted CR measures were not confounded by head motion, additional checks were conducted at the analysis stage to assess whether the CR residual was correlated with mean FWD and whether the network strength predicted CR measures were correlated with mean FWD. The possible influence of other confounds, specifically age and sex, on functional connectivity were assessed by correlating age with global functional connectivity and assessing gender differences in global functional connectivity. To ensure that the network strength predicted CR measures were not confounded by covariates, CPM was repeated including the covariates age, sex, and mean FWD at the *feature selection* step. This was implemented by using a partial correlation to relate functional connectivity in each edge to the CR residual, including age, sex, and mean FWD as covariates.

4.2.11 Supplementary analyses

As a further validation of the main results, the analysis was repeated in the training set using k-fold cross-validation schemes (5-fold and 10-fold cross-validation) instead of LOOCV. While LOOCV is the standard cross-validation scheme in studies applying CPM (Greene et al., 2018; Rosenberg et al., 2016), LOOCV can result in overly optimistic estimates of model accuracy and can provide more variable predictions when applied to external datasets (Dwyer et al., 2018; Varoquaux et al., 2017). As such, k-fold cross-validation has been recommended as a preferable cross-validation scheme (Poldrack et al., 2020; Varoquaux et al., 2017). In k-fold cross-validation, the data is randomly split into k subsets. One subset is then set aside (as the left out participant is set aside in LOOCV) as a test set for the *model application* step in CPM and the k-1 subsets are used to fit the model (i.e., *edge selection, network strength calculation, model fitting* steps in CPM). 100 iterations were run of each k-fold model, and the models were evaluated by averaging their Pearson's r and R^2 across datasets across the 100 iterations.

4.3 Results

4.3.1 Creation of cognitive reserve residual

The CR residuals accounted for 30% and 19% of the variance in global cognition in CR/RANN and TILDA respectively (see Table 4.2). In Chapter 3, verbal intelligence explained a mean additional 16.8% of the variance in global cognition, across both datasets, after accounting for age, sex, and hippocampal volume. This was the largest effect identified. The CR residuals accounted for a mean additional 57.9% of the variance in global cognition across both datasets (see Table S8 in 7.3.2 Supplemental Results). As such, the CR residuals were a preferable target variable for CPM as they reflected a larger amount of the variance in global cognition that was not explained by brain structure or demographics. In both datasets, CR residuals were approximately normally distributed and were significantly positively correlated with NART scores (see Fig. 4.3). As such, CR residuals in both datasets displayed face validity as measures of CR.

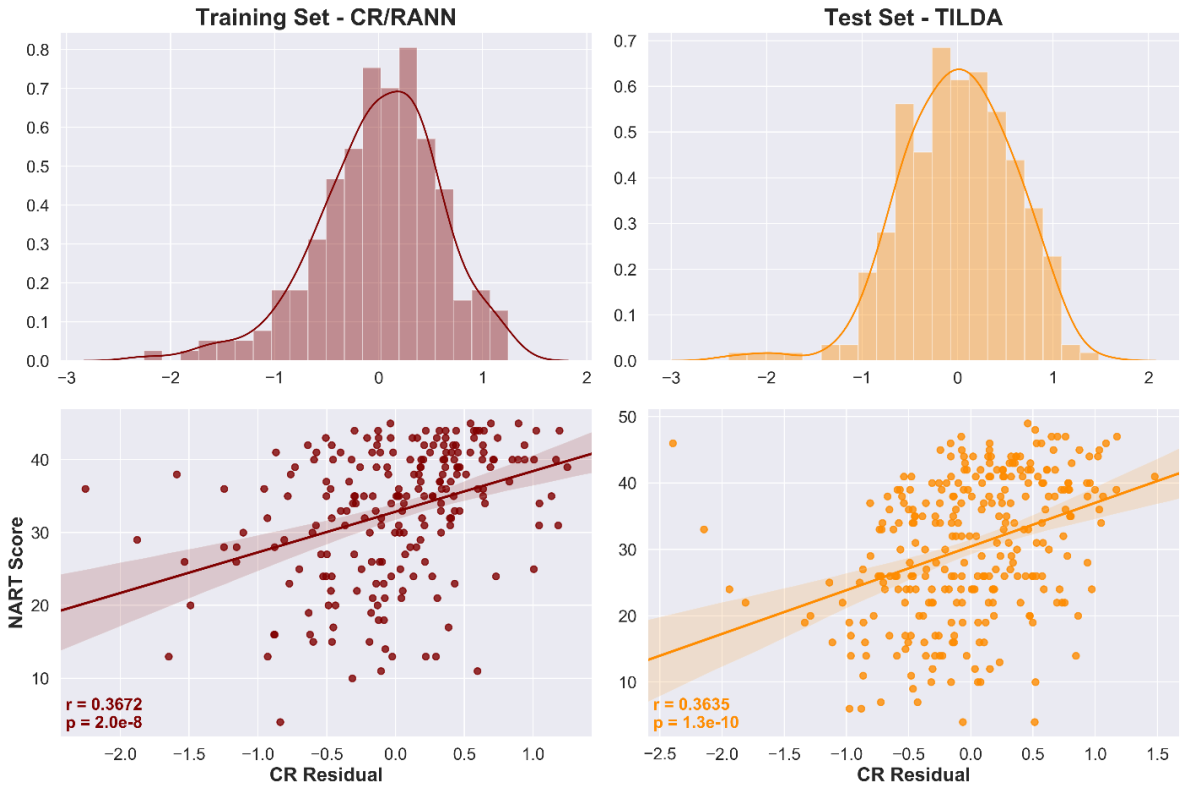


Figure 4.3. Normality and face validity of CR residuals. *Histograms of CR residuals with kernel density estimates (top row) show that the CR residuals are approximately normally distributed. Scatterplots with regression lines (bottom row) show significant positive relationships between CR residuals and NART scores, demonstrating the face validity of CR residuals as CR measures.*

4.3.2 Connectome-based prediction of cognitive reserve

The connectome-based predictive models significantly predicted the CR residuals of novel participants (i.e., each left-out participant in the LOOCV) from task-based functional connectivity data in the training set (see Fig. 4.4 and Table 4.3). The combined network strength model had the best predictive accuracy as it had the highest R value for the correlation between network strength predicted CR and CR residuals, the highest R^2, and the lowest MAE.

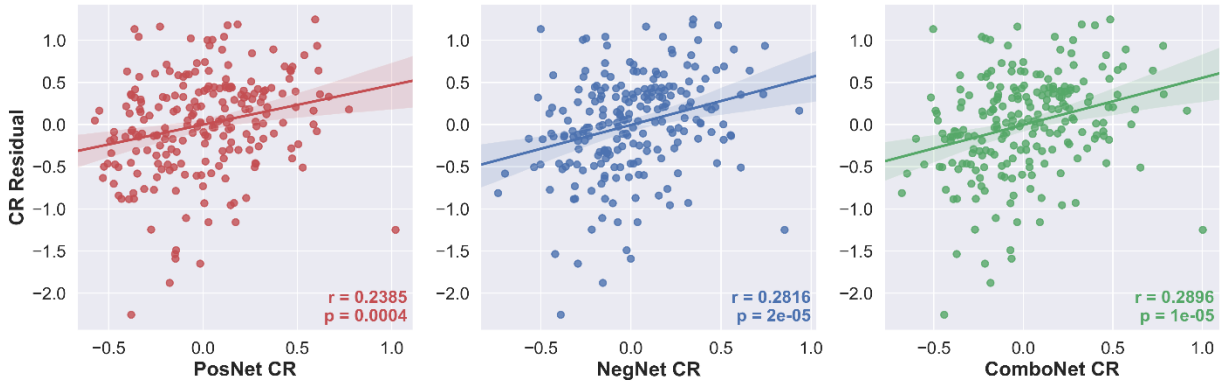


Figure 4.4. CR residual vs positive-, negative-, and combined network strength predicted CR in the training set. *PosNet CR = Positive network strength predicted CR; NegNet CR = Negative network strength predicted CR; ComboNet CR = Combined network strength predicted CR.*

4.3.3 Validation of network strength predicted CR in the training set

The network strength predicted CR values generated by the connectome-based predictive models displayed face validity as measures of CR, as all models were significantly positively correlated with a CR proxy – verbal intelligence as measured by NART scores (see Fig. 4.5 and Table 4.4). The network strength predicted CR values also satisfied the cognitive benefit criterion for measures of CR, as all were positively associated with global cognition, controlling for the effects of mean cortical thickness, age, and sex (see Table 4.4).

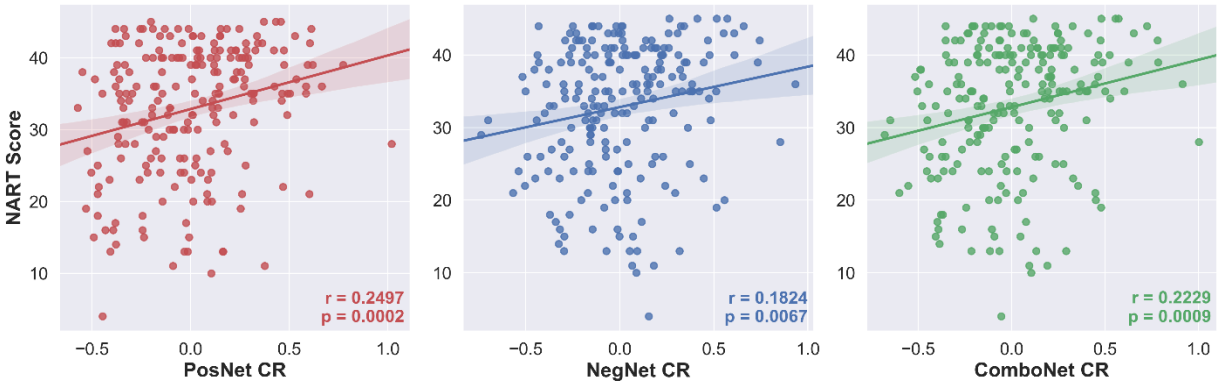


Figure 4.5 NART scores vs positive-, negative-, and combined network strength predicted CR in the training set. *PosNet CR = Positive network strength predicted CR; NegNet CR = Negative network strength predicted CR; ComboNet CR = Combined network strength predicted CR.*

4.3.4 Motion control and confounds

A Pearson's correlation indicated that, in the training set, mean FWD was not significantly associated with the target variable (i.e., the CR residual, r = -.0786, p = 0.2459). However, mean FWD was significantly negatively associated with positive network strength predicted CR (r = -0.1542, p = 0.0221); negative network strength predicted CR (r = -0.1390, p = 0.0394); and combined network strength predicted CR (r = -0.1551, p = 0.0214). In the test set, mean FWD was not significantly associated with the CR residual (r = -0.0270, p = 0.6442); positive network strength predicted CR (r = 0.0236, p = 0.6875); negative network strength predicted CR (r = 0.0365, p = 0.5333); and combined network strength predicted CR (r = 0.0441, p = 0.4508).

A Pearson's correlation further established that age was significantly positively associated with global functional connectivity in the training set, r = 0.1430, p = 0.0340. An independent samples t-test showed that there were no significant differences in the training set between males and females in mean global functional connectivity (t = -.1099, p = 0.2734). However, in the test set, age was significantly negatively associated with global functional connectivity (r = - 0.1842, p = 0.0015) and there was a significant sex difference in global functional connectivity (t = 2.8917, p = 0.0041) such that males (mean = 0.01, SD = 0.0063) had higher mean global functional connectivity than females (mean = 0.0081, SD = 0.0049).

Given the associations between network strength predicted CR and mean FWD in the training set, age and global functional connectivity in both datasets, and the sex differences in global functional connectivity in the test set, these three variables (FWD, age, and sex) were considered confounds. As such, an adjusted connectome-based predictive model was applied where, age, sex, and mean FWD were included as covariates at the feature selection stage, using a partial correlation between functional connectivity in each edge and the CR residual. After this adjustment, mean FWD was no longer significantly associated with negative network strength predicted CR ($r = -0.0728$, $p = 0.2822$) or combined network strength predicted CR ($r = -0.1023$, $p = 0.1302$). While mean FWD was still significantly associated with positive network strength predicted CR ($r = -0.1339$, $p = 0.0474$), the strength of the association was reduced as compared to the original (unadjusted) model. In the test set, as was also the case prior to adjusting for possible confounds, mean FWD was not significantly associated with positive network strength predicted CR ($r = 0.0473$, $p = 0.4188$); negative network strength predicted CR ($r = 0.0377$, $p = 0.5199$); and combined network strength predicted CR ($r = 0.0587$, $p = 0.3154$).

The connectome-based predictive models remained statistically significant when adjusting for age, sex, and mean FWD at the feature selection stage (see Fig. 4.6 and Table 4.3). Furthermore, network strength predicted CR values generated from the adjusted connectome-based predictive models also displayed face validity as, and satisfied the cognitive benefit criterion for, measures of CR (see Table 4.4 and Fig. 4.7).

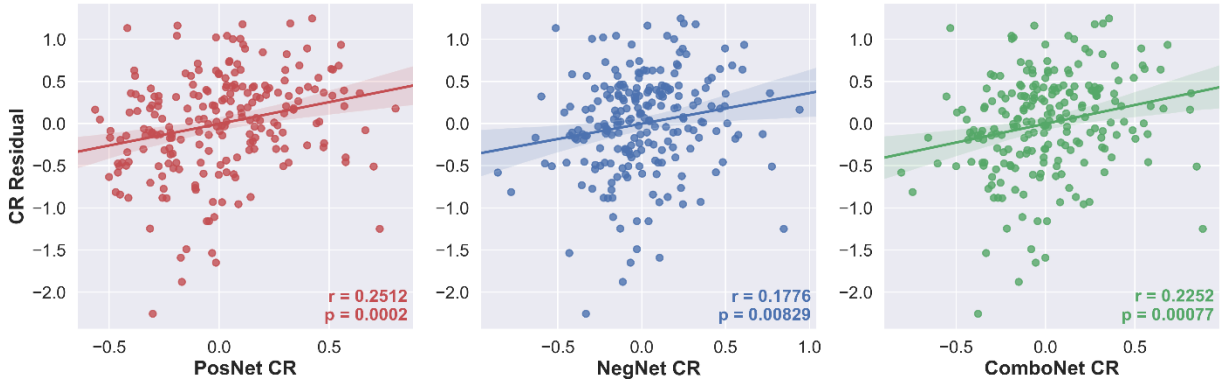


Figure 4.6. CR residual vs positive-, negative-, and combined network strength predicted CR using adjusted CPM in the training set. *PosNet CR = Positive network strength predicted CR; NegNet CR = Negative network strength predicted CR; ComboNet CR = Combined network strength predicted CR.*

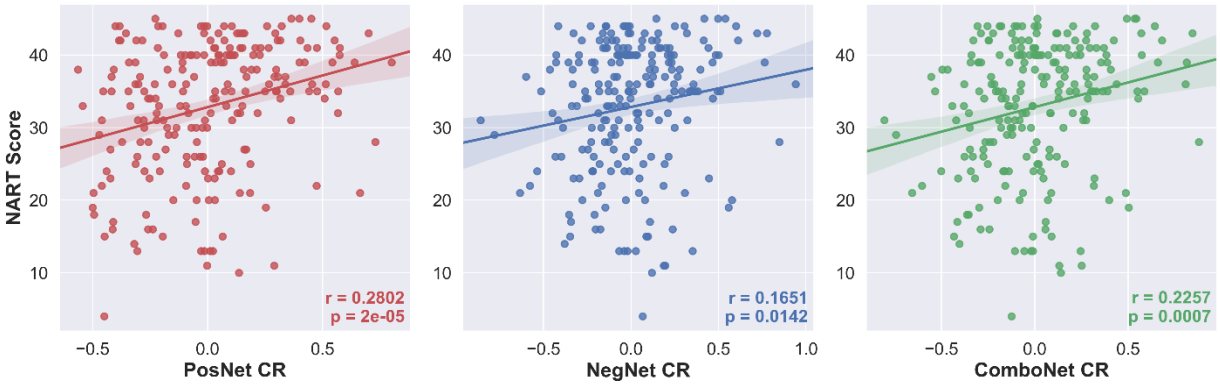


Figure 4.7. NART scores vs positive-, negative-, and combined network strength predicted CR using adjusted CPM in the training set. *PosNet CR = Positive network strength predicted CR; NegNet CR = Negative network strength predicted CR; ComboNet CR = Combined network strength predicted CR.*

Table 4.2 Results of multiple regressions used to create CR residuals in the training and test datasets.

Dataset	Model Statistics			GM Volume	Hippocampal Volume	Mean Cortical Thickness	Sex	Age
	n	R^2	*f*	*β*	*β*	*β*	*β*	*β*
CR/RANN (Training)	220	0.304	18.723*	.088	.001	.080	-.018	-.433*
TILDA (Test)	294	0.190	13.483*	-.045	.077	.040	.095	-.395*

*Note: * = p < .0001.*

Table 4.3 CPM performance for prediction of CR residuals in the training set.

	Positive Network Strength				Negative Network Strength				Combined Network Strength			
	r	*p*	R^2	*MAE*	*r*	*p*	R^2	*MAE*	*r*	*p*	R^2	*MAE*
Original	.239	3.6e-4	.057	.453	.282	2.3e-5	.079	.447	.290	1.3e-5	.084	.444
Adjusted	.251	1.7e-4	.063	.452	.178	.008	.032	.475	.225	7.6e-4	.051	.463

Table 4.4 Validation of network strength predicted CR in the training set.

	Positive Network Strength						Negative Network Strength						Combined Network Strength					
	Face Validity		*Cognitive Benefit*				*Face Validity*		*Cognitive Benefit*				*Face Validity*		*Cognitive Benefit*			
	NART		*Ind. Effect*		*Mod. Effect*		*NART*		*Ind. Effect*		*Mod. Effect*		*NART*		*Ind. Effect*		*Mod. Effect*	
	r	*p*	R^2	*p*	R^2	*p*	*r*	*p*	R^2	*p*	R^2	*p*	*r*	*p*	R^2	*p*	R^2	*p*
Original	.250	1.8e-4	.043	2.2e-4	2.5e-5	.928	.182	.007	.060	1.2e-5	.005	.184	.223	8.7e-4	.063	6.2e-6	.026	.348
Adjusted	.280	2.5e-5	.030	.022	1.7e-6	.981	.165	.014	.056	2.3e-5	.003	.329	.226	7.4e-4	.054	3.2e-5	.002	.446

Note: NART = Correlation of predicted values with NART scores; Ind. Effect = Independent effect of predicted values on global cognition, controlling for age, sex, and mean cortical thickness; Mod. Effect = Moderation effect of predicted values on relationship between brain structure and global cognition.

4.3.5 Functional network anatomy

Both positive and negative CR networks were sparse with 9 common edges (0.04% of total edges) that were statistically significant in every iteration of the positive network, and 12 common edges (0.06% of total edges) that were statistically significant in every iteration of the negative network. Fig. 4.8 shows circle plots visualising the edges that constituted the positive and negative networks/connectomes. Fig. 4.9 shows these same edges visualised within the brain.

The left dorsolateral prefrontal cortex contained two key nodes within the positive CR network, with number of edges (k) = 3 and 2 respectively (see Table 4.5). Other nodes containing more than one edge in the positive CR network were located in the left premotor/supplementary cortex (k = 2) and the right angular gyrus (k = 2). The remaining nodes in the positive CR network were connected by single edges. The left temporal pole contained the most highly connected node in the negative CR network (k = 5), followed by the right angular gyrus (k = 3) and the left precentral gyrus (k = 2; see Table 4.6). The remaining nodes in the negative CR network were connected by single edges.

The connectivity patterns of the positive and negative CR networks were further explored in relation to the distribution of connectivity within and between 10 different canonical functional networks as previously defined in an independent sample (Noble et al., 2017). Noble et al. labelled these networks as follows: medial frontal network (MFN), FPN, DMN, motor network, visual I network, visual II network, visual association network, limbic network, basal ganglia network, and cerebellar network. In BioImageSuite Web and more recent studies by the same group (Greene et al., 2018, 2020), the limbic network is referred to as the salience network and the basal ganglia network is referred to as the subcortical network (see Fig. S3 in 7.3.2 Supplemental Results). The latter definitions are used here. It should also be noted that while the cerebellar network is presented in figures here, it was not included in any models analysed. The positive CR network was largely characterised by connectivity within the FPN, connectivity of the FPN to other networks, and of the motor network to other networks (see Fig. 4.10). This pattern of connectivity remained the same when adjusting for age, sex, and mean FWD, although there were a lower number of common edges across all iterations of the positive network. The negative CR network was mostly characterised by connectivity of a single node in the MFN, the left temporal pole, to other networks, connectivity within the motor network, and connectivity of the motor network

to other networks. There was a similar, but reduced, pattern of connectivity in models adjusting for age, sex, and mean FWD.

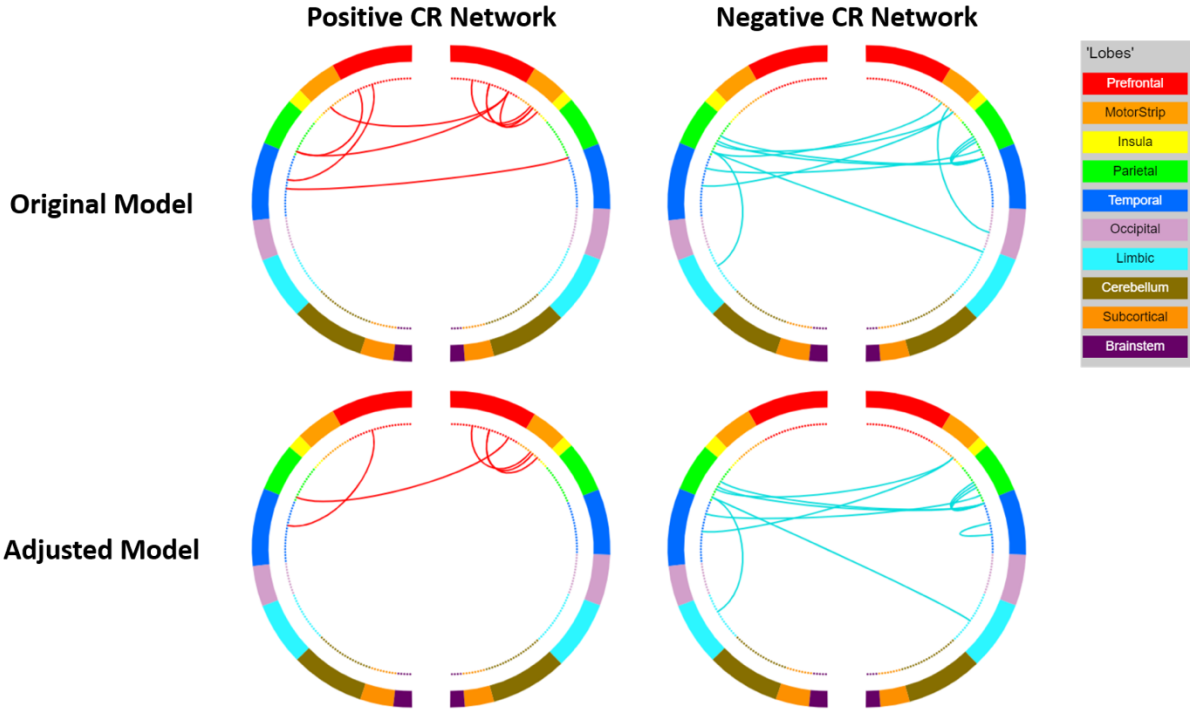


Figure 4.8. Circle plots illustrating the positive and negative CR connectomes. *Positive connections (red) and negative connections (blue) in original CPM (top panel) and adjusted CPM (bottom panel), controlling for age, sex, and mean FWD. These circle plots are inverted such that the right side of each plot corresponds to the left hemisphere and the left side to the right hemisphere. Image created with BioImageSuite Web.*

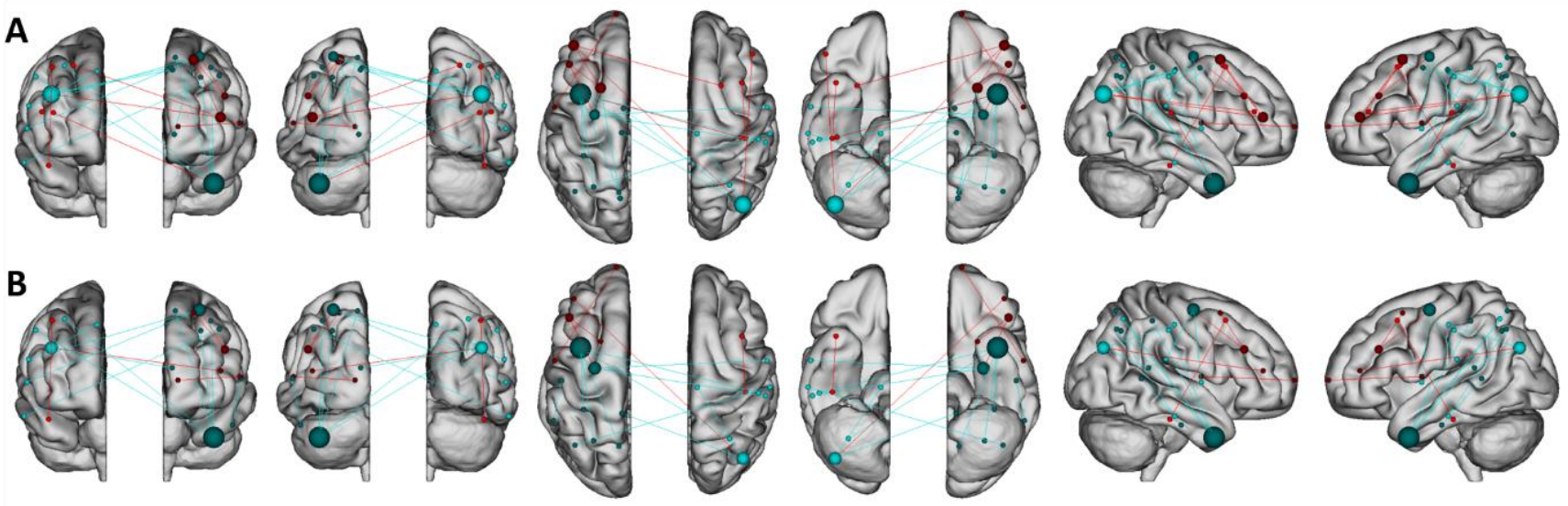


Figure 4.9. Glass brain visualising the patterns of connectivity within the brain. *Positive connections (red) and negative connections (blue) in original CPM (A, top panel) and adjusted CPM (B, bottom panel), controlling for age, sex, and mean FWD. Image created with BioImageSuite Web.*

Table 4.5. Positive CR network functional neuroanatomy with nodes sorted by degree strength.

#	BIS Label (BA)	Tailarach Label	K	Net-work	Lobe (L/R)	MNI co-ordinates		
						x	y	z
154	Dorsolateral prefrontal cortex (46)	Middle Frontal Gyrus	3	FP	L Prefrontal	-42.97	42.04	11.04
164	Premotor/supplementary motor cortex (6)	Middle Frontal Gyrus	2	FP	L MotorStrip	-23.22	10.66	53.61
147	Dorsolateral prefrontal cortex (9)	Middle Frontal Gyrus	2	FP	L Prefrontal	-46.12	28.15	26.79
49	Angular Gyrus (39)	Middle Temporal Gyrus	2	DMN	R Parietal	41.39	-75.34	27.98
185	Temporal Pole (38)	Superior Temporal Gyrus	1	MF	L Temporal	-38.01	6.07	-37.86
166	Premotor/supplementary motor cortex (6)	Precentral Gyrus	1	Mot	L MotorStrip	-27.58	-9.08	55.86
163	Premotor/supplementary motor cortex (6)	Superior Temporal Gyrus	1	Mot	L MotorStrip	-56.98	-3.43	6.82
141	Anterior Prefrontal Cortex (10)	Superior Frontal Gyrus	1	DMN	L Prefrontal	-11.7	65.09	4.18
62	Primary Auditory Cortex (41)	Insula	1	Mot	R Temporal	39.86	-25.56	14.38
59	Inferior Temporal Gyrus (20)	Sub-gyral Temporal Lobe	1	VAs	R Temporal	43.36	-26.48	-24.63
30	Premotor/supplementary motor cortex (6)	Sub-gyral Frontal Lobe	1	FP	R MotorStrip	25.22	12.41	49.39
19	Dorsolateral prefrontal cortex (46)	Middle Frontal Gyrus	1	FP	R Prefrontal	48.29	35.68	15.15
14	Frontal Eye Fields (8)	Middle Frontal Gyrus	1	FP	R Prefrontal	40.68	14.51	48.21

Note: # = Node number; BIS Label (BA) = Brodmann Area label and number for node as listed in BioImageSuite; Tailarach Label = Anatomic label for node from Tailarach atlas (Salehi et al., 2020); K = degree strength (i.e., number of connections) of node in positive network; Network = canonical networks defined in an independent sample (Noble et al., 2017); Lobe (L/R) = Left or right hemisphere and Lobe as listed in BioImageSuite; MNI co-ordinates= Montreal Neurological Institute co-ordinates; FP = Frontoparietal Network; DMN = Default Mode Network; MF = Medial Frontal Network; Mot = Motor Network; VAs = Visual Association Network.

Table 4.6. Negative CR network functional neuroanatomy with nodes sorted by degree strength.

#	BIS Label (BA)	Tailarach Label	K	Net-work	Lobe (L/R)	MNI co-ordinates		
						x	y	z
185	Temporal Pole (38)	Superior Temporal Gyrus	5	MF	L Temporal	-38.01	6.07	-37.86
49	Angular Gyrus (39)	Middle Temporal gyrus	3	DMN	R Parietal	41.39	-75.34	27.98
166	Premotor/supplementary motor cortex (6)	Precentral Gyrus	2	Mot	L MotorStrip	-27.58	-9.08	55.86
164	Premotor/supplementary motor cortex (6)	Middle Frontal Gyrus	1	FP	L MotorStrip	-23.22	10.66	53.61
218	Premotor/supplementary motor cortex (6)	Cingulate Gyrus	1	Mot	L Limbic	-7.75	-22.37	46.05
211	Secondary Visual Cortex (18)	Lingual Gyrus	1	Vis I	L Occipital	-8.88	-70.65	-1.67
182	Angular Gyrus (39)	Angular Gyrus	1	FP	L Parietal	-42.05	-65.62	41.73
179	Supramarginal Gyrus (40)	Sub-gyral Parietal Lobe	1	Mot	L Parietal	-35.72	-39.37	47.75
178	Visual Motor Co-ordination (7)	Precuneus	1	SAL	L Parietal	-9.83	-66.34	55.14
177	Visual Motor Co-ordination (7)	Sub-gyral Parietal Lobe	1	VAs	L Parietal	-28.41	-62.35	40.42
161	Premotor/supplementary motor cortex (6)	Cingulate Gyrus	1	Mot	L MotorStrip	-6.47	-4.31	47.6
89	Dorsal Posterior Cingulate Cortex (31)	Cingulate Gyrus	1	Mot	R Limbic	7.83	-23.07	44.93
61	Primary Auditory Cortex (41)	Superior Temporal Gyrus	1	Mot	R Temporal	59.18	-3.36	2.74
55	Middle Temporal Gyrus (21)	Inferior Temporal Gyrus	1	FP	R Temporal	61.28	-22.87	-22.38
46	Supramarginal Gyrus (40)	Inferior Parietal Lobule	1	Mot	R Parietal	58	-29.28	19.53
45	Supramarginal Gyrus (40)	Inferior Parietal Lobule	1	Mot	R Parietal	52.84	-27.25	40.93
43	Visual Motor Co-ordination (7)	Precuneus	1	VAs	R Parietal	31.62	-60.79	49.21

Note: # = Node number; BIS Label (BA) = Brodmann Area label and number for node as listed in BioImageSuite; Tailarach Label = Anatomic label for node from Tailarach atlas (Salehi et al., 2020); K = degree strength (i.e., number of connections) of node in negative network; Network = canonical networks defined in an independent sample (Noble et al., 2017); Lobe (L/R) = Left or right hemisphere and Lobe as listed in BioImageSuite; MNI co-ordinates= Montreal Neurological Institute co-ordinates; MF = Medial Frontal Network; DMN = Default Mode Network; Mot = Motor Network; FPN = Frontoparietal Network; Vis I = Visual I Network; SAL = Salience Network; VAs = Visual Association Network.

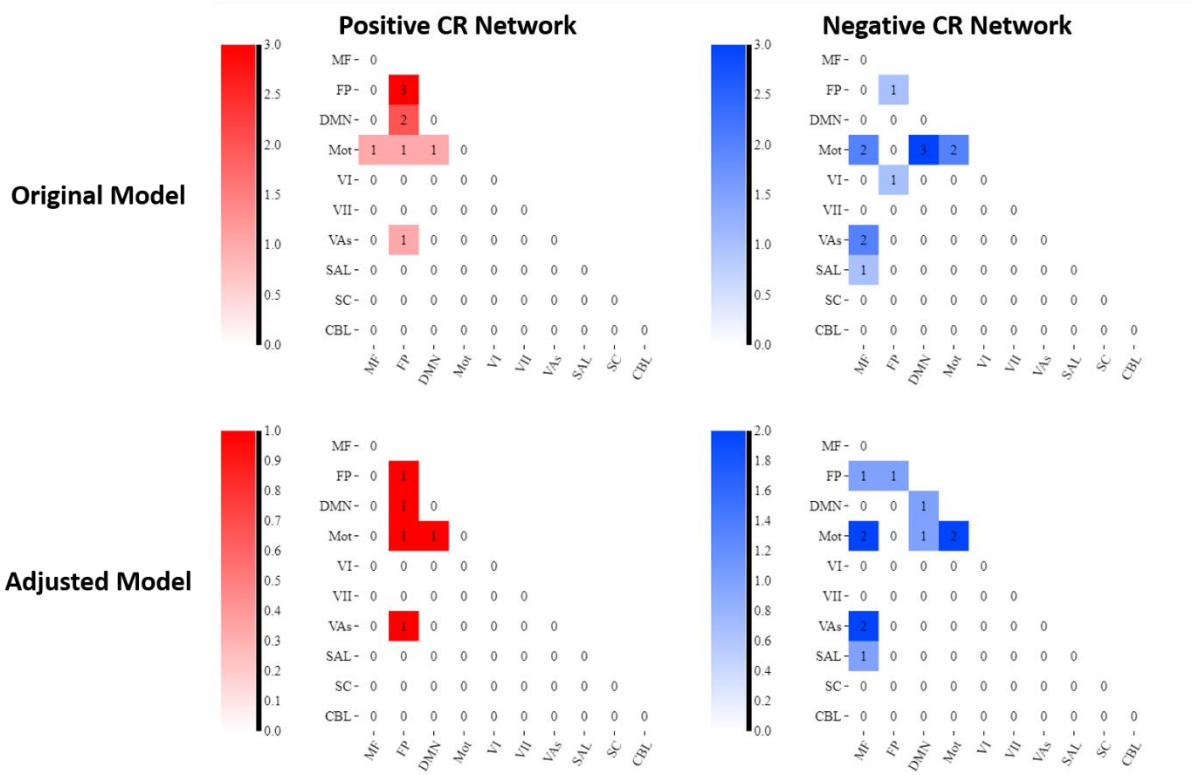


Figure 4.10. Connectivity matrices summarising the connectivity patterns within and between different functional networks. *Note: Darker shades represent stronger connectivity (i.e., larger number of edges in that network). MF = Medial Frontal Network; FP = Frontoparietal Network; DMN = Default Mode Network; Mot = Motor Network; Vis I = Visual I Network; Vis II = Visual II Network; VAs = Visual Association Network; SAL = Salience Network; SC = Subcortical Network; CBL = Cerebellar Network. Image created with BioImageSuite Web.*

4.3.6 Generalisability of network strength predicted CR: Application to the test set

The network strength predicted CR values in the test set, which were generated by network strength models calculated over the same edges identified in the training set, were not related to the CR residual (see Fig. 4.11A and Table 4.7). While the correlations of the CR residual with negative- and combined-network strength predicted CR had p-values < 0.05, these relationships are not meaningful as the negative correlation indicates that the network strength predicted CR values are negatively related to the CR residual in the test set (Ren et al., 2021). The results were similar when CPM, controlling for age, sex, and mean FWD,

was applied to the test set (see Fig. 4.11B and Table 4.7). These results show that the predictive models developed in the training set on task-based fMRI data did not generalise to resting-state data in the test set.

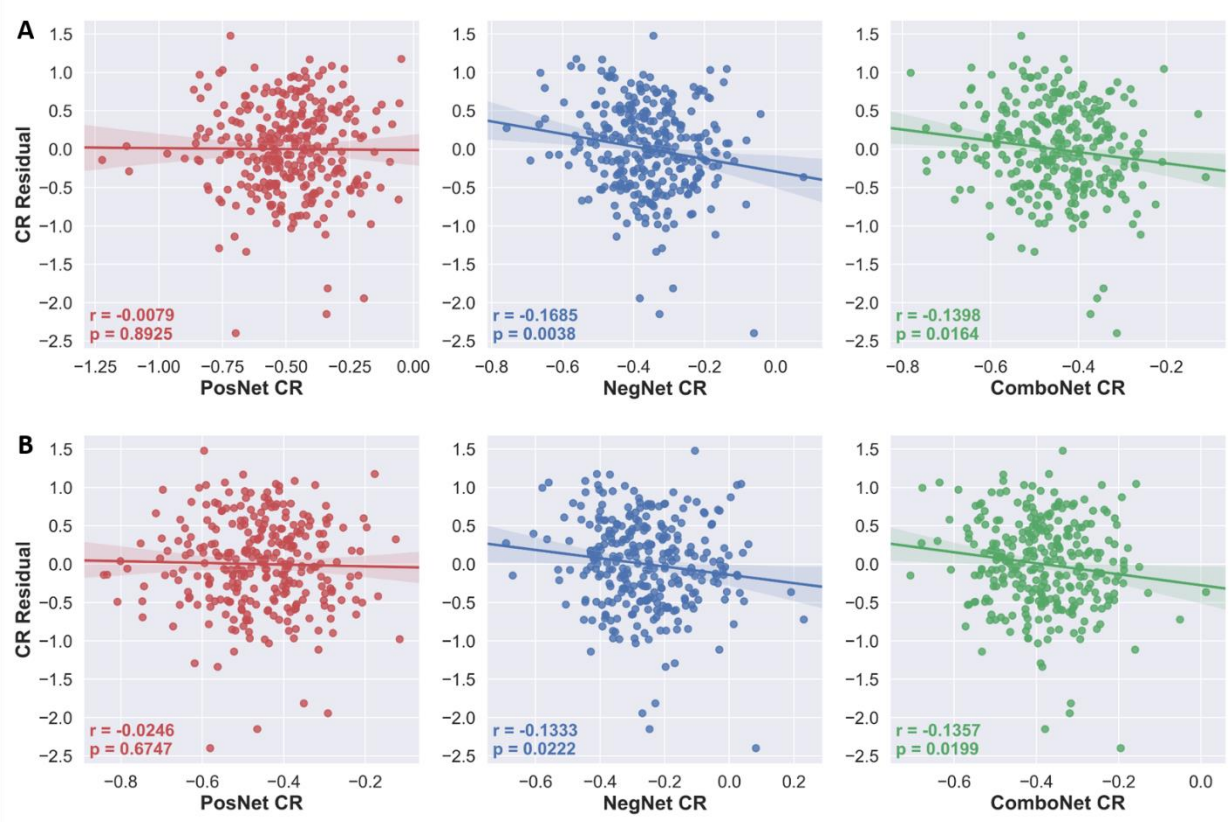


Figure 4.11. CR residual vs positive-, negative-, and combined network strength predicted CR in the test set. ***A*** = *predicted CR values from original network strength models;* ***B*** = *predicted CR values from adjusted network strength models, controlling for age, sex, and mean FWD. PosNet CR = Positive network strength predicted CR; NegNet CR = Negative network strength predicted CR; ComboNet CR = Combined network strength predicted CR.*

4.3.7 Validation of network strength predicted CR in the test set

The network strength predicted CR values generated by the connectome-based predictive models did not display face validity as measures of CR, as they were not significantly positively correlated with a CR proxy, verbal intelligence as measured by NART scores (see Fig. 4.12A and Table 4.8). These results were also observed in network strength models controlling for age, sex, and mean FWD (see Fig. 4.12B). Furthermore, the network strength predicted CR values did not satisfy the cognitive benefit criterion for measures of

CR, as they did not moderate the relationship between mean cortical thickness and global cognition nor were they significantly positively associated with global cognition, controlling for the effects of mean cortical thickness, age, and sex (see Table 4.8). The adjusted predictive models similarly did not satisfy the cognitive benefit criterion.

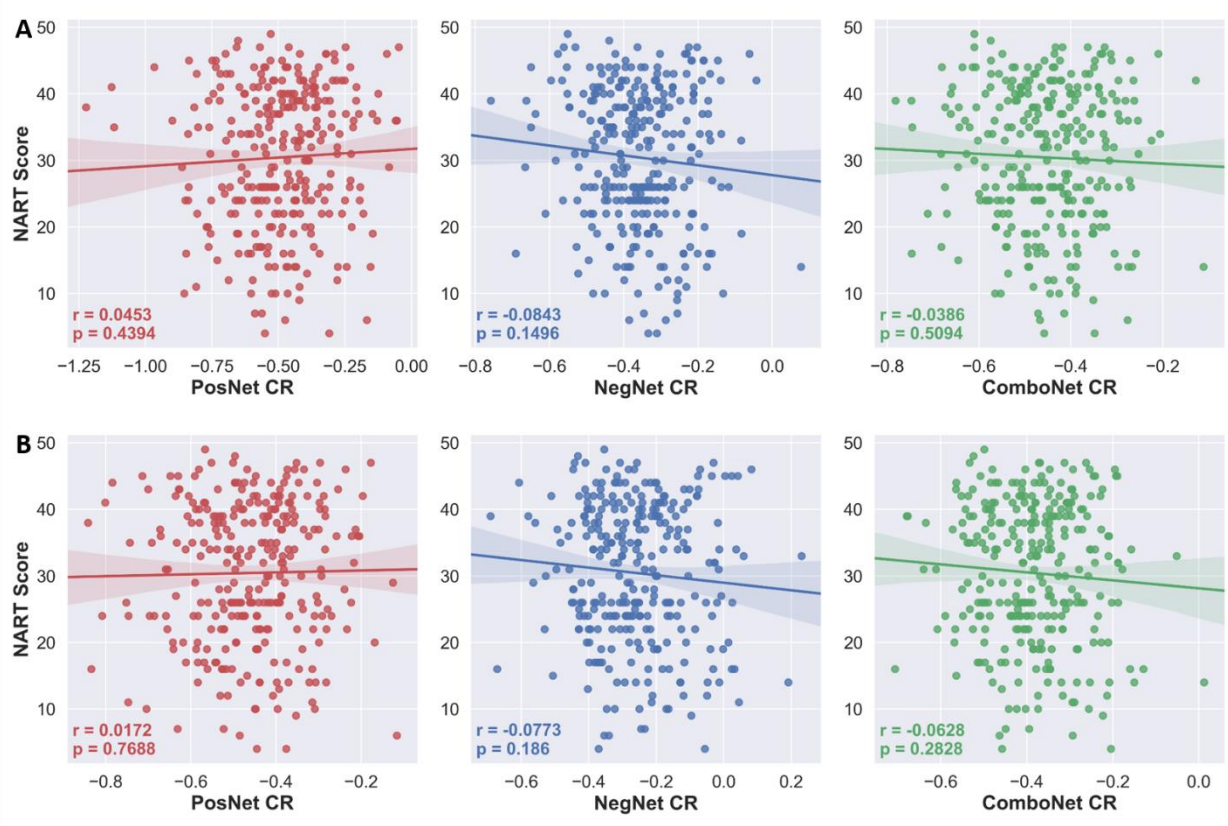


Figure 4.12. NART scores vs positive-, negative-, and combined network strength predicted CR in the test set. ***A*** *= predicted CR values from original network strength models;* ***B*** *= predicted CR values from adjusted network strength models, controlling for age, sex, and mean FWD. PosNet CR = Positive network strength predicted CR; NegNet CR = Negative network strength predicted CR; ComboNet CR = Combined network strength predicted CR.*

4.3.8 Supplementary analyses

In the training set, the negative and combined network strength predicted CR values remained statistically significant when applying k-fold cross-validation methods instead of LOOCV (see Table S9 in 7.3.2 Supplemental Results). However, the positive network strength predicted CR values did not remain statistically significant in the training set. Application of the 5-fold and 10-fold cross-validation models to the test set generated negative- and combined-network strength predicted CR values that were statistically

significant but negatively correlated with the CR residual (see Table S9 in 7.3.2 Supplemental Results). As was observed using LOOCV, the negative correlations between network strength predicted CR values and the CR residual indicated that these predictions were not meaningful and that the models did not generalise.

To investigate why the predictive models did not generalise to resting-state data in an independent dataset, a further exploratory analysis was conducted. To investigate whether differences in the age ranges of the datasets may have explained the inability to generalise, the training set was restricted to adults aged 50 years or older in order to more closely match the age range of the training set. This reduced the sample size of the training set to n = 128 (mean age = 64.42 years, SD = 8.49 years). CPM was then repeated, adjusting for age, sex, and FWD as covariates at the feature selection stage. In the training set, the CR residuals of novel participants were only significantly predicted by negative and combined network strength predicted CR values (see Table S10 in 7.3.2 Supplemental Results). In the test set, the CR residuals were not significantly predicted by positive and combined network strength predicted CR (see Table S10 in 7.3.2 Supplemental Results). While the correlation between CR residuals and negative network strength predicted CR was statistically significant, the correlation was negative, indicating that the models failed to generalise to resting-state data in the test set. This suggested that differences in the age range in each dataset were not responsible for the inability of the models to generalise to the test set.

A negative correlation between network strength predicted CR values and observed values (i.e., the CR residual) is interpreted as the predicted values failing to explain any variance in the observed values (Greene et al., 2018) and is therefore considered meaningless (Ren et al., 2021). However, to further explore the negative correlation between network strength predicted CR values and the CR residual in the test set, the correlation between the CR residual and the thresholded edges in the test set was investigated (see Table S11 in 7.3.2 Supplemental Results). 12 edges were selected in the negative network (i.e., edges where connectivity was negatively correlated to the CR residual) in the training set. The average correlation between connectivity in these 12 edges and the CR residual in the training set was r = - 0.2728. However, in the test set, none of these edges were negatively correlated with the CR residual and the average correlation in the test set was r = 0.0696. In fact, 3 of the thresholded edges (25%) were significantly positively correlated with the CR residual.

Table 4.7 CPM performance for prediction of CR residuals in the test set.

	Positive Network Strength				Negative Network Strength				Combined Network Strength			
	r	***p***	***R^2***	***MAE***	***r***	***p***	***R^2***	***MAE***	***r***	***p***	***R^2***	***MAE***
Original	-.008	.893	6.3e-5	.648	-.169	.004	.028	.570	-.140	.016	.020	.617
Adjusted	-.025	.675	.001	.620	-.133	.022	.018	.536	-.136	.020	.018	.580

Table 4.8 Validation of network strength predicted CR in the test set.

	Positive Network Strength						***Negative Network Strength***						***Combined Network Strength***					
	Face Validity		***Cognitive Benefit***				***Face Validity***		***Cognitive Benefit***				***Face Validity***		***Cognitive Benefit***			
	NART		***Ind. Effect***		***Mod. Effect***		***NART***		***Ind. Effect***		***Mod. Effect***		***NART***		***Ind. Effect***		***Mod. Effect***	
	r	***p***	***R^2***	***p***	***R^2***	***p***	***r***	***p***	***R^2***	***p***	***R^2***	***p***	***r***	***p***	***R^2***	***p***	***R^2***	***p***
Original	.045	.439	2.9e-5	.918	.002	.374	-.084	.150	.024	.003	.001	.483	-.039	.509	.016	.016	.005	.166
Adjusted	.017	.769	4e-4	.700	4e-4	.698	-.077	.186	.015	.021	6.1e-6	.963	-.063	.283	.015	.020	3.5e-6	.972

Note: NART = Correlation of predicted values with NART scores; Ind. Effect = Independent effect of predicted values on global cognition, controlling for age, sex, and mean cortical thickness; Mod. Effect = Moderation effect of predicted values on relationship between brain structure and global cognition.

4.4 Discussion

CPM was applied to task-based functional connectivity to predict a CR residual in order to develop functional neuroimaging measures of CR, namely *positive network strength predicted CR, negative network strength predicted CR,* and *combined network strength predicted CR.* All three measures were shown to predict the CR residual in unseen individuals within the same dataset based on a sparse set of edges. Negative and combined network strength predicted CR values remained statistically significant when different cross-validation schemes were used. However, positive network strength predicted CR values were only statistically significant when LOOCV was used, suggesting possible overfitting within this network. The network strength predicted CR values met the criteria necessary for neuroimaging measures of CR, as they displayed face validity and were positively associated with cognition beyond the effects of brain structure. However, these measures did not generalise to resting-state functional connectivity data from an independent dataset.

The validation of the task-based fMRI measure on novel data within the same dataset provides further evidence in support of the objective measurement of CR using task-based fMRI. As was demonstrated in previous task-based fMRI studies, the network strength predicted CR measures here displayed face validity and protective effects on cognition (Stern et al., 2018; van Loenhoud et al., 2020). As previous studies used task-related activations (Stern et al., 2018) and task potency (van Loenhoud et al., 2020), the present study is the first to demonstrate that functional connectivity during task performance can predict CR in unseen data (Stern et al., 2018), albeit using internal cross-validation.

The CR connectomes identified here were sparse, as the positive and negative connectomes together reflected connectivity strength from only 0.1% of all edges. This level of sparsity is much greater than observed in connectomes underlying cognitive phenotypes in previous CPM studies. For example, 3.88% and 1.99% of edges were included across positive and negative networks underlying sustained attention (Rosenberg et al., 2016) and processing speed (M. Gao et al., 2020), respectively. This sparsity may have occurred due to the stricter feature selection threshold used here, $p < .0009$, than in previous CPM studies (M. Gao et al., 2020; Rosenberg et al., 2016). However, previous connectivity studies of CR have also identified sparse underlying networks, only 0.17% of edges were included in the task potency measure described by van Loenhoud et al. (2020), for example. While the present study used the same dataset as van Loenhoud et al. (2020), the latter study observed this sparse network across 12 different task-based fMRI scans, meaning the

sparsity was not just specific to the overlapping task fMRI scan used in both studies. While it still cannot be ruled out that this sparsity may be an idiosyncrasy of functional connectivity within this dataset, these findings suggest that the network of functional connectivity underlying CR is indeed sparse (see Chapter 5, section 5.4 for discussion of the functional anatomy of the sparse networks underlying CR in relation to previous findings).

Although the pattern of connectivity underlying the positive and negative CR connectomes, and sparsity of the connectomes themselves, were in line with previous research, the relationship between connectivity of the positive CR connectome and the CR residual in the present study was not particularly robust. Positive network strength predicted CR only accurately predicted the CR residual in unseen data when LOOCV was used in CPM, and not when k-fold cross-validation schemes were used (see Table S9 in 7.3.2 Supplemental Results). LOOCV can produce estimates that have high variance (Efron, 1983), particularly compared to 10-fold cross-validation (Kohavi, 1995). As such, LOOCV can result in overfitting (Lever et al., 2016) and consequently the positive network strength predicted CR values may have somewhat reflected noise in the data (Poldrack et al., 2020; R. Whelan & Garavan, 2014).

The network strength predicted CR measures did not generalise to resting-state data in an independent dataset. One possible reason for this may have been the different age profiles of the training set and test set samples. Gao et al. (2020) reported that a CPM measure of processing speed, developed in a training set of older adults, could predict processing speed in a cohort of older adults but not middle-aged or younger adults in an external dataset. In an exploratory analysis to investigate if a similar age-difference may have affected the generalisability here, the training set was restricted to adults aged 50 years or older to more closely match the age range of the training set. Still, the network strength measures of CR did not generalise to the external resting-state dataset. As such, age differences were not responsible for the inability of the model to generalise.

Another exploratory analysis investigated why the negative and combined network strength predicted CR values were negatively related to CR residuals. This was not a meaningful prediction because the predicted values were in the opposite direction to the observed values and similar findings have been treated as meaningless (Greene et al., 2018; Ren et al., 2021). However, differences in connectivity from task to rest conditions may be relevant to CR. Inspection of the negative CR connectome in the test set revealed that all edges had positive, albeit mostly non-significant, correlations with CR. As these

same edges were negatively correlated with CR in the training set, this may suggest that CR is associated with a change or reorganisation of brain connectivity in response to task demands, as previously shown by the relationship between task potency and CR (van Loenhoud et al., 2020). A practical implication of this is that while CR is associated with both task-based and resting-state connectivity, it may not be possible for measures developed solely on task-based data to generalise to resting-state data as has been demonstrated for CPM measures of cognitive phenotypes like sustained attention (Rosenberg et al., 2016).

The inability to generalise to resting-state data may also have arisen due to the nature of the data in both the training and test set datasets. Single task connectomes with static univariate functional connectivity were used in the training set but CPM studies have reported more accurate predictions of cognitive phenotypes with training sets consisting of multiple task connectomes (S. Gao et al., 2019), multivariate connectivity data (Yoo et al., 2019), and dynamic functional connectivity data (Zhu et al., 2021). In the test set, the resting-state fMRI scan was approximately five minutes in duration. This length is sufficient to obtain stable correlations for functional connectivity (Van Dijk et al., 2010), but longer durations further reduce the amount of noise in, and the reliability of, functional connectivity data (Birn et al., 2013; Van Dijk et al., 2010). The degree of individual variability in functional connectivity matrices is also greatly reduced in scans with fewer than 500 time points (Finn et al., 2015). As the test set resting-state scan contained only 200 time points, more time points may be needed for connectivity matrices to have sufficient variation across individuals in order to accurately predict phenotypes such as cognitive reserve.

Of course, the model may not have generalised to the test here simply because the model itself was suboptimal. Various additions or alterations to CPM have now been described that may improve model accuracy and generalisability. One improvement may be bootstrap aggregating (*bagging*), where bootstrapped samples of the training set (i.e., randomly sampled subsets of the training set) are used to create multiple estimates of the predicted value which are then aggregated (Jollans et al., 2019). Bagging has been previously shown to improve the performance of multiple regression models applied to neuroimaging data, particular in datasets with samples sizes of less than 400 participants (Jollans et al., 2019). Bagging applied to CPM was found to improve the generalisability of a model predicting fluid intelligence from task-based connectivity data to an external dataset (O'Connor et al., 2020). Another improvement could be the use of partial least squares regression within CPM which assigns every edge in the connectivity matrix a weight for

subsequent prediction, instead of selecting only a subset of edges as is the case in standard CPM where correlation and linear regression are used (Yoo et al., 2018). CPM with partial least squares regression outperformed standard CPM for the prediction of attentional function with respect to generalisability to external datasets (Yoo et al., 2018).

There were some important limitations in the present study. Due to incomplete coverage of the cerebellum in the majority of the training set (56% of the training set sample), nodes within the cerebellum and brainstem were removed from the functional connectivity matrices analysed here. After removing these edges, the connectivity matrices had 20,910 edges which was 58% of their original size, with 35,778 edges. Task-related activation and functional connectivity studies have identified positive relationships between the cerebellum and CR proxies (Belleville et al., 2021; Marques et al., 2016; Stern et al., 2018). Consequently, while removing nodes from the cerebellum and brainstem was necessary to avoid drastic reductions of the sample size in the training set, it removed information from a region that has been previously associated with CR. As positive associations between task-related activation and connectivity of the cerebellum have been reported, the loss of this information in the training set may have impeded the ability of the positive CR network to produce robust accurate predictions of the CR residual (i.e., across k-fold validations).

Other limitations arose due to the use of the CR residual as the target variable in the predictive model. First, there were no statistically significant independent associations between any of the brain structure variables and global cognition in the multiple regression models used to create the CR residual in each dataset. As such, while the CR residual reflected the variance in cognition that was not explained by age, sex, or brain structure; the variance uniquely attributable to brain structure was minimal. Therefore, it could be argued that the CR residual simply reflected age- and sex-adjusted global cognition. Subsequently, network-strength predicted CR measures may then have just represented individual differences in global cognition. This leads to a circular reasoning fallacy, as the candidate measure of CR, a construct which may explain individual differences in cognitive ageing, is simply just a measure of cognition. Furthermore, the validation of the candidate measure requires demonstration of a positive association between the measure and cognition, above and beyond the effects of brain structure. However, if the variables representing brain structure are not independently associated with cognition themselves, then, in effect, the test of validation simply requires an association between the candidate measure and cognition. In the present study, this was already effectively conducted during CPM.

The issue of circularity could be mitigated in future studies by developing a CR residual using variables representing brain structure that are statistically significantly associated with cognition, independent of age and sex. Using variables such as white matter hyperintensity volume, fractional anisotropy and neuropathological indices may enable this. A future study could attempt to develop an optimal CR residual by using the first principal component of an extensive set of variables representing brain structure and pathology, including regionally specific measures, as the brain structure variable. Moreover, in future studies assessing the cognitive benefit criterion of candidate CR measures, it should be established that the measure of brain structure, adjusted for in the association between the measure and cognition, is itself significantly correlated with cognition. This would ensure that the candidate CR measure is in fact associated with cognition above and beyond the effects of brain structure.

Second, because CR residuals inevitably contain a large amount of measurement error (Ewers, 2020), the use of the CR residual as a target variable, introduced irreducible error (i.e., noise in the dependent variable) into the predictive model. Because irreducible error affects generalisability (Janssen et al., 2018), even an optimal predictive model may not have generalised to the test set as the predictor variables may not contain all the information required to reconstruct the target variable, due to the amount of noise contained within the target variable. The decision to use a CR residual instead of a CR proxy was justified on the basis that CR residuals are considered more direct measures of CR than proxy variables (Stern et al., 2020) and as it was demonstrated that the most robust CR proxy explained no more than 16.8% of the variance in cognition, after accounting for age, sex, and brain structure. Nonetheless, a more nuanced approach than simply predicting a CR residual might be needed to achieve generalisability. One approach may be to predict cognition from functional connectivity, controlling for brain structure at the edge selection step of CPM. This would minimise the effects of measurement error in comparison to the approach described here, where measurement error was introduced in both the target variable (i.e., creation of the CR residual) and during CPM.

Third, there was an overlap in the measures used to create the CR residual and to assess the cognitive benefit, or protective effects, of the network strength predicted CR measures. Specifically, the CR residual was created from the regression of global cognition on age, sex, total GM volume, hippocampal volume, and mean cortical thickness. The cognitive benefit of network strength predicted CR measures was assessed in hierarchical

regressions of global cognition on age, sex, mean cortical thickness, and network strength predicted CR measures. Consequently, there was a degree of circularity in assessing the cognitive benefit of the network strength predicted CR measures. This circularity was accepted as it was important to create the CR residual with as much detailed brain structure information as possible. This circularity somewhat weakens the validity of the network strength predicted CR measures as neuroimaging measures of CR. However, the strong face validity of the network strength predicted CR measures suggests that they may still be valid. Developing CR measures by applying CPM to predict global cognition from functional connectivity, controlling for measures of brain structure at the edge selection step, may be another approach which could avoid this circularity.

There were, however, a number of strengths to the current study. A whole-brain approach was taken that considered functional connectivity across the brain, except for the cerebellum due to poor scanner coverage. As such, the model and results were not biased by a priori predictions. A cross-validation framework was applied to assess whether the model could make accurate predictions in unseen data. An external validation dataset, with functional connectivity obtained from a different fMRI condition on which the model was trained, was used to provide a rigorous test of the generalisability of the developed measures across datasets and conditions. The gold-standard recommendations for deriving measures of CR were rigorously applied by assessing the face validity of the measures in respect to their association with a robust CR proxy, validated in Chapter 3, as well assessing their protective effects on cognition, above and beyond the effects of brain structure.

In sum, the present results demonstrated that task-based functional connectivity data can be used to create objective summary measures of CR (i.e., network strength predicted CR values) that are significantly associated with a CR residual, positively correlated with a CR proxy, and demonstrate a protective effect on cognition, beyond the effects of brain structure. These findings were demonstrated on unseen data within the training set (i.e., the same dataset used to develop the measures). However, the findings did not replicate when the model was applied to the test set (i.e., resting-state data from an independent dataset). The present study presents a framework for future attempts to develop measures that can generalise across datasets and fMRI conditions such that objective measures of CR can be developed, shared, and used by the wider research community with the ultimate aim of validating their clinical potential.

5 Chapter 5: General Discussion

5.1 Summary

The aim of the present thesis was to develop and validate objective measures of BM and CR in cognitively healthy adults. While the global burden of cognitive decline and dementia is increasing, there is a lack of effective treatments. Consequently, there is a growing emphasis on preventative strategies that rely on early identification of individuals at risk of severe cognitive decline and/or dementia. However, early identification is complicated by the significant variability in cognitive decline across individuals and the lack of a one-to-one relationship between cognitive function and brain structure and/or pathology. BM and CR are two constructs that may explain this variability and therefore their accurate measurement is vital in order to improve early identification and to improve the design and evaluation of preventative interventions.

The first experimental chapter, Chapter 2, set out to clarify the relationship between brain-predicted age difference score (brainPAD) and cognitive function in order to validate brainPAD as a measure of BM. BrainPAD reflects the relative preservation of brain structural health or integrity and has been previously associated with earlier mortality, poorer physical ageing, and cognitive impairment (Cole et al., 2018; Franke & Gaser, 2012; Gaser et al., 2013; Liem et al., 2017). However, the validity of brainPAD as a measure of BM was previously uncertain as the relationship between brainPAD and cognitive function has been obscured by a lack of studies in cognitively healthy older adults that statistically controlled for the effects of age and corrected for multiple comparisons. Here, a penalised linear regression with cross-validation was applied to structural MRI data to develop a model of brainPAD that generalised to three independent datasets. Across these three datasets, brainPAD was robustly associated with general cognitive status, semantic verbal fluency, executive function, and executive function (without processing speed). The replication of these associations across multiple datasets demonstrated that lower brainPADs – reflecting better relative preservation of brain structural health – was related to better cognitive function. Therefore, brainPAD was established as a valid measure of BM.

In the second experimental chapter, Chapter 3, the validity and robustness of standard CR proxy variables as measures of CR were systematically analysed and compared. While CR is most commonly measured using socio-behavioural proxy variables, the validity of these variables as measures of CR is often not assessed within studies. Moreover, the choice of proxy tends to vary across studies. As such, the validity and relative

efficacy of proxy variables as measures of CR is unclear. This hinders the measurement of CR for researchers who cannot assess the validity of a CR proxy themselves due to a lack of neuroimaging data. It also impairs the comparability of future candidate neuroimaging measures of CR, as researchers may end up using different CR proxies for the assessment of face validity. Across two datasets, the validity and robustness of five standard CR proxies were assessed using hierarchical moderated linear regressions in complete CR models containing a measure of cognitive function and a measure of brain structure. The five proxies were educational attainment, occupational complexity, verbal intelligence, engagement in leisure activities, and engagement in physical activities. Composite proxies, created from all possible combinations of these five CR proxies, were also investigated. The analysis revealed that verbal intelligence had the largest independent associations with cognition and was the most robust CR proxy as it was validated in all CR models across both datasets. Smaller and less robust associations with cognition were observed for educational attainment and some composite proxies. These findings demonstrated that verbal intelligence should be used when measuring CR with proxies in cross-sectional studies of cognitively healthy older adults and when assessing the face validity of candidate neuroimaging measures.

Chapter 3 demonstrated that some CR proxies – specifically verbal intelligence – are robust and valid measures of CR but proxies are nonetheless suboptimal for the accurate measurement of CR (Stern et al., 2020). Measuring neural networks that vary as a function of CR – using functional neuroimaging – can provide a more direct measure of CR (Stern & Barulli, 2019). More accurate predictions of individual cognitive phenotypes have been reported using task-based fMRI in comparison to resting-state fMRI (Greene et al., 2018; Yoo et al., 2018) and to-date, candidate functional neuroimaging measures of CR have solely relied on task-based fMRI. However, measures of CR that can be generated using resting-state fMRI may have greater clinical and research utility. CPM is a method that has been applied to task-based functional connectivity to develop measures of cognitive phenotypes, such as sustained attention, that can also then be generated from resting-state functional connectivity (Rosenberg et al., 2016). This potentially enables the development of measures with high accuracy using task-based data that can be applied to resting-state data and therefore have wider clinical and research utility. The final experimental chapter, *Chapter 4*, examined whether valid functional neuroimaging measures of CR could be developed using CPM from task-based fMRI data and applied to resting-state fMRI. Application of CPM to task-based functional connectivity data developed three measures of

CR, *positive-, negative-,* and *combined-network strength predicted CR*. These three measures predicted CR, as measured by a CR residual, in unseen individuals within the same dataset. The three measures reflected the summed connectivity strength of a sparse set of edges with key nodes including the left temporal pole of the MFN, the right angular gyrus of the DMN, and the left dorsolateral prefrontal cortex and premotor/supplementary motor cortex of the FPN. These measures displayed face validity as measures of CR as they were positively associated with a CR proxy, verbal intelligence, and they also satisfied the cognitive benefit criterion as they were positively associated with global cognition, independent of mean cortical thickness. However, critically, these measures did not generalise to resting-state functional connectivity in an independent dataset. Nonetheless, Chapter 4 demonstrated a framework for developing and validating functional neuroimaging measures of CR that could be further improved in future research studies.

5.2 Implications

5.2.1 Relationship between brainPAD and cognition

In Chapter 2, robust associations between brainPAD and specific domains of cognition were observed. These robust associations demonstrated validity for brainPAD as an operational measure of BM. These associations were in line with previous findings from the literature, although previous associations were mostly in non-cognitively healthy samples and may have been affected by statistical issues (i.e., failure to statistically control for age and/or correct for multiple comparisons). For some of the cognitive domains, there were some previous studies which reported conflicting evidence regarding their association with brainPAD. In this section, the robust associations between brainPAD and specific domains of cognition, identified in Chapter 2, are reviewed in the context of previous evidence from the literature[21].

BrainPAD was negatively correlated with general cognitive status, as measured using the MMSE and DRS, in DEU and CR/RANN. Previous studies have reported that

[21] In Chapter 2, verbal intelligence was considered as a cognitive outcome measure which was assessed in terms of its relationship with brainPAD. At the outset of the work presented in this thesis, I considered verbal intelligence to be a measure of cognitive function. However, after the completion of the work presented in Chapter 2 and after learning more about the area of BM and CR, it became apparent that verbal intelligence can also be considered as a contributor to, or indicator of, BM and CR (e.g., Fleck et al., 2017; Habeck et al., 2017; Oosterman et al., 2020). Accordingly, verbal intelligence was treated as an indicator, or proxy, of CR in Chapters 3 and 4. I elected to include verbal intelligence in Chapter 2, for consistency with Boyle et al., 2021, which contains the core of Chapter 2. However, in future studies, verbal intelligence should be treated as an indicator of CR and BM rather than a cognitive outcome measure given its use as such in the literature.

brainPAD is related to general cognitive status, albeit in samples including individuals with MCI or dementia (Beheshti et al., 2018; Kaufmann et al., 2019), and without adjusting for the effect of age or controlling for multiple comparisons (Beheshti et al., 2018; Cole, Underwood, et al., 2017). In contrast to these findings, Gaser et al. (2013) reported that brainPAD was correlated with the CDR and Alzheimer's Disease Assessment Scale but not the MMSE in an MCI sample. However, Gaser et al. (2013) did not account for the effect of age. While Löwe et al. (2016) reported that brainPAD was negatively correlated with the MMSE across mixed samples of APOE e4 carriers and non-carriers (including healthy controls, MCI, and AD), it was not significantly correlated with the MMSE within healthy control and MCI subgroups. Sample sizes within these subgroups were relatively small, ranging from 14 to 81 participants. Consequently, the correlations between brainPAD and the MMSE in these subgroups may not have been adequately powered to reach significance. This study is the first to report a relationship between brainPAD and measures of general cognitive status in cognitively healthy adults while controlling for the effects of age and correcting for multiple comparisons. These findings provide strong support for the existence of a significant negative relationship between brainPAD and general cognitive status.

BrainPAD was significantly negatively correlated with semantic verbal fluency, as measured using the Animals task, in both DEU and CR/RANN. This finding contradicts non-significant correlations between brainPAD and composite measures of semantic and phonemic verbal fluency (Cole, Underwood, et al., 2017; Richard et al., 2018). However, the former study used age-adjusted t-scores to control for the age-cognition relationship rather than adding age as a covariate to the brainPAD-fluency association (cf. Le et al., 2018). As semantic verbal fluency is associated with age (Clark et al., 2009; Santos Nogueira et al., 2016), the failure to appropriately adjust for age may have obscured a significant effect. Alternatively, these previously reported non-significant correlations could be explained by the use of composite measures of both phonemic and semantic fluency as the present study did not find strong evidence for a relationship between phonemic verbal fluency and brainPAD (although it was significant in DEU, this correlation was not replicated in CR/RANN). Therefore, it is possible that a non-significant relationship between phonemic fluency and brainPAD in the Cole et al. (2017) and Richard *et* al. (2018) studies may have diluted a possible significant relationship between semantic fluency and brainPAD. The significant negative correlation between brainPAD and semantic verbal fluency observed here is supported by a previous negative association, which was adjusted for age (Franke

et al., 2013). This therefore further strengthens the validity of brainPAD as an operationally valid measure of BM and indicates that semantic verbal fluency may be supported by better BM.

Across all three datasets, brainPAD was negatively correlated with executive function as measured by trail-making tests (TMT B or CTT 2). The TMT B has previously been described as a relatively sensitive measure of cognitive decline as completion times were shown to be significantly different between healthy controls, MCI, and AD (Ashendorf et al., 2008). Likewise, the CTT 2 was also found to be sensitive to cognitive decline, with differences between AD and healthy controls (Lin et al., 2014), and between healthy controls, MCI, and AD (Guo et al., 2010). Moreover, preservation of brain structure, measured by medial temporal lobe atrophy, deep WM hyperintensities, and periventricular hyperintensities, has been previously associated with the TMT B (Oosterman et al., 2010). Therefore, it is no surprise that executive function was negatively correlated here with an index of accelerated brain ageing. Indeed, previous studies have reported similar results for trail-making versus brainPAD; however, these studies used clinical samples (traumatic brain injury; Cole et al., 2015) or did not correct for multiple comparisons (Cole, Underwood, et al., 2017). The present results therefore augment these previous findings by replicating this result across three independent datasets. This provides further support for the validity of brainPAD as a measure of BM and suggests that executive function may be supported by better BM.

BrainPAD was also negatively correlated with executive function, without the confound of processing speed (TMT B minus A), in DEU and CR/RANN. The replicated association between relative preservation of brain structure and the TMT B minus A is supported by previous studies reporting associations with medial temporal lobe atrophy, deep WM hyperintensities, and periventricular hyperintensities (Oosterman et al., 2010). In regards to previous associations with brainPAD, only one previous study investigated this relationship, in older adults with traumatic brain injury (Cole et al., 2015), where a significant positive correlation was reported. This robust association provides further support for the validity of brainPAD as a measure of BM and suggests that the association between executive function and BM is not solely driven by the influence of processing speed. This finding may also suggest that the association between brainPAD and cognition is primarily driven by a strong association with executive function, as it was also robustly associated with both the TMT B/CTT 2 as well as the Animals task, which measures semantic verbal

fluency but is also influenced by executive function (Aita et al., 2019; Ardila et al., 2006; but cf. Whiteside et al., 2016).

While the robust associations between brainPAD and general cognitive status and semantic verbal fluency observed across datasets in Chapter 2 contradict some previous findings, there are plausible explanations for those contradictory findings, as outlined above. These issues in previous studies were addressed in Chapter 2 via the use of large non-clinical samples and appropriate statistical controls. Moreover, other studies have reported associations with these cognitive domains that are in line with the associations observed here. Therefore, there is firm evidence that brainPAD is indeed robustly associated with general cognitive status and semantic verbal fluency. Moreover, given the findings support and extend previous associations with executive function, with and without the influence of processing speed, Chapter 2 further demonstrated that brainPAD is robustly associated with executive function. Together, these robust associations between the relative preservation of brain structural health and cognition establish strong validity for the use of brainPAD as a measure of BM.

5.2.2 Verbal intelligence, not educational attainment, robustly assesses CR

In Chapter 3, it was demonstrated that verbal intelligence was the most robust CR proxy with the largest associations with cognition, beyond the effects of brain structure, as measured by GM volume, hippocampal volume, and mean cortical thickness. In particular, verbal intelligence had larger and more robust CR effects than educational attainment. This finding was important and especially interesting because educational attainment is the most commonly used CR proxy (Opdebeeck et al., 2016). The relative superiority of verbal intelligence, in terms of effect size and robustness, convincingly supports an argument favouring the use of verbal intelligence over education (Perneczky et al., 2019). This argument was previously broadly supported by evidence that, compared to educational attainment, verbal intelligence was a stronger predictor of cognitive function/decline (Manly et al., 2004, 2005) and had greater protective effects on the onset of clinical symptoms of MCI/AD (Pettigrew et al., 2013, 2017). More specifically, Malek-Ahmadi et al. (2017) directly compared educational attainment and verbal intelligence in a mixed autopsy sample, consisting of adults with diagnoses of no cognitive impairment, MCI and AD. In complete CR models, including neuropathological indices and measures of episodic memory and executive function, positive evidence was found for verbal intelligence, but not education, as a CR proxy, leading to the conclusion that verbal intelligence measures are superior to

educational attainment as CR proxies. In Chapter 3, it was demonstrated that verbal intelligence is also a superior CR proxy when using in-vivo measures of GM volume, hippocampal volume, and mean cortical thickness and when assessed in respect to additional cognitive outcome measures, including verbal fluency, processing speed, and global cognition. Importantly, these results demonstrated that this conclusion held when tested across two separate samples of cognitively healthy older adults.

The larger and more robust effects of verbal intelligence reported here and elsewhere could be explained by two key factors. First, verbal intelligence may be a closer reflection of the *quality*, benefit, or outcomes of educational attainment (Manly et al., 2002) than years of education, which simply reflects the *quantity* of educational attainment. Quality of education can differ greatly among individuals with the same quantity of education due to various socioeconomic and systemic factors (Chin et al., 2012), such as class size (Ehrenberg et al., 2001), and also due to individual level factors such as intrinsic learning motivation and academic self-efficacy (Hsieh, 2014). Second, measures of verbal intelligence may reflect wider lifetime educational and cognitive experiences as compared to years of education which is generally restricted to early-life formal education (Baker et al., 2017; Oh et al., 2018; Perneczky et al., 2019; Schwartz et al., 2016) and typically neglects to consider later-life education which has been positively associated with cognitive function (Anatürk, Suri, Smith, et al., 2020; Peeters et al., 2020). In this sense, verbal intelligence could be considered a dynamic CR proxy which can change over time (Deary et al., 1998; Giambra et al., 1995; McHutchison et al., 2019) whereas years of education may be considered a static CR proxy (Malek-Ahmadi et al., 2017). Despite the widespread use of educational attainment as an individual CR proxy, these results suggest that it should only be used as an individual proxy in cross-sectional studies of cognitively healthy older adults where verbal intelligence is not available.

5.2.3 Functional anatomy of sparse networks underlying CR

In Chapter 4, it was shown that task-based functional connectivity can be used to objectively measure CR, although the measures did not generalise to an external resting-state dataset. An interesting finding was that CR could be predicted in the training set by a very sparse network of functional connections within specific functional networks. This suggested that the edges and functional networks identified in the positive and negative CR connectomes are particularly relevant to CR. To establish whether the relationship of these edges and functional networks with CR was indeed sensible, such that the model was

identifying patterns of connectivity truly relevant for CR, they must be considered in terms of the broader literature. In this section, these edges and functional networks, are discussed in terms of their previously identified associations with CR.

The negative CR connectome was largely influenced by connectivity within the motor network and of the motor network to the DMN and MFN. This suggested that stronger connectivity of the motor network is related to lower CR. A key node of the motor network in the negative CR connectome was the left precentral gyrus, where functional connectivity has been previously negatively related to a CR residual (D. H. Lee et al., 2019). Activation of the left precentral gyrus during a passive language comprehension task (Bosch et al., 2009) and a working memory task (Steffener et al., 2011) has also been negatively related to CR proxy composites. Negative relationships between connectivity of the motor network and CR proxies have been previously reported, although these associations were between specific nodes of the network, including the insula and post central gyrus (Anatürk, Suri, Zsoldos, et al., 2020; Conti et al., 2021), that were not present in the negative CR connectome here.

The MFN was the other functional network with a major influence on the negative CR connectome. Stronger connectivity of a single node in the MFN, the left temporal pole, to other networks including the FPN, salience network, and visual association network, was associated with lower CR. This negative relationship is supported by previous studies where functional connectivity of the left temporal pole in adults with MCI decreased after participating in an education programme intervention (Simon et al., 2020) and where lower task-related activation of the left temporal pole was associated with higher education and better memory performance in younger adults (Springer et al., 2005). Interestingly, both left precentral gyrus and the left temporal pole were identified as two regions where functional activity, measured by metabolic activity using FDG-PET, was most negatively affected during normal aging (Kalpouzos et al., 2009). The role of these two nodes in the negative CR connectome identified here might therefore tentatively suggest that individuals with higher CR are less reliant on connectivity to these nodes of the motor network and MFN during task performance.

Aside from the left precentral gyrus and temporal pole, only one node, the right angular gyrus, had more than one significant edge in the negative CR connectome. However, this node also had multiple edges in the positive CR connectome. One possible explanation for the differential relationship of connectivity of this node to CR may be that it

depends on the levels of neuropathology present in the brain. This differential relationship was demonstrated in an FDG-PET study where higher CR, measured by education, was associated with higher metabolic functional activity in amyloid-beta negative participants but with lower activity in amyloid-beta positive participants (Ewers et al., 2013). This relationship between education and metabolic activity of the angular gyrus has also been found to differ depending on sex, with a negative relationship identified for males, but not for females, with Alzheimer's disease (Malpetti et al., 2017). One region in the negative CR connectome, the supramarginal gyrus of the bilateral inferior parietal lobules, contained multiple nodes with single edges negatively related to CR. This finding supports a previous negative association between functional connectivity here and a CR residual in cognitively healthy adults (D. H. Lee et al., 2019). Moreover, task-related activation of this region contributed to a pattern of activation across multiple tasks that had protective effects on cognition and was negatively associated with a CR proxy, verbal intelligence (Stern et al., 2018).

The negative CR connectome was robust as it accurately predicted the CR residual in unseen data in the training set when generated using different k-fold cross-validation schemes (see Table S11 in 7.3.2 Supplemental Results) as well as LOOCV. The positive CR connectome was not as robust as it did not accurately predict the CR residual in unseen data in the training set when generated using k-fold cross-validation (see Table S11 in 7.3.2 Supplemental Results). Therefore, this network needs to be interpreted with caution. However, the positive CR connectome did accurately predict the CR residual in the training set using LOOCV and the edges and functional networks identified within the positive CR connectome have been previously implicated in CR.

The positive CR connectome was primarily comprised of edges within the FPN, and from the FPN to the DMN, motor network, and visual association network. This suggested that stronger connectivity of the FPN is related to higher CR, in agreement with a previous finding where stronger functional connectivity during task performance was related to higher CR as measured by education and a CR residual (Franzmeier, Hartmann, et al., 2017). Resting-state studies have also implicated connectivity of the FPN in CR, as stronger connectivity of the FPN attenuated the impact of high amyloid burden on global cognitive decline (Buckley et al., 2017) and of WM lesions on executive function (Benson et al., 2018). This relationship may persist even after the onset of cognitive decline, as positive correlations between FPN connectivity and CR proxies (Franzmeier, Caballero, et al., 2017; Serra et al., 2016), in addition to protective effects on cognition (Franzmeier, Duering, et al.,

2017), have been reported in adults with MCI. A key node of the positive CR connectome was Brodmann area 6, the premotor/supplementary motor area in the left frontal cortex which overlaps with the specific hub of the FPN, Brodmann area 6/44 in the left frontal cortex, that has been implicated in CR in the series of studies by Franzmeier et al. Connectivity of the supplementary motor area has also been positively associated with a CR residual in cognitively healthy adults (D. H. Lee et al., 2019).

D. H. Lee et al. (2019) also reported a positive association between a CR residual and connectivity of the left temporal pole. A node in the left temporal pole was identified in the positive CR connectome, although not in the adjusted connectome generated by CPM controlling for age, sex, and head motion. An edge that did survive adjustment for confounds was identified between the FPN and the right inferior temporal gyrus of the visual association network. Increased task-related activation of this region was positively associated with a CR proxy composite as well as with protective effects on memory in the face of lower hippocampal volume (Belleville et al., 2021). Moreover, the degree strength of this node was positively related to a CR residual in a network-based analysis (Marques et al., 2016). Together, these findings suggest that the right inferior temporal gyrus may be an important hub of a CR network.

Connectivity between the FPN and DMN also contributed to the positive CR connectome, as did connectivity between the DMN and motor network. Stronger resting-state functional connectivity of the DMN has been previously implicated in CR, as it was associated with reduced cognitive decline in the face of amyloid burden (Buckley et al., 2017). Moreover, CR, as measured by education, has been positively associated with resting-state functional connectivity and metabolic activity of the anterior cingulate cortex (Arenaza-Urquijo et al., 2013), a node of the DMN (Buckner et al., 2008; Washington & VanMeter, 2015). Stronger resting-state functional connectivity of this node was subsequently associated with higher levels of verbal fluency. However, a previous study using the training set sample here, reported a negative association between task potency of edges in the DMN and CR, as measured by verbal intelligence (van Loenhoud et al., 2020). The opposite direction of that finding may be due to the difference in measures used in both studies. The present study used functional connectivity during task performance, reflecting resting-state connectivity and the changes in connectivity in response to a task, whereas task potency reflects the change in connectivity without the influence of resting-state connectivity (van Loenhoud et al., 2020). As suppression of the DMN is vital for

successful cognitive performance (Anticevic et al., 2012), an explanation could be that individuals with higher CR are better able to reduce connectivity of the DMN to other networks during task performance, which would reflect lower task potency.

In sum, while the positive CR connectome was not as robust as the negative CR connectome to different cross-validation schemes, it did predict the CR residual in the training set using LOOCV and was comprised of edges and functional networks that have been previously associated with CR. The lack of robustness of the positive CR network here may have therefore arisen due to limitations of the present study rather than a lack of edges and functional networks where stronger connectivity is associated with higher CR. In contrast, the negative CR connectome was robust. Various studies have previously identified negative associations between CR and connectivity or task-related activation of key nodes in the negative CR connectome identified here. Therefore, this suggests that the negative CR connectome identified patterns of connectivity meaningful to CR. More careful consideration of factors such as neuropathology and sex may be needed to elucidate the contributions of specific nodes, such as the right angular gyrus. Nonetheless, the key functional networks and nodes identified in the negative CR connectome strongly supports previous studies suggesting that individuals with greater connectivity and task-related activation of these networks and nodes tend to have lower CR.

5.2.4 Implications summary

Together, the findings outlined in the present thesis demonstrate that brainPAD can be used as an operational measure of BM and that verbal intelligence can be used as a proxy measure of CR. While evidence was found that network strength predicted CR may be a valid neuroimaging measure of CR, this measure did not generalise to independent resting-state data. Nonetheless, an important implication of this measure is that it established a promising framework that can be used to develop valid measures of CR using task-based functional connectivity. Future studies can use this framework and attempt to develop a robust measure that generalises to resting-state data in independent datasets and therefore has wide clinical and research potential.

The stringent validation of brainPAD as a measure of BM has important implications for cognitive ageing research. One implication of a validated measure of BM is that it may enable improved early identification of individuals at risk of cognitive decline by identifying individuals who have lower BM despite normal cognition. This might enable more effective early intervention by directing interventions and/or resources towards these individuals

before the onset of cognitive decline. A second implication is that brainPAD can be used to identify lifestyle factors and interventions that influence BM. Interventions based on these lifestyle factors and experiences can then be designed to target BM in order to slow or reduce cognitive decline. A third implication is that brainPAD could be used to more effectively measure the impact of such interventions, as instead of relying on measurement of the effect on cognition, researchers can now assess the effect on BM directly, in addition to the effect on cognition.

More generally, given the robust associations between brainPAD and specific domains of cognitive function, brainPAD could also be used as a supplementary or alternative outcome measure in studies of general cognitive ageing, as it is not confounded by the various biases and effects (e.g., low reliability, practice effects) that limit the MMSE (Galasko et al., 1993; Pfeffer et al., 1984; Tombaugh & McIntyre, 1992) and the DRS (Emery et al., 1996; Green et al., 1995). Similarly, in studies investigating cognitive ageing in the domains of semantic verbal fluency and executive function, brainPAD could be also be included as an outcome measure because it is not affected by possible biases including scoring and administration procedures (Woods et al., 2016) and practice effects (Cooper et al., 2001; J. E. Harrison et al., 2000; B. A. Wilson et al., 2000) that can impact the Animals task or factors that can bias trail-making performance, including practice effects (Bartels et al., 2010), rater effects (Feeney et al., 2016) and participant literacy (Vaucher et al., 2014).

The validation of verbal intelligence as a robust proxy measure of CR has important implications for CR research in particular. Given that candidate neuroimaging measures of CR must be shown to display face validity via an association with a CR proxy, the evidence presented here concludes that verbal intelligence should be used for the assessment of face validity. An important implication here is that this proxy can be used to study CR-related questions by researchers without access to neuroimaging data. A more general implication is that verbal intelligence could be used to improve the screening of individuals at risk for dementia and cognitive decline. While cognitive screening measures, such as the MMSE, are often adjusted for education (Franco-Marina et al., 2010; Ylikoski et al., 1992), the finding that verbal intelligence was a more robust CR proxy than educational attainment suggests that such measures could be improved by adjusting for verbal intelligence. This would enable more accurate assessments of an individual's cognitive status. Improvements to screening of individuals at risk of dementia and cognitive decline could improve clinical trials and intervention studies by ensuring that recruited participants are in fact likely to

experience severe cognitive decline or dementia. A further implication of the present thesis, that could improve clinical trials and intervention studies, is that such trials could effectively control for CR-related differences in outcomes by matching participants based on verbal intelligence or statistically controlling for verbal intelligence when evaluating intervention efficacy.

5.3 Limitations and future directions

The results outlined in the present thesis are robust and have been rigorously assessed in terms of their replicability. As outlined above, these findings also have important implications for cognitive ageing research. Nonetheless, there are some limitations. Limitations arose due to the use of cross-sectional data, different age ranges in the CR/RANN dataset across chapters, the selection of specific cognitive measures, collapsing across sexes for all analyses, and only using GM data to represent brain structure. As discussed below, while these limitations required clarification, they were not critical faults and the conclusions drawn from the results remain valid. These limitations can be addressed in future studies that may further illuminate our understanding, and improve our measurement, of BM and CR.

The main limitation of the present thesis is that the results in each experimental chapter were cross-sectional. As such, the developed measures of BM and CR, in Chapters 2 and 4, were associated with individual differences in cognition, but not cognitive decline or change. While the measures were validated in relation to cross-sectional associations with cognition, to be completely validated as measures of BM and CR, longitudinal data is required. Similarly, the validation of proxy measures of CR in Chapter 3, did not provide information about purported protective effects of those proxies on cognitive decline. Moreover, the use of cross-sectional data prevented inferences from being made about the causal direction of the relationships between the robust proxies and cognitive function. While the reliance on cross-sectional data in the present thesis is a significant limitation, longitudinal data was not available. However, cross-sectional data is still informative. Cross-sectional data can be particularly useful for the development of candidate measures, as these can be developed and validated cross-sectionally using the available data, and then optimised later when longitudinal data becomes available. Similarly, given that many CR studies investigate cross-sectional associations with CR proxies, assessing the validity of CR proxies using cross-sectional data was important in order to provide empirically driven recommendations for future researchers.

With longitudinal data on cognitive function, a future study could assess the validity of the BM measure, CR proxies, and CR neuroimaging measure in relation to cognitive decline rather than individual differences in cognition. The development of a CR functional neuroimaging measure might not even necessarily require longitudinal neuroimaging data. Instead, functional connectivity could be used to predict a CR residual created from the regression of longitudinal cognitive function on brain structure, demographics, and baseline cognitive function. This would enable the creation of network strength predicted CR measures that accounted for the variance in cognitive change that is not explained by brain structure, demographics, and baseline cognition. In contrast, for an optimal neuroimaging measure of BM, longitudinal neuroimaging data may be required. The brain age prediction model described in Chapter 2 generated a measure of BM, brainPAD, that reflected the preservation of an individual's structural brain health relative to the norm for their age. While this is a useful cross-sectional measure, this approach somewhat failed to account for individual differences in structural brain health. With longitudinal data, brain age could be predicted at two time points and the difference between the rate of brain ageing and chronological ageing could be calculated. Like brainPAD, if this difference score is positive it would reflect accelerated brain ageing, and therefore lower BM. However, this measure would be specific to each individual and would potentially provide a more precise measure of BM.

In Chapters 2 and 4, the full age range of CR/RANN was used whereas in Chapter 3 a restricted age range was used. The rationale for this was that the development of neuroimaging measures of BM and CR in Chapters 2 and 4, respectively, would be better served by the larger range in brain structure/function and cognitive function afforded by the larger age range. In contrast, in Chapter 3, already established proxy measures of CR were assessed in terms of their robustness and validity, so in order to draw comparisons across two datasets, it was necessary to match their age ranges. While this difference in age range may be incongruent across experimental chapters, there was a clear rationale for the age range used in each dataset. Moreover, there were protections in Chapters 2 and 4 against the possibility of any findings arising as a consequence of the increased age range of CR/RANN. In Chapter 2, a third dataset, DEU, was assessed, which had a restricted age range more closely matching the TILDA dataset. Furthermore, age was included as a covariate in the analysis of the associations between brainPAD and cognitive function. In Chapter 4, a supplementary analysis with a restricted age range of CR/RANN was conducted (see Section 4.3.8 and Table S10 in 7.3.1. Supplementary Results) and adjusted

measures of network strength predicted CR were created in which age was included as a covariate when relating the functional connectivity of edges in connectivity matrices to CR. Given these controls, it is unlikely that the difference in the age range of CR/RANN across the experimental chapters had undue effects on the findings outlined here.

A limitation relevant to all three experimental chapters was the use of different cognitive measures to assess the putatively same cognitive processes. Differences in the cognitive measures could have potentially resulted in differential associations of that cognitive domain to BM/CR across datasets. For instance, the verbal episodic memory (delayed) variables were measured by delayed recall of word lists after a 40 minute delay in DEU, a 15 minute delay in CR/RANN, and a 20-25 minute delay in TILDA. In Chapter 2, brainPAD was only significantly associated with verbal episodic memory (delayed) in DEU. One possible explanation for this may be that the shorter delay periods in CR/RANN and TILDA may have resulted in ceiling effects (Uttl et al., 2002) that attenuated any statistical association with brainPAD. The longer delay periods in DEU may have been less affected by ceiling effects as word list recall decays with increasing delay durations (Geffen et al., 1997; Saloner et al., 2018). However, the durations assessed by Geffen et al. and Saloner et al. were on the scale of days to weeks, whereas the increased duration in DEU was a maximum of 25 minutes, versus CR/RANN. Increases in delay duration of up to 45 minutes have previously been found not to affect immediate memory performance, albeit for figures rather than word lists (Berry & Carpenter, 1992). As such, despite slight differences in this cognitive measure, differences in the association between brainPAD and verbal episodic memory across datasets may have been due to differences in confounding factors and age range (see Chapter 2, section 2.4 for discussion of these factors).

Another example of a cognitive domain that was assessed by different cognitive measures was executive function, which was assessed by the TMT B in DEU and CR/RANN but by the CTT 2 in TILDA. Although the CTT 2 has been described as a direct 'culture-free' analogue of the TMT B (Elkin-Frankston et al., 2007; Messinis et al., 2011), the CTT 2 has different stimuli (shapes and colours vs numbers and letters) and takes longer because it has more stimuli (Mitrushina et al., 2005). Consequently, some have argued, based on findings of significant difference in mean scores on CTT 2 and TMT B, that the tests are not direct equivalents (Dugbartey et al., 2000; Strauss et al., 2006). However, mean scores for both measures are calculated as time to completion and therefore a difference in means between both measures primarily reflects a difference in test length. A more appropriate

measure of test equivalence would be correlations between mean scores, and various studies report significant correlations between both measures (Dugbartey et al., 2000; Elkin-Frankston et al., 2007; T. Lee et al., 2000; Messinis et al., 2011).

Although it would have been preferable to use identical measures across datasets, this was not possible because the studies were designed after data collection so they made use of the pre-existing data. Nonetheless, the measures used here were broadly comparable in that they are apparent measures of the same underlying cognitive constructs. It could also be argued that any slight differences in cognitive measures used across datasets might actually strengthen the evidence supporting a robust association between BM and a specific domain of cognition, or a robust effect of a particular CR proxy. Nonetheless, to ensure that differences in the cognitive measures used to assess specific cognitive domains do not bias the relationship of cognitive function with BM and CR, a future study could ensure agreement in the measures used across datasets.

Another limitation in relation to the cognitive measures in the present thesis, is that the use of a specific cognitive measure, the TMT/CTT difference score, may not have assessed the intended underlying cognitive domain. This measure, calculated via the subtraction of TMT A/CTT 1 from TMT B/CTT 2, was intended to provide a measure of executive function without the influence of processing speed (Strauss et al., 2006). However, evidence suggests that the difference score is in fact influenced by processing speed: processing speed, as well as attention, episodic memory, and working memory, were independently associated with TMT difference scores in a stepwise regression (Oosterman et al., 2010). In contrast, the ratio score, calculated by the division of TMT B by TMT A, was only associated with executive function. Given these findings, a purer measure of executive function may have been obtained by using the ratio score instead of the difference score in this thesis. In particular, this may have introduced redundant information into the global cognition variable used in Chapters 3 and 4, as processing speed may have essentially been accounted for twice in that measure.

Previous work has shown that there is an influence of processing speed on the TMT difference score (Oosterman et al., 2010). However, a pattern of results observed in Chapter 2 suggests that the difference score may have accounted for executive function as intended, without the influence of processing speed. Processing speed (TMT A/CTT 1) was associated with brainPAD in TILDA, but not DEU or CR/RANN. In contrast, executive function with the influence of processing speed (TMT B/CTT 2) was associated with

brainPAD in all three datasets. Finally, executive function without the influence of processing speed (i.e., the difference score) was associated with brainPAD in DEU and CR/RANN but not TILDA. Given that processing speed, and executive function with the influence of processing speed were associated with brainPAD in TILDA but executive function without the influence of processing speed was not, this suggests that the difference score was not unduly influenced by processing speed. Nonetheless, based on the findings reported by Oosterman et al. (2010), the ratio score may be a preferable measure of executive function without the influence of processing speed. Given that executive function without processing speed was one of only four cognitive domains reliably associated with brainPAD, future research should clarify that this association holds when using the TMT/CTT ratio score instead of the difference score.

A key limitation in the present thesis was that all analyses collapsed across sexes. In Chapter 2, brainPAD was developed using a training set comprised of males and females. Brain age prediction models have been developed in males and females separately, and have resulted in greater training set accuracy, with Pearson's r = 0.93 for females and 0.94 for males (Kaufmann et al., 2019), than reported here, r = 0.85. However, the model reported here had comparable accuracy in external test sets, ranging from r = 0.65 to 0.87, than for models trained within each sex separately, r = 0.53 to 0.86 as reported by Kaufmann et al. As such, developing models separately within sexes may not improve model performance. Moreover, this may only be possible when extremely large training sets are available as training set sample size is associated with model performance (de Lange et al., 2020; Schnack & Kahn, 2016) and Kaufmann et al. had a training set consisting of 35,474 participants, which was 26 times larger than the training set of 1,359 participants studied here.

Although developing models separately within sexes may not result in improved model performance when applied to external datasets, there may still be some value in doing so when using brainPAD to investigate the association between BM and cognitive, lifestyle, or health factors. Significant sex differences in brainPAD were identified in two of the three test sets assessed here and have been reported elsewhere (Cole et al., 2018; Franke et al., 2013; Luders et al., 2016; Smith et al., 2019). These sex differences in brainPAD may be relevant for the subsequent investigation of associations between BM and specific factors, as the pattern of relationships between brainPAD and health and lifestyle factors was found to differ in males and females (Franke et al., 2014). As such, it may be

worthwhile to develop sex-specific brain age prediction models when using brainPAD to investigate the influence of lifestyle or health factors on BM, or indeed the influence of BM on cognition. In the present study, while sex-specific brain age prediction models were not created in order to maintain a sufficiently large training set sample size, potential sex differences between BM and cognition were accounted for. These potential differences were mitigated by the inclusion of sex as a covariate when assessing the relationship between brainPAD and cognition. In future studies, it would be informative to identify the associations between BM and cognition that differ by sex as this may be relevant for the effective design of lifestyle interventions that target BM in order to reduce cognitive decline.

Sex was also included as a covariate in Chapter 3 when investigating the association between CR proxies and cognition. This approach mitigated the effect of potential sex differences in the relationship between CR proxies and cognition. However, there are sex differences in the factors that contribute to CR (Subramaniapillai et al., 2021). These differences may be particularly relevant in older cohorts, including the two samples assessed in Chapter 3, as men in these cohorts, on average, had greater educational and occupational opportunities in the earlier years of their lives (Subramaniapillai et al., 2021). This could be particularly pertinent in the Irish context, as Irish women were subject to a marriage bar, whereby they were obliged to retire from employment upon marriage (O'Leary, 1987). The marriage bar in Ireland was in place in some sectors until the 1970s, much later than other countries where they were by abolished the 1950s (Mosca & Wright, 2019). As a result, women in Ireland may have had less exposure to greater occupational complexity, thereby reducing the contribution of this factor to CR. While the analysis in Chapter 2 collapsed across sexes, but mitigated any sex effects by adjusting for sex as a covariate in the CR proxy-cognition relationship, a future study could conduct these analyses separately within sexes. This might elucidate sex-specific differences in the validity of certain proxies as measures of CR.

The predictive models developed for CR using CPM in Chapter 4 were also developed across sexes. This may have been a suboptimal approach as the connectomes underlying certain cognitive phenotypes, such as intelligence, have been reported to differ across sexes (Greene et al., 2018; Jiang et al., 2020). For the prediction of intelligence, these differences were sufficiently substantial such that predictive models developed in males could not predict intelligence in females, and vice versa (Jiang et al., 2020). However, developing the predictive models separately within sexes would have substantially

decreased the training set sample size. Given the relationship between sample size and model performance (de Lange et al., 2020; Schnack & Kahn, 2016), a reduction in sample size may have impaired model performance. Instead, an adjusted model was developed whereby sex was included as a covariate when assessing the relationship between each edge in the connectivity matrix and the CR residual. The adjusted models did not substantially differ from the original model. While this was considered a satisfactory correction for potential sex differences, given that different patterns of connectivity have been reported to underlie successful predictions of intelligence, similar sex-differences in the CR connectome may exist. As such, if sufficient data was available, a future study could apply CPM separately within sexes in order to account for possible sex-differences in CR-related functional connectivity.

Another key limitation of all analyses in the present thesis is that only GM neuroimaging data was used. This may have resulted in lower accuracy of the brain age prediction model described in Chapter 2. More complex feature sets using combinations of GM and WM voxel-wise density information (Cole et al., 2015, 2018; Cole, Underwood, et al., 2017), GM and WM volumetric and thickness information (Gutierrez Becker et al., 2018) and DTI metrics (Richard et al., 2018) have been used to create brainPAD scores. However, such feature sets typically require dimension reduction techniques such as PCA (Gutierrez Becker et al., 2018) or even dot products to combine GM and WM data (Cole et al., 2015, 2018; Cole, Underwood, et al., 2017). These methods can reduce the interpretability of the relationship between the original feature and brain age (Mateos-Pérez et al., 2018), although methods exist for making such feature sets interpretable (Honeine & Richard, 2009; Kwok & Tsang, 2004; Snyder et al., 2013). A specific aim of Chapter 2 was to develop an interpretable model with a relatively straight-forward method, an aim which required a simple feature set. While this approach may have limited the model's accuracy as larger and more complex feature sets often produce more accurate predictions (Scheinost et al., 2019), accuracy was still comparable to other models reported to-date in the literature. Nonetheless, future investigations of the association between BM, using brainPAD, and cognition may be improved by including WM information in the brain age prediction model. Inclusion of such information may reveal further associations. For instance, brainPAD, developed here using only GM information, was not robustly associated with processing speed. Inclusion of WM could potentially reveal an association between BM and processing speed, given that faster processing speed is associated with higher fractional anisotropy,

an index of WM microstructural integrity, in cognitively healthy older adults (Vernooij et al., 2009).

In Chapter 3, the complete CR models solely contained information relating to GM, specifically total GM volume, hippocampal volume, and mean cortical thickness. These models did not contain measures of WM microstructural integrity, WM hyperintensity volume, or AD-related neuropathology, as such data was not available across both datasets. Nonetheless, CR proxies have been previously reported to moderate the relationship between these measures and cognition (Baker et al., 2017; Dufouil et al., 2003; Joannette et al., 2019; Rentz et al., 2017; Zahodne et al., 2019). Future studies could assess proxies in complete CR models containing these brain structure variables to extend the conclusions made here to a wider spectrum of brain-cognition relationships. Furthermore, inclusion of WM measures, such as WM hyperintensity volume, would explain additional variance in cognition (Tsapanou et al., 2019). In the hierarchical moderated regressions used to analyse the complete CR models, this additional variance in cognition explained by brain structure would equate to a larger additive effect on cognition. For moderation effects of an ordinal nature, such as the theorised CR moderation effect, the size of the moderation effect is constrained by the size of the additive effect (Rogers, 2002; Whisman & McClelland, 2005). As such, a larger additive effect can reveal a larger moderation effect. Therefore, including variables reflecting WM information in the CR models may improve the ability to detect a moderation effect.

In Chapter 4, the CR residual used as the target variable for prediction only contained brain structure information pertaining to GM. As such, variance in global cognition attributable to other features of brain structure, such as WM hyperintensity volume, and neuropathology, such as amyloid and tau burden, were inadvertently retained in the CR residual. CR residuals have been described including measures of WM hyperintensity volume (Bettcher et al., 2019) and in-vivo amyloid and tau burden (D. H. Lee et al., 2019). Including such measures increases the precision of the CR residual as it minimises the amount of variance retained in the residual that can be attributed to brain structure and/or pathology. Although the CR residual used in the present study was restricted to the data available, by not including these other features, it was inevitably sub-optimal. A future study may obtain more precise network strength predicted CR measures by predicting target variables that include measures of WM microstructural integrity, WM hyperintensity volume, and in-vivo measures of neuropathology.

Addressing the limitations outlined above in future research could provide more informative and accurate measures of BM and CR and could establish further validity of specific CR proxies in relation to cognitive decline. Nonetheless, the measure of BM described here could itself be used in future studies to further investigate BM in order to better understand the variability in cognitive decline. If improvements to CPM and or richer neuroimaging data enabled validation of network strength predicted CR in an external dataset, then it too could be used in future studies. First, both measures could be used to identify the specific exposures or life experiences that influence both constructs. Identification of such factors could enable the design of particularly effective interventions to prevent dementia or reduce cognitive decline by targeting both BM and CR in comparison to just a single mechanism.

Second, the measures of BM and CR could be used to identify the cognitive domains associated with, and possibly influenced by, BM and CR. This was assessed in the present thesis for BM, but not for CR. BM, as measured using brainPAD, was robustly associated with various measures assessing executive function. This suggests that BM may be particularly relevant to the protection of executive function, although of course this was only assessed cross-sectionally. As such, this information could be used when designing intervention studies targeted at improving BM, as the results here suggest that a suitable cognitive outcome measure would be some measure of executive function. This would enable assessment of the intervention in a cognitive domain previously shown to be associated to the mechanism or construct (e.g., BM, targeted by the intervention). BrainPAD itself could be used as an outcome measure too. Future studies could identify the specific cognitive domains associated with network strength predicted CR in order to obtain similar insights.

A third future direction of the work presented here is that the measures of BM and CR could be used to further investigate the sex-related differences in cognitive decline. Among clinically normal older adults, men have been reported to experience significantly faster cognitive decline than women (McCarrey et al., 2016). This relationship is different when examined among individuals with preclinical AD, as females with higher levels of amyloid beta showed faster cognitive decline than males (Buckley et al., 2018; McCarrey et al., 2016) and females with MCI had better verbal memory than males with MCI despite similar levels of pathology, measured by temporal lobe glucose hypometabolism (Sundermann et al., 2016). These findings suggest sex-differences may contribute to the

variability in cognitive decline. A future study could further examine these sex differences, using brainPAD and network strength predicted CR, in order to establish the construct via which these sex differences are manifested in cognitive decline.

Finally, as BM and CR should account for unexplained variance in cognitive decline, it follows that accurate and valid measurement of these two constructs may improve the prediction of future cognitive decline. To-date, machine learning applied to neuroimaging data has largely failed to predict future cognitive decline (Marinescu, Bron, et al., 2020), Therefore a future study, accounting for BM and CR using brainPAD and network strength predicted CR, in addition to the different neuroimaging features reflecting brain structure as used in the TADPOLE challenge study (Marinescu, Bron, et al., 2020), may provide more accurate predictions of cognitive decline. Such a study would further enable the comparison of the relative prediction power of BM and CR by directly comparing the associations between brainPAD, network strength predicted CR and cognitive decline, independent of brain structure (or brain reserve). This would inform the design of lifestyle interventions to prevent dementia or slow cognitive decline, as greater emphasis could be placed on targeting the more predictive construct.

5.4 Conclusion

In summary, this thesis attempted to develop and validate objective measures of BM and CR. Data-driven analyses, including machine learning, were applied to structural MRI, fMRI, cognitive measures, and socio-behavioural variables in order to achieve these aims. The first study established firm evidence for a robust association between brainPAD and cognition. This clarified previously unclear associations between brainPAD and cognition. As such, this study demonstrated the validity of brainPAD as an operational measure of CR. In the second study, verbal intelligence was established as a robust and valid socio-behavioural proxy measure of CR. In the third and final study, task-based functional connectivity was used to develop an objective measure of CR that displayed good accuracy and validity within the same dataset in which it was developed. However, the developed measure was not able to accurately predict CR when generated from resting-state functional connectivity in an independent dataset. Nonetheless, a promising candidate task-based measure was developed and potential areas for improvement to this measure that may enable generalisability were outlined**.** Together, these findings establish valid measures of BM and CR as well as a promising framework for future attempts to measure CR using functional connectivity data. Using these measures, and further developing and refining

them, can improve our understanding of the variability in cognitive ageing and decline. This may have important clinical and policy implications via the better design and implementation of lifestyle interventions and preventative strategies to prevent dementia and slow cognitive decline.

6 References

Agroskin, D., Klackl, J., & Jonas, E. (2014). The Self-Liking Brain: A VBM Study on the Structural Substrate of Self-Esteem. *PLoS ONE*, *9*(1), e86430. https://doi.org/10.1371/journal.pone.0086430

Aguinis, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology*, *82*(1), 192–205. https://doi.org/10.1037/0021-9010.82.1.192

Aita, S. L., Beach, J. D., Taylor, S. E., Borgogna, N. C., Harrell, M. N., & Hill, B. D. (2019). Executive, language, or both? An examination of the construct validity of verbal fluency measures. *Applied Neuropsychology: Adult*, *26*(5), 441–451. https://doi.org/10.1080/23279095.2018.1439830

Akbaraly, T. N., Portet, F., Fustinoni, S., Dartigues, J.-F., Artero, S., Rouaud, O., Touchon, J., Ritchie, K., & Berr, C. (2009). Leisure activities and the risk of dementia in the elderly: results from the Three-City Study. *Neurology*, *73*(11), 854–861. https://doi.org/10.1212/WNL.0b013e3181b7849b

Almairac, F., Duffau, H., & Herbet, G. (2018). Contralesional macrostructural plasticity of the insular cortex in patients with glioma. *Neurology*, *91*(20), e1902–e1908. https://doi.org/10.1212/WNL.0000000000006517

Amodio, P., Montagnese, S., Spinelli, G., Schiff, S., & Mapelli, D. (2017). Cognitive reserve is a resilience factor for cognitive dysfunction in hepatic encephalopathy. *Metabolic Brain Disease*, *32*(4), 1287–1293. https://doi.org/10.1007/s11011-017-0032-2

Anatürk, M., Suri, S., Smith, S., Ebmeier, K., & Sexton, C. (2020). Mid-life and late life activities and their relationship with MRI measures of brain structure and functional connectivity in the UK Biobank cohort. *BioRxiv*, 2020.04.10.035451. https://doi.org/10.1101/2020.04.10.035451

Anatürk, M., Suri, S., Zsoldos, E., Filippini, N., Mahmood, A., Singh-Manoux, A., Kivimäki, M., Mackay, C. E., Ebmeier, K. P., & Sexton, C. E. (2020). Associations Between Longitudinal Trajectories of Cognitive and Social Activities and Brain Health in Old Age. *JAMA Network Open*, *3*(8), e2013793. https://doi.org/10.1001/jamanetworkopen.2020.13793

Andel, R., Crowe, M., Pedersen, N. L., Mortimer, J., Crimmins, E., Johansson, B., & Gatz, M. (2005). Complexity of Work and Risk of Alzheimer's Disease: A Population-Based Study of Swedish Twins. *The Journals of Gerontology: Series B*, *60*(5), P251–P258.

https://doi.org/10.1093/geronb/60.5.P251

Anthony, M., & Lin, F. (2017). A Systematic Review for Functional Neuroimaging Studies of Cognitive Reserve Across the Cognitive Aging Spectrum. *Archives of Clinical Neuropsychology*, *33*(8), 937–948. https://doi.org/10.1093/arclin/acx125

Anticevic, A., Cole, M. W., Murray, J. D., Corlett, P. R., Wang, X.-J., & Krystal, J. H. (2012). The Role of Default Network Deactivation in Cognition and Disease. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2012.10.008

Ardila, A., Ostrosky-Solís, F., Bernal, B., Ostrosky-Solís, F., & Bernal, B. (2006). Cognitive testing toward the future: The example of Semantic Verbal Fluency (ANIMALS). *International Journal of Psychology*, *41*(5), 324–332. https://doi.org/10.1080/00207590500345542

Arenaza-Urquijo, E. M., de Flores, R., Gonneaud, J., Wirth, M., Ourry, V., Callewaert, W., Landeau, B., Egret, S., Mézenge, F., Desgranges, B., & Chételat, G. (2017). Distinct effects of late adulthood cognitive and physical activities on gray matter volume. *Brain Imaging and Behavior*, *11*(2), 346–356. https://doi.org/10.1007/s11682-016-9617-3

Arenaza-Urquijo, E. M., Landeau, B., La Joie, R., Mevel, K., Mézenge, F., Perrotin, A., Desgranges, B., Bartrés-Faz, D., Eustache, F., & Chételat, G. (2013). Relationships between years of education and gray matter volume, metabolism and functional connectivity in healthy elders. *NeuroImage*, *83*, 450–457. https://doi.org/10.1016/j.neuroimage.2013.06.053

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*, 40–79. https://doi.org/10.1214/09-SS054

Arvaneh, M., Robertson, I. H., & Ward, T. E. (2019). A P300-Based Brain-Computer Interface for Improving Attention. *Frontiers in Human Neuroscience*, *12*, 524. https://doi.org/10.3389/fnhum.2018.00524

Ashendorf, L., Jefferson, A. L., O'connor, M. K., CHAISSON, C., Green, R. C., Stern, R. A., O'Connor, M., Chaisson, C., Green, R. C., & Stern, R. A. (2008). Trail Making Test errors in normal aging, mild cognitive impairment, and dementia. *Archives of Clinical Neuropsychology*, *23*(2), 129–137. https://doi.org/10.1016/j.acn.2007.11.005

Azor, A. M., Cole, J. H., Holland, A. J., Dumba, M., Patel, M. C., Sadlon, A., Goldstone, A. P., & Manning, K. E. (2019). Increased brain age in adults with Prader-Willi syndrome. *NeuroImage: Clinical*, *21*, 101664. https://doi.org/https://doi.org/10.1016/j.nicl.2019.101664

Baker, L. M., Laidlaw, D. H., Cabeen, R., Akbudak, E., Conturo, T. E., Correia, S., Tate, D. F., Heaps-Woodruff, J. M., Brier, M. R., Bolzenius, J., Salminen, L. E., Lane, E. M., McMichael, A. R., & Paul, R. H. (2017). Cognitive reserve moderates the relationship between neuropsychological performance and white matter fiber bundle length in healthy older adults. *Brain Imaging and Behavior*, *11*(3), 632–639. https://doi.org/10.1007/s11682-016-9540-7

Baker, L. M., Laidlaw, D. H., Conturo, T. E., Hogan, J., Zhao, Y., Luo, X., Correia, S., Cabeen, R., Lane, E. M., Heaps, J. M., Bolzenius, J., Salminen, L. E., Akbudak, E., McMichael, A. R., Usher, C., Behrman, A., & Paul, R. H. (2014). White matter changes with age utilizing quantitative diffusion MRI. *Neurology*, *83*(3), 247—252. https://doi.org/10.1212/wnl.0000000000000597

Balart-Sánchez, S. A., Bittencourt-Villalpando, M., van der Naalt, J., & Maurits, N. M. (2021). Electroencephalography, Magnetoencephalography, and Cognitive Reserve: A Systematic Review. *Archives of Clinical Neuropsychology*, *00*, 1–18. https://doi.org/10.1093/arclin/acaa132

Bartels, C., Wegrzyn, M., Wiedl, A., Ackermann, V., & Ehrenreich, H. (2010). Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC Neuroscience*, *11*, 118. https://doi.org/10.1186/1471-2202-11-118

Beardsall, L., & Brayne, C. (1990). Estimation of verbal intelligence in an elderly community: A prediction analysis using a shortened NART. *British Journal of Clinical Psychology*. https://doi.org/10.1111/j.2044-8260.1990.tb00851.x

Beheshti, I., Maikusa, N., & Matsuda, H. (2018). The association between “Brain-Age Score” (BAS) and traditional neuropsychological screening tools in Alzheimer’s disease. *Brain and Behavior*, *8*(8), e01020. https://doi.org/10.1002/brb3.1020

Belleville, S., Mellah, S., Cloutier, S., Dang-Vu, T. T., Duchesne, S., Maltezos, S., Phillips, N., & Hudon, C. (2021). Neural correlates of resilience to the effects of hippocampal atrophy on memory. *NeuroImage: Clinical*, *29*, 102526. https://doi.org/10.1016/j.nicl.2020.102526

Bennett, D. A., Schneider, J. A., Arvanitakis, Z., Kelly, J. F., Aggarwal, N. T., Shah, R. C., & Wilson, R. S. (2006). Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology*, *66*(12), 1837–1844. https://doi.org/10.1212/01.wnl.0000219668.47116.e6

Bennett, D. A., Schneider, J. A., Tang, Y., Arnold, S. E., & Wilson, R. S. (2006). The effect of social networks on the relation between Alzheimer’s disease pathology and level of

cognitive function in old people: a longitudinal cohort study. *Lancet Neurology*, *5*(5), 406–412. https://doi.org/10.1016/S1474-4422(06)70417-3

Benson, G., Hildebrandt, A., Lange, C., Schwarz, C., Köbe, T., Sommer, W., Flöel, A., & Wirth, M. (2018). Functional connectivity in cognitive control networks mitigates the impact of white matter lesions in the elderly. *Alzheimer's Research and Therapy*, *10*(1), 109. https://doi.org/10.1186/s13195-018-0434-3

Benton, A. L., Varney, N. R., & Hamsher, K. D. (1978). Visuospatial judgment. A clinical test. *Archives of Neurology*, *35*(6), 364–367.

Berkman, L. F., & Syme, S. L. (1979). Social networks, host resistance, and mortality: A nine-year follow-up study of alameda county residents. *American Journal of Epidemiology*, *109*(2), 186–204. https://doi.org/10.1093/oxfordjournals.aje.a112674

Berry, D. T. R., & Carpenter, G. S. (1992). Effect of four different delay periods on recall of the Rey-Osterrieth Complex Figure by older persons. *Clinical Neuropsychologist*, *6*(1), 80–84. https://doi.org/10.1080/13854049208404119

Bettcher, B. M., Gross, A. L., Gavett, B. E., Widaman, K. F., Fletcher, E., Dowling, N. M., Buckley, R. F., Arenaza-Urquijo, E. M., Zahodne, L. B., Hohman, T. J., Vonk, J. M. J., Rentz, D. M., & Mungas, D. (2019). Dynamic change of cognitive reserve: associations with changes in brain, cognition, and diagnosis. *Neurobiology of Aging*, *83*, 95–104. https://doi.org/10.1016/j.neurobiolaging.2019.08.016

Bhome, R., Berry, A. J., Huntley, J. D., & Howard, R. J. (2018). Interventions for subjective cognitive decline: Systematic review and meta-analysis. *BMJ Open*, 8(7), p. 21610. https://doi.org/10.1136/bmjopen-2018-021610

Biddle, K. D., d'Oleire Uquillas, F., Jacobs, H. I. L., Zide, B., Kirn, D. R., Rentz, D. M., Johnson, K. A., Sperling, R. A., & Donovan, N. J. (2019). Social Engagement and Amyloid-β-Related Cognitive Decline in Cognitively Normal Older Adults. *American Journal of Geriatric Psychiatry*, *27*(11), 1247–1256. https://doi.org/10.1016/j.jagp.2019.05.005

Birn, R. M., Molloy, E. K., Patriat, R., Parker, T., Meier, T. B., Kirk, G. R., Nair, V. A., Meyerand, M. E., & Prabhakaran, V. (2013). The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *NeuroImage*, *83*, 550–558. https://doi.org/10.1016/j.neuroimage.2013.05.099

Bland, J. M., & Altman, D. G. (2011). Correlation in restricted ranges of data. *BMJ (Clinical Research Ed.)*, *342*, d556. https://doi.org/10.1136/bmj.d556

Bloom, D. E., Chatterji, S., Kowal, P., Lloyd-Sherlock, P., McKee, M., Rechel, B.,

Rosenberg, L., & Smith, J. P. (2015). Macroeconomic implications of population ageing and selected policy responses. *The Lancet*, *385*(9968), 649–657. https://doi.org/10.1016/S0140-6736(14)61464-1

Boots, E. A., Schultz, S. A., Almeida, R. P., Oh, J. M., Koscik, R. L., Dowling, M. N., Gallagher, C. L., Carlsson, C. M., Rowley, H. A., Bendlin, B. B., Asthana, S., Sager, M. A., Hermann, B. P., Johnson, S. C., & Okonkwo, O. C. (2015). Occupational Complexity and Cognitive Reserve in a Middle-Aged Cohort at Risk for Alzheimer's Disease. *Archives of Clinical Neuropsychology*, *30*(7), 634–642.

Borroni, B., Premi, E., Agosti, C., Alberici, A., Garibotto, V., Bellelli, G., Paghera, B., Lucchini, S., Giubbini, R., Perani, D., & Padovani, A. (2009). Revisiting brain reserve hypothesis in frontotemporal dementia: Evidence from a brain perfusion study. *Dementia and Geriatric Cognitive Disorders*, *28*(2), 130–135. https://doi.org/10.1159/000235575

Bosch, B., Bartrés-Faz, D., Rami, L., Arenaza-Urquijo, M., Ferná Ndez-Espejo, D., Junqué, C., Solé-Padullés, C., Sá Nchez-Valle, R., Ria Bargalló, N., Falcón, C., & Molinuevo, J. L. (2009). Cognitive reserve modulates task-induced activations and deactivations in healthy elders, amnestic mild cognitive impairment and mild Alzheimer's disease. *CORTEX*, *46*, 451–461. https://doi.org/10.1016/j.cortex.2009.05.006

Bosgelmez, S., Yildiz, M., Yazici, E., Inan, E., Turgut, C., Karabulut, U., Kircali, A., Tas, H. I., Yakisir, S. S., Cakir, U., & Sungur, M. Z. (2015). Reliability and Validity of The Turkish Version of Cognitive Assessment Interview (CAI-TR). *Klinik Psikofarmakoloji Bülteni-Bulletin of Clinical Psychopharmacology*, *25*(4), 365–380. https://doi.org/10.5455/bcp.20150502064017

Bowen, M. E. (2012). A Prospective Examination of the Relationship between Physical Activity and Dementia Risk in Later Life. *American Journal of Health Promotion*, *26*(6), 333–340. https://doi.org/10.4278/ajhp.110311-QUAN-115

Boyle, P. A., Wilson, R. S., Yu, L., Barr, A. M., Honer, W. G., Schneider, J. A., & Bennett, D. A. (2013). Much of late life cognitive decline is not due to common neurodegenerative pathologies. *Annals of Neurology*, *74*(3), 478–489. https://doi.org/10.1002/ana.23964

Briels, C. T., Briels, C. T., Schoonhoven, D. N., Schoonhoven, D. N., Stam, C. J., De Waal, H., Scheltens, P., & Gouw, A. A. (2020). Reproducibility of EEG functional connectivity in Alzheimer's disease. *Alzheimer's Research and Therapy*.

https://doi.org/10.1186/s13195-020-00632-3

Buchman, A. S., Yu, L., Wilson, R. S., Lim, A., Dawe, R. J., Gaiteri, C., Leurgans, S. E., Schneider, J. A., & Bennett, D. A. (2019). Physical activity, common brain pathologies, and cognition in community-dwelling older adults. *Neurology*, *92*(8), E811–E822. https://doi.org/10.1212/WNL.0000000000006954

Buckley, R. F., Mormino, E. C., Amariglio, R. E., Properzi, M. J., Rabin, J. S., Lim, Y. Y., Papp, K. V., Jacobs, H. I. L., Burnham, S., Hanseeuw, B. J., Doré, V., Dobson, A., Masters, C. L., Waller, M., Rowe, C. C., Maruff, P., Donohue, M. C., Rentz, D. M., Kirn, D., … Sperling, R. A. (2018). Sex, amyloid, and APOE ε4 and risk of cognitive decline in preclinical Alzheimer's disease: Findings from three well-characterized cohorts. *Alzheimer's and Dementia*, *14*(9), 1193–1203. https://doi.org/10.1016/j.jalz.2018.04.010

Buckley, R. F., Schultz, A. P., Hedden, T., Papp, K. V., Hanseeuw, B. J., Marshall, G., Sepulcre, J., Smith, E. E., Rentz, D. M., Johnson, K. A., Sperling, R. A., & Chhatwal, J. P. (2017). Functional network integrity presages cognitive decline in preclinical Alzheimer disease. *Neurology*, *89*(1), 29–37. https://doi.org/10.1212/WNL.0000000000004059

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The Brain's Default Network: Anatomy, Function, and Relevance to Disease. *Annals of the New York Academy of Sciences*, *1124*(1), 1–38. https://doi.org/10.1196/annals.1440.011

Bunea, F., She, Y., Ombao, H., Gongvatana, A., Devlin, K., & Cohen, R. (2011). Penalized least squares regression methods and applications to neuroimaging. *NeuroImage*, *55*(4), 1519–1527. https://doi.org/10.1016/j.neuroimage.2010.12.028

Buschke, H., & Fuld, P. A. (1974). Evaluating storage, retention, and retrieval in disordered memory and learning. *Neurology*, *24*(11), 1019–1025. https://doi.org/10.1212/WNL.24.11.1019

Cabeza, R., Albert, M., Belleville, S., Craik, F. I. M., Duarte, A., Grady, C. L., Lindenberger, U., Nyberg, L., Park, D. C., Reuter-Lorenz, P. A., Rugg, M. D., Steffener, J., & Rajah, M. N. (2018). Maintenance, reserve and compensation: the cognitive neuroscience of healthy ageing. *Nature Reviews Neuroscience*, *19*(11). 701–710). https://doi.org/10.1038/s41583-018-0068-2

Carey, D., Nolan, H., Kenny, R. A., & Meaney, J. (2019). Cortical covariance networks in ageing: Cross-sectional data from the Irish Longitudinal Study on Ageing (TILDA). *Neuropsychologia*, *122*, 51–61.

https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2018.11.013

Casaletto, K. B., Rentería, M. A., Pa, J., Tom, S. E., Harrati, A., Armstrong, N. M., Rajan, K. B., Mungas, D., Walters, S., Kramer, J., & Zahodne, L. B. (2020). Late-life physical and cognitive activities independently contribute to brain and cognitive resilience. *Journal of Alzheimer's Disease*, *74*(1), 363–376. https://doi.org/10.3233/JAD-191114

Cassani, R., Estarellas, M., San-Martin, R., Fraga, F. J., & Falk, T. H. (2018). Systematic Review on Resting-State EEG for Alzheimer's Disease Diagnosis and Progression Assessment. *Disease Markers*, *2018*, 5174815. https://doi.org/10.1155/2018/5174815

Cervilla, J. A., Prince, M., Joels, S., Lovestone, S., & Mann, A. (2000). Long-term predictors of cognitive outcome in a cohort of older people with hypertension. *British Journal of Psychiatry*, *177*(JUL.), 66–71. https://doi.org/10.1192/bjp.177.1.66

Champoux, J. E., & Peters, W. S. (1987). Form, effect size and power in moderated regression analysis. *Journal of Occupational Psychology*, *60*(3), 243–255. https://doi.org/10.1111/j.2044-8325.1987.tb00257.x

Chan, D., Shafto, M., Kievit, R., Matthews, F., Spink, M., Valenzuela, M., & Henson, R. N. (2018). Lifestyle activities in mid-life contribute to cognitive reserve in late-life, independent of education, occupation, and late-life activities. *Neurobiology of Aging*, *70*, 180–183. https://doi.org/10.1016/j.neurobiolaging.2018.06.012

Chanraud, S., Martelli, C., Delain, F., Kostogianni, N., Douaud, G., Aubin, H. J., Reynaud, M., & Martinot, J. L. (2007). Brain Morphometry and Cognitive Performance in Detoxified Alcohol-Dependents with Preserved Psychosocial Functioning. *Neuropsychopharmacology*, *32*(2), 429–438. https://doi.org/10.1038/sj.npp.1301219

Chapko, D., McCormack, R., Black, C., Staff, R., & Murray, A. (2018). Life-course determinants of cognitive reserve (CR) in cognitive aging and dementia–a systematic literature review. In *Aging and Mental Health*, *22*(8), 915–926. https://doi.org/10.1080/13607863.2017.1348471

Cheng, S. T. (2016). Cognitive Reserve and the Prevention of Dementia: the Role of Physical and Cognitive Activities. *Current Psychiatry Reports, 18*(9). https://doi.org/10.1007/s11920-016-0721-2

Chin, A. L., Negash, S., Xie, S., Arnold, S. E., & Hamilton, R. (2012). Quality, and not just quantity, of education accounts for differences in psychometric performance between African Americans and white non-hispanics with Alzheimer's disease. *Journal of the International Neuropsychological Society*, *18*(2), 277–285.

https://doi.org/10.1017/S1355617711001688

Christensen, H., Anstey, K. J., Leach, L. S., & Mackinnon, A. J. (2008). Intelligence, education, and the brain reserve hypothesis. In *The handbook of aging and cognition, 3rd ed.* (pp. 133–188). Psychology Press.

Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., Shinohara, R. T., Elliott, M. A., Eickhoff, S. B., Davatzikos, C., Gur, R. C., Gur, R. E., Bassett, D. S., & Satterthwaite, T. D. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage*, *154*, 174–187. https://doi.org/10.1016/j.neuroimage.2017.03.020

Clark, L. J., Gatz, M., Zheng, L., Chen, Y.-L., McCleary, C., & Mack, W. J. (2009). Longitudinal Verbal Fluency in Normal Aging, Preclinical, and Prevalent Alzheimer's Disease. *American Journal of Alzheimer's Disease & Other Dementiasr*, *24*(6), 461–468. https://doi.org/10.1177/1533317509345154

Clewett, D. V., Lee, T. H., Greening, S., Ponzio, A., Margalit, E., & Mather, M. (2016). Neuromelanin marks the spot: identifying a locus coeruleus biomarker of cognitive reserve in healthy aging. *Neurobiology of Aging*, *37*, 117–126. https://doi.org/10.1016/j.neurobiolaging.2015.09.019

Cole, J. H., Franke, K., & Cherbuin, N. (2019). Quantification of the Biological Age of the Brain Using Neuroimaging. In A. Moskalev (Ed.), *Biomarkers of Human Aging* (pp. 293–328). Springer International Publishing. https://doi.org/10.1007/978-3-030-24970-0_19

Cole, J. H., Leech, R., & Sharp, D. J. (2015). Prediction of brain age suggests accelerated atrophy after traumatic brain injury. *Annals of Neurology*, *77*(4), 571–581. https://doi.org/10.1002/ana.24367

Cole, J. H., Poudel, R. P. K., Tsagkrasoulis, D., Caan, M. W. A., Steves, C., Spector, T. D., & Montana, G. (2017). Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*, *163*, 115–124. https://doi.org/10.1016/j.neuroimage.2017.07.059

Cole, J. H., Ritchie, S. J., Bastin, M. E., Valdés Hernández, M. C., Muñoz Maniega, S., Royle, N., Corley, J., Pattie, A., Harris, S. E., Zhang, Q., Wray, N. R., Redmond, P., Marioni, R. E., Starr, J. M., Cox, S. R., Wardlaw, J. M., Sharp, D. J., & Deary, I. J. (2018). Brain age predicts mortality. *Molecular Psychiatry*, *23*(5), 1385–1392. https://doi.org/10.1038/mp.2017.62

Cole, J. H., Underwood, J., Caan, M. W. A. A., De Francesco, D., van Zoest, R. A., Leech,

R., Wit, F. W. N. M. N. M., Portegies, P., Geurtsen, G. J., Schmand, B. A., Schim van der Loeff, M. F., Franceschi, C., Sabin, C. A., Majoie, C. B. L. M. L. M., Winston, A., Reiss, P., Sharp, D. J., & COBRA collaboration. (2017). Increased brain-predicted aging in treated HIV disease. *Neurology*, *88*(14), 1349–1357. https://doi.org/10.1212/WNL.0000000000003790

Conneely, K. N., & Boehnke, M. (2007). So Many Correlated Tests, So Little Time! Rapid Adjustment of P Values for Multiple Correlated Tests. *The American Journal of Human Genetics*, *81*(6), 1158–1168. https://doi.org/10.1086/522036

Conti, L., Riccitelli, G. C., Preziosa, P., Vizzino, C., Marchesi, O., Rocca, M. A., & Filippi, M. (2021). Effect of cognitive reserve on structural and functional MRI measures in healthy subjects: a multiparametric assessment. *Journal of Neurology*. https://doi.org/10.1007/s00415-020-10331-6

Cooper, D. B., Epker, M., Lacritz, L., Weiner, M., Rosenberg, R. N., Honig, L., & Cullum, C. M. (2001). Effects of Practice on Category Fluency in Alzheimers Disease. *The Clinical Neuropsychologist*, *15*(1), 125–128. https://doi.org/10.1076/clin.15.1.125.1914

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J. F., & Oja, P. (2003). International Physical Activity Questionnaire: 12-Country Reliability and Validity. *Medicine & Science in Sports & Exercise*, *35*(8), 1381–1395. https://doi.org/10.1249/01.MSS.0000078924.61453.FB

Crowe, M., Andel, R., Pedersen, N. L., Johansson, B., & Gatz, M. (2003). Does participation in leisure activities lead to reduced risk of Alzheimer's disease? A prospective study of Swedish twins. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *58*(5), P249–P255. https://doi.org/10.1093/geronb/58.5.P249

Cruz-Almeida, Y., Fillingim, R. B., Riley, J. L. I. I. I., Woods, A. J., Porges, E., Cohen, R., & Cole, J. (2019). Chronic pain is associated with a brain aging biomarker in community-dwelling older adults. *PAIN*, *160*(5). https://doi.org/https://doi.org/10.1097/j.pain.0000000000001491

Cummings, J. L., Morstorf, T., & Zhong, K. (2014). Alzheimer's disease drug-development pipeline: Few candidates, frequent failures. *Alzheimer's Research and Therapy*, *6*(4). https://doi.org/10.1186/alzrt269

D'Elia, L. F., Satz, P., Uchiyama, C. L., & White, T. (1996). *Color Trials Test. Professional*

Manual. Psychological Assessment Resources.

Daniels, J., Gaebler, M., Lamke, J.-P., & Walter, H. (2015). Grey matter alterations in patients with depersonalization disorder: a voxel-based morphometry study. *Journal of Psychiatry & Neuroscience*, *40*(1), 397–406. https://doi.org/10.1503/jpn.130284

De Jager, C. (2005). Predicting Cognitive Decline in Healthy Older Adults. *American Journal of Geriatric Psychiatry*, *13*(8), 735–740. https://doi.org/10.1176/appi.ajgp.13.8.735

de Lange, A. M. G., Anatürk, M., Suri, S., Kaufmann, T., Cole, J. H., Griffanti, L., Zsoldos, E., Jensen, D. E. A., Filippini, N., Singh-Manoux, A., Kivimäki, M., Westlye, L. T., & Ebmeier, K. P. (2020). Multimodal brain-age prediction and cardiovascular risk: The Whitehall II MRI sub-study. *NeuroImage*, *222*, 117292. https://doi.org/10.1016/j.neuroimage.2020.117292

De Marco, M., Meneghello, F., Duzzi, D., Rigon, J., Pilosio, C., & Venneri, A. (2016). Cognitive stimulation of the default-mode network modulates functional connectivity in healthy aging. *Brain Research Bulletin*, *121*, 26–41. https://doi.org/10.1016/j.brainresbull.2015.12.001

Dean, W., & Morgan, R. F. (1988). In defense of the concept of biological aging measurement--current status. *Archives of Gerontology and Geriatrics*, *7*(3), 191–210. https://doi.org/https://doi.org/10.1016/0167-4943(88)90002-7

Deary, I. J., Starr, J. M., & MacLennan, W. J. (1998). Is age kinder to the initially more able?: Differential ageing of a verbal ability in the healthy old people in Edinburgh study. *Intelligence*, *26*(4), 357–375. https://doi.org/10.1016/S0160-2896(99)00005-7

Dekhtyar, S., Wang, H.-X., Fratiglioni, L., & Herlitz, A. (2016). Childhood school performance, education and occupational complexity: a life-course study of dementia in the Kungsholmen Project. *International Journal of Epidemiology*, *45*(4), 1207–1215. https://doi.org/10.1093/ije/dyw008

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Dohmatob, E., Eickenberg, M., Thirion, B., & Varoquaux, G. (2015). Speeding-Up Model-Selection in Graphnet via Early-Stopping and Univariate Feature-Screening. *2015 International Workshop on Pattern Recognition in NeuroImaging*, 17–20.

https://doi.org/10.1109/PRNI.2015.19

Donoghue, O. A., McGarrigle, C. A., Foley, M., Fagan, A., Meaney, J., & Kenny, R. A. (2018). Cohort Profile Update: The Irish Longitudinal Study on Ageing (TILDA). *International Journal of Epidemiology*, *47*(5), 1398-1398l. https://doi.org/10.1093/ije/dyy163

Dorrian, J., Rogers, N., & Dinges, D. (2005). Psychomotor Vigilance Performance: Neurocognitive Assay Sensitive to Sleep Loss. In C. A. Kushida (Ed.), *Sleep Deprivation: Clinical Issues, Pharmacology, and Sleep Loss Effects* (Vol. 193, pp. 39–70). CRC Press.

Doshi-Velez, F., & Kim, B. (2017). *Towards A Rigorous Science of Interpretable Machine Learning*. Retrieved from http://arxiv.org/abs/1702.08608

Dubois, M., Hadj-Selem, F., Lofstedt, T., Perrot, M., Fischer, C., Frouin, V., & Duchesnay, E. (2014). Predictive support recovery with TV-Elastic Net penalty and logistic regression: An application to structural MRI. *Proceedings - 2014 International Workshop on Pattern Recognition in Neuroimaging, PRNI 2014*, 1–4. https://doi.org/10.1109/PRNI.2014.6858517

Dudoit, S., Shaffer, J. P., & Boldrick, J. C. (2003). Multiple Hypothesis Testing in Microarray Experiments. *Statistical Science*, *18*(1).

Dufouil, C., Alpérovitch, A., Tzourio, C., & C Dufouil, A Alpérovitch, C. T. (2003). Influence of education on the relationship between white matter lesions and cognition. *Neurology*, *60*(5), 831–836. https://doi.org/10.1212/01.WNL.0000049456.33231.96

Dugbartey, A. T., Townes, B. D., & Mahurin, R. K. (2000). Equivalence of the Color Trails Test and Trail Making Test in Nonnative English-Speakers. *Archives of Clinical Neuropsychology*, *15*(5), 425–431. https://doi.org/10.1016/S0887-6177(99)00034-7

Dunlap, W. P., & Kemery, E. R. (1988). Effects of predictor intercorrelations and reliabilities on moderated multiple regression. *Organizational Behavior and Human Decision Processes*, *41*(2), 248–258. https://doi.org/10.1016/0749-5978(88)90029-5

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology*, *14*, 91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

Efron, B. (1983). Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation. *Journal of the American Statistical Association*, *78*(382), 316. https://doi.org/10.2307/2288636

Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Willms, J. D. (2001). Class Size and

Student Achievement. *Psychological Science in the Public Interest : A Journal of the American Psychological Society*, *2*(1), 1–30. https://doi.org/10.1111/1529-1006.003

Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). *Manual for the Kit of Factor-Referenced Cognitive Tests (1976)*.

Elkin-Frankston, S., Lebowitz, B. K., Kapust, L. R., Hollis, H. H., & O'Connor, M. G. (2007). The use of the Color Trails Test in the assessment of driver competence: Preliminary report of a culture-fair instrument. *Archives of Clinical Neuropsychology*, *22*(5), 631–635. https://doi.org/10.1016/j.acn.2007.04.004

Elman, J. A., Jak, A. J., Panizzon, M. S., Tu, X. M., Chen, T., Reynolds, C. A., Gustavson, D. E., Franz, C. E., Hatton, S. N., Jacobson, K. C., Toomey, R., McKenzie, R., Xian, H., Lyons, M. J., & Kremen, W. S. (2018). Underdiagnosis of mild cognitive impairment: A consequence of ignoring practice effects. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *10*, 372–381. https://doi.org/10.1016/J.DADM.2018.04.003

Emek-Savaş, D., Yerlikaya, D., Yener, G., & Öktem, Ö. (2020). Validity, Reliability and Normative Data of the Stroop Test Çapa Version. *Turkish Journal of Psychiatry, 31*(1), 9-21. https://doi.org/10.5080/u23549

Emery, V. O. B., Gillie, E. X., & Smith, J. A. (1996). Reclassification of the Vascular Dementias: Comparisons of Infarct and Noninfarct Vascular Dementias. *International Psychogeriatrics*, *8*(1), 33–61. https://doi.org/10.1017/S1041610296002475

Enache, D., Winblad, B., & Aarsland, D. (2011). Depression in dementia. *Current Opinion in Psychiatry*, *24*(6), 461–472. https://doi.org/10.1097/YCO.0b013e32834bb9d4

Erickson, K. I., Leckie, R. L., & Weinstein, A. M. (2014). Physical activity, fitness, and gray matter volume. *Neurobiology of Aging*, *35*(2), s20-s28. https://doi.org/10.1016/j.neurobiolaging.2014.03.034

Everson-Rose, S. A. (2003). Early Life Conditions and Cognitive Functioning in Later Life. *American Journal of Epidemiology*, *158*(11), 1083–1089. https://doi.org/10.1093/aje/kwg263

Ewers, M. (2020). Reserve in Alzheimer's disease. *Current Opinion in Psychiatry*, *33*(2), 178–184. https://doi.org/10.1097/YCO.0000000000000574

Ewers, M., Insel, P. S., Stern, Y., & Weiner, M. W. (2013). Cognitive reserve associated with FDG-PET in preclinical Alzheimer disease. *Neurology*, *80*(13), 1194–1201. https://doi.org/10.1212/WNL.0b013e31828970c2

Farina, F. R., Emek-Savaş, D. D., Rueda-Delgado, L., Boyle, R., Kiiski, H., Yener, G., &

Whelan, R. (2020). A comparison of resting state EEG and structural MRI for classifying Alzheimer's disease and mild cognitive impairment. *NeuroImage*, *215*, 116795. https://doi.org/10.1016/j.neuroimage.2020.116795

Farokhian, F., Yang, C., Beheshti, I., Matsuda, H., & Wu, S. (2017). Age-Related Gray and White Matter Changes in Normal Adult Brains. *Aging and Disease*, *8*(6), 899–909. https://doi.org/10.14336/AD.2017.0502

Feeney, J., Savva, G. M., O'Regan, C., King-Kallimanis, B., Cronin, H., Kenny, R. A., O'Regan, C., King-Kallimanis, B., Cronin, H., & Kenny, R. A. (2016). Measurement Error, Reliability, and Minimum Detectable Change in the Mini-Mental State Examination, Montreal Cognitive Assessment, and Color Trails Test among Community Living Middle-Aged and Older Adults. *Journal of Alzheimer's Disease*, *53*(3), 1107–1114. https://doi.org/10.3233/JAD-160248

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., & Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, *18*(11), 1664–1671. https://doi.org/10.1038/nn.4135

Fischer, C., Yeung, E., Hansen, T., Gibbons, S., Fornazzari, L., Ringer, L., & Schweizer, T. A. (2009). Impact of socioeconomic status on the prevalence of dementia in an inner city memory disorders clinic. *International Psychogeriatrics*, *21*(6), 1096–1104. https://doi.org/10.1017/S1041610209990846

Fischl, B. (2012). FreeSurfer. *NeuroImage*, 62(2), 774–781). https://doi.org/10.1016/j.neuroimage.2012.01.021

Fjell, A. M., Westlye, L. T., Grydeland, H., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Holland, D., Dale, A. M., Walhovd, K. B., & Alzheimer Disease Neuroimaging Initiative. (2013). Critical ages in the life course of the adult brain: nonlinear subcortical aging. *Neurobiology of Aging*, *34*(10), 2239–2247. https://doi.org/10.1016/j.neurobiolaging.2013.04.006

Fleck, J. I., Kuti, J., Mercurio, J., Mullen, S., Austin, K., & Pereira, O. (2017). The impact of age and cognitive reserve on resting-state brain connectivity. *Frontiers in Aging Neuroscience*, *9*(DEC). https://doi.org/10.3389/fnagi.2017.00392

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Fountain-Zaragoza, S., Samimy, S., Rosenberg, M. D., & Prakash, R. S. (2019).

Connectome-based models predict attentional control in aging adults. *NeuroImage*, *186*, 1–13. https://doi.org/10.1016/j.neuroimage.2018.10.074

Fox, M. D., & Greicius, M. (2010). Clinical applications of resting state functional connectivity. *Frontiers in Systems Neuroscience, 4*, 19. https://doi.org/10.3389/fnsys.2010.00019

Franco-Marina, F., García-González, J. J., Wagner-Echeagaray, F., Gallo, J., Ugalde, O., Sánchez-García, S., Espinel-Bermúdez, C., Juárez-Cedillo, T., Rodríguez, M. Á. V., & García-Peña, C. (2010). The Mini-mental state examination revisited: Ceiling and floor effects after score adjustment for educational level in an aging Mexican population. *International Psychogeriatrics*, *22*(1), 72–81. https://doi.org/10.1017/S1041610209990822

Franke, K., & Gaser, C. (2012). Longitudinal Changes in Individual BrainAGE in Healthy Aging, Mild Cognitive Impairment, and Alzheimer's Disease. *GeroPsych*, *25*(4), 235–245. https://doi.org/10.1024/1662-9647/a000074

Franke, K., Gaser, C., Manor, B., & Novak, V. (2013). Advanced BrainAGE in older adults with type 2 diabetes mellitus. *Frontiers in Aging Neuroscience*, *5*, 90. https://doi.org/10.3389/fnagi.2013.00090

Franke, K., Ristow, M., Gaser, C., & Alzheimer's Disease Neuroimaging Initiative. (2014). Gender-specific impact of personal health parameters on individual brain aging in cognitively unimpaired elderly subjects. *Frontiers in Aging Neuroscience*, *6*, 94. https://doi.org/10.3389/fnagi.2014.00094

Franke, K., Ziegler, G., Klöppel, S., & Gaser, C. (2010). Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. *NeuroImage*, *50*(3), 883–892. https://doi.org/10.1016/j.neuroimage.2010.01.005

Franzmeier, N., Caballero, M. Á. A., Taylor, A. N. W., Simon-Vermot, L., Buerger, K., Ertl-Wagner, B., Mueller, C., Catak, C., Janowitz, D., Baykara, E., Gesierich, B., Duering, M., & Ewers, M. (2017). Resting-state global functional connectivity as a biomarker of cognitive reserve in mild cognitive impairment. *Brain Imaging and Behavior*, *11*(2), 368–382. https://doi.org/10.1007/s11682-016-9599-1

Franzmeier, N., Duering, M., Weiner, M., Dichgans, M., & Ewers, M. (2017). Left frontal cortex connectivity underlies cognitive reserve in prodromal Alzheimer disease. *Neurology*, *88*(11), 1054–1061. https://doi.org/10.1212/WNL.0000000000003711

Franzmeier, N., Göttler, J., Grimmer, T., Drzezga, A., Áraque-Caballero, M. A., Simon-

Vermot, L., Taylor, A. N. W., Bürger, K., Catak, C., Janowitz, D., Müller, C., Duering, M., Sorg, C., & Ewers, M. (2017). Resting-State Connectivity of the Left Frontal Cortex to the Default Mode and Dorsal Attention Network Supports Reserve in Mild Cognitive Impairment. *Frontiers in Aging Neuroscience*, *9*(AUG), 264. https://doi.org/10.3389/fnagi.2017.00264

Franzmeier, N., Hartmann, J. C., Taylor, A. N. W. W., Caballero, M. Á. A., Simon-Vermot, L., Buerger, K., Kambeitz-Ilankovic, L. M., Ertl-Wagner, B., Mueller, C., Catak, C., Janowitz, D., Stahl, R., Dichgans, M., Duering, M., Ewers, M., Araque Caballero, M. A., Simon-Vermot, L., Buerger, K., Kambeitz-Ilankovic, L. M., … Ewers, M. (2017). Left Frontal Hub Connectivity during Memory Performance Supports Reserve in Aging and Mild Cognitive Impairment. *Journal of Alzheimer's Disease : JAD*, *59*(4), 1381–1392. https://doi.org/10.3233/JAD-170360

Franzmeier, N., Hartmann, J., Taylor, A. N. W., Araque-Caballero, M., Simon-Vermot, L., Kambeitz-Ilankovic, L., Bürger, K., Catak, C., Janowitz, D., Müller, C., Ertl-Wagner, B., Stahl, R., Dichgans, M., Duering, M., & Ewers, M. (2018). The left frontal cortex supports reserve in aging by enhancing functional network efficiency. *Alzheimer's Research and Therapy*, *10*(1), 28. https://doi.org/10.1186/s13195-018-0358-y

Galasko, D., Abramson, I., Corey-Bloom, J., & Thal, L. J. (1993). Repeated exposure to the Mini-Mental State Examination and the Information-Memory-Concentration Test results in a practice effect in Alzheimer's disease. *Neurology*, *43*(8), 1559–1563. https://doi.org/https://doi.org/10.1212/wnl.43.8.1559

Gao, M., Wong, C. H. Y., Huang, H., Shao, R., Huang, R., Chan, C. C. H., & Lee, T. M. C. (2020). Connectome-based models can predict processing speed in older adults. *NeuroImage*, *223*, 117290. https://doi.org/10.1016/j.neuroimage.2020.117290

Gao, S., Greene, A. S., Constable, R. T., & Scheinost, D. (2019). Combining multiple connectomes improves predictive modeling of phenotypic measures. *NeuroImage*, *201*, 116038. https://doi.org/10.1016/j.neuroimage.2019.116038

Gaser, C., Franke, K., Klöppel, S., Koutsouleris, N., Sauer, H., & Alzheimer's Disease Neuroimaging Initiative. (2013). BrainAGE in Mild Cognitive Impaired Patients: Predicting the Conversion to Alzheimer's Disease. *PLoS ONE*, *8*(6), e67346. https://doi.org/10.1371/journal.pone.0067346

Gauthier, S., Albert, M., Fox, N., Goedert, M., Kivipelto, M., Mestre-Ferrandiz, J., & Middleton, L. T. (2016). Why has therapy development for dementia failed in the last two decades? In *Alzheimer's and Dementia*, *12*(1), 60-64.

https://doi.org/10.1016/j.jalz.2015.12.003

Gazzina, S., Grassi, M., Premi, E., Cosseddu, M., Alberici, A., Archetti, S., Gasparotti, R., Van Swieten, J., Galimberti, D., Sanchez-Valle, R., Laforce, R. J., Moreno, F., Synofzik, M., Graff, C., Masellis, M., Tartaglia, M. C., Rowe, J. B., Vandenberghe, R., Finger, E., … Borroni, B. (2019). Education modulates brain maintenance in presymptomatic frontotemporal dementia. *Journal of Neurology, Neurosurgery and Psychiatry*, *90*(10), 1124–1130. https://doi.org/10.1136/jnnp-2019-320439

Ge, Y., Grossman, R. I., Babb, J. S., Rabin, M. L., Mannon, L. J., & Kolson, D. L. (2002). Age-related total gray matter and white matter changes in normal adult brain. Part I: volumetric MR imaging analysis. *AJNR. American Journal of Neuroradiology*, *23*(8), 1327–1333.

Geffen, G. M., Geffen, L., Bishop, K., & Manning, L. (1997). Extended Delayed Recall of Avlt Word Lists: Effects of Age and Sex on Adult Performance. *Australian Journal of Psychology*, *49*(2), 78–84. https://doi.org/10.1080/00049539708259857

Giambra, L. M., Arenberg, D., Zonderman, A. B., Kawas, C., & Costa, P. T. (1995). Adult Life Span Changes in Immediate Visual Memory and Verbal Intelligence. *Psychology and Aging*, *10*(1), 123–139. https://doi.org/10.1037/0882-7974.10.1.123

Gillan, C. M., & Whelan, R. (2017). What big data can do for treatment in psychiatry. *Current Opinion in Behavioral Sciences*, *18*, 34–42. https://doi.org/10.1016/j.cobeha.2017.07.003

Godin, G., & Shephard, R. J. (1985). A simple method to assess exercise behavior in the community. *Canadian Journal of Applied Sport Sciences. Journal Canadien Des Sciences Appliquees Au Sport*, *10*(3), 141–146.

Golden, C. J. (1978). Stroop Color and Word Test: A manual for clinical and experimental uses. In *Chicago: Stoelting*. https://doi.org/10.1002

Good, P. (1994). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer New York. https://doi.org/10.1007/978-1-4757-2346-5

Goodwin, L. D., & Leech, N. L. (2006). Understanding Correlation: Factors That Affect the Size of r. *The Journal of Experimental Education*, *74*(3), 249–266. https://doi.org/10.3200/JEXE.74.3.249-266

Goyal, A. R., Bergh, S., Engedal, K., Kirkevold, M., & Kirkevold, Ø. (2018). Trajectories of quality of life and their association with anxiety in people with dementia in nursing homes: A 12-month follow-up study. *PLOS ONE*, *13*(9), e0203773. https://doi.org/10.1371/journal.pone.0203773

Green, R. C., Woodard, J. L., & Green, J. (1995). Validity of the Mattis Dementia Rating Scale for detection of cognitive impairment in the elderly. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *7*(3), 357–360. https://doi.org/10.1176/jnp.7.3.357

Greene, A. S., Gao, S., Noble, S., Scheinost, D., & Constable, R. T. (2020). How Tasks Change Whole-Brain Functional Organization to Reveal Brain-Phenotype Relationships. *Cell Reports*, *32*(8), 108066. https://doi.org/10.1016/j.celrep.2020.108066

Greene, A. S., Gao, S., Scheinost, D., & Constable, R. T. (2018). Task-induced brain state manipulation improves prediction of individual traits. *Nature Communications*, *9*(1), 1–13. https://doi.org/10.1038/s41467-018-04920-3

Grober, E., & Sliwinski, M. (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *Journal of Clinical and Experimental Neuropsychology*. https://doi.org/10.1080/01688639108405109

Groot, C., van Loenhoud, A. C., Barkhof, F., Van Berckel, B. N. M., Koene, T., Teunissen, C. C., Scheltens, P., Van Der Flier, W. M., & Ossenkoppele, R. (2018). Differential effects of cognitive reserve and brain reserve on cognition in Alzheimer disease. *Neurology*, *90*(2), e149–e156. https://doi.org/10.1212/WNL.0000000000004802

Gu, L., Chen, J., Gao, L., Shu, H., Wang, Z., Liu, D., Yan, Y., Li, S., & Zhang, Z. (2018). Cognitive reserve modulates attention processes in healthy elderly and amnestic mild cognitive impairment: An event-related potential study. *Clinical Neurophysiology*, *129*(1), 198–207. https://doi.org/https://doi.org/10.1016/j.clinph.2017.10.030

Guggenmos, M., Schmack, K., Sekutowicz, M., Garbusow, M., Sebold, M., Sommer, C., Smolka, M. N., Wittchen, H.-U., Zimmermann, U. S., Heinz, A., & Sterzer, P. (2017). Quantitative neurobiological evidence for accelerated brain aging in alcohol dependence. *Translational Psychiatry*, *7*(12), 1279. https://doi.org/10.1038/s41398-017-0037-y

Guo, Q. H., Cao, X. Y., Zhou, Y., Zhao, Q. H., Ding, D., & Hong, Z. (2010). Application study of quick cognitive screening test in identifying mild cognitive impairment. *Neuroscience Bulletin*, *26*(1), 47–54. https://doi.org/10.1007/s12264-010-0816-4

Gutierrez Becker, B., Klein, T., & Wachinger, C. (2018). Gaussian process uncertainty in age estimation as a measure of brain abnormality. *NeuroImage*, *175*, 246–258. https://doi.org/10.1016/j.neuroimage.2018.03.075

Habeck, C., Gazes, Y., Razlighi, Q., Steffener, J., Brickman, A., Barulli, D., Salthouse, T.,

& Stern, Y. (2016). The Reference Ability Neural Network Study: Life-time stability of reference-ability neural networks derived from task maps of young adults. *NeuroImage*, *125*, 693–704. https://doi.org/10.1016/j.neuroimage.2015.10.077

Habeck, C., Razlighi, Q., Gazes, Y., Barulli, D., Steffener, J., & Stern, Y. (2017). Cognitive Reserve and Brain Maintenance: Orthogonal Concepts in Theory and Practice. *Cerebral Cortex*, *27*(8), 3962–3969. https://doi.org/10.1093/cercor/bhw208

Han, L. K. M., Dinga, R., Hahn, T., Ching, C., Eyler, L., Aftanas, L., Aghajani, M., Aleman, A., Baune, B., Berger, K., Brak, I., Busatto Filho, G., Carballedo, A., Connolly, C., Couvy-Duchesne, B., Cullen, K., Dannlowski, U., Davey, C., Dima, D., … Schmaal, L. (2019). Brain Aging in Major Depressive Disorder: Results from the ENIGMA Major Depressive Disorder working group. *BioRxiv*, 560623. https://doi.org/10.1101/560623

Hannigan, C. B., Robertson, I., & Brennan, S. (2015). *Cognitive reserve: An investigation of construct validity and relationships with cognitive function in generally healthy older adults*. Trinity College (Dublin, Ireland). School of Psychology.

Hanssen, H., Heldmann, M., Prasuhn, J., Tronnier, V., Rasche, D., Diesta, C. C., Domingo, A., Rosales, R. L., Jamora, R. D., Klein, C., Münte, T. F., & Brüggemann, N. (2018). Basal ganglia and cerebellar pathology in X-linked dystonia-parkinsonism. *Brain*, *141*(10), 2995–3008. https://doi.org/10.1093/brain/awy222

Harrison, J. E., Buxton, P., Husain, M., & Wise, R. (2000). Short test of semantic and phonological fluency: normal performance, validity and test-retest reliability. *British Journal of Clinical Psychology*, *39*(2), 181–191. https://doi.org/10.1348/014466500163202

Harrison, J. K., Noel-Storr, A. H., Demeyere, N., Reynish, E. L., & Quinn, T. J. (2016). Outcomes measures in a decade of dementia and mild cognitive impairment trials. *Alzheimer's Research and Therapy*, *8*(1), 1–10. https://doi.org/10.1186/s13195-016-0216-8

Harrison, S. L., Sajjad, A., Bramer, W. M., Ikram, M. A., Tiemeier, H., & Stephan, B. C. M. (2015). Exploring strategies to operationalize cognitive reserve: A systematic review of reviews. *Journal of Clinical and Experimental Neuropsychology*, *37*(3), 253–264. https://doi.org/10.1080/13803395.2014.1002759

Heaton, R. K. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test Manual: Revised and expanded*. Psychological Assessment Resources. https://doi.org/10.2466/pr0.1995.76.2.623

Hedden, T., Schultz, A. P., Rieckmann, A., Mormino, E. C., Johnson, K. A., Sperling, R.

A., & Buckner, R. L. (2014). Multiple Brain Markers are Linked to Age-Related Variation in Cognition. *Cerebral Cortex*, *26*(4), 1388–1400. https://doi.org/10.1093/cercor/bhu238

Henneghan, A. M., Gibbons, C., Harrison, R. A., Edwards, M. L., Rao, V., Blayney, D. W., Palesh, O., & Kesler, S. R. (2020). Predicting Patient Reported Outcomes of Cognitive Function Using Connectome-Based Predictive Modeling in Breast Cancer. *Brain Topography*, *33*(1), 135–142. https://doi.org/10.1007/s10548-019-00746-4

Henretta, J. C., & McCrory, C. (2016). Childhood Circumstances and Mid-Life Functional Mobility. *Journal of Aging and Health*, *28*(3), 440–459. https://doi.org/10.1177/0898264315594135

Hodes, J. F., Oakley, C. I., O'Keefe, J. H., Lu, P., Galvin, J. E., Saif, N., Bellara, S., Rahman, A., Kaufman, Y., Hristov, H., Rajji, T. K., Fosnacht Morgan, A. M., Patel, S., Merrill, D. A., Kaiser, S., Meléndez-Cabrero, J., Melendez, J. A., Krikorian, R., & Isaacson, R. S. (2019). Alzheimer's "Prevention" vs. "Risk Reduction": Transcending Semantics for Clinical Practice. *Frontiers in Neurology*, *9*(JAN), 1179. https://doi.org/10.3389/fneur.2018.01179

Honeine, P., & Richard, C. (2009). Solving the pre-image problem in kernel machines: A direct method. *2009 IEEE International Workshop on Machine Learning for Signal Processing*, 1–6. https://doi.org/10.1109/MLSP.2009.5306204

Horien, C., Shen, X., Scheinost, D., & Constable, R. T. (2019). The individual functional connectome is unique and stable over months to years. *NeuroImage*, *189*, 676–687. https://doi.org/10.1016/j.neuroimage.2019.02.002

Hsieh, T. L. (2014). Motivation matters? The relationship among different types of learning motivation, engagement behaviors and learning outcomes of undergraduate students in Taiwan. *Higher Education*, *68*(3), 417–433. https://doi.org/10.1007/s10734-014-9720-6

Irimia, A., Torgerson, C. M., Goh, S.-Y. M., & Van Horn, J. D. (2015). Statistical estimation of physiological brain age as a descriptor of senescence rate during adulthood. *Brain Imaging and Behavior*, *9*(4), 678–689. https://doi.org/10.1007/s11682-014-9321-0

Janssen, R. J., Mourão-Miranda, J., & Schnack, H. G. (2018). Making Individual Prognoses in Psychiatry Using Neuroimaging and Machine Learning. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(9), 798–808. https://doi.org/10.1016/j.bpsc.2018.04.004

Jiang, R., Calhoun, V. D., Fan, L., Zuo, N., Jung, R., Qi, S., Lin, D., Li, J., Zhuo, C., Song,

M., Fu, Z., Jiang, T., & Sui, J. (2020). Gender Differences in Connectome-based Predictions of Individualized Intelligence Quotient and Sub-domain Scores. *Cerebral Cortex*, *30*(3), 888–900. https://doi.org/10.1093/cercor/bhz134

Jo, T., Nho, K., & Saykin, A. J. (2019). Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. *Frontiers in Aging Neuroscience*, *11*, 220. https://doi.org/10.3389/fnagi.2019.00220

Joannette, M., Bocti, C., Dupont, P. S., Lavallée, M. M., Nikelski, J., Vallet, G. T., Chertkow, H., & Joubert, S. (2019). Education as a Moderator of the Relationship Between Episodic Memory and Amyloid Load in Normal Aging. *The Journals of Gerontology: Series A*, *75*(10), 1820–1826. https://doi.org/10.1093/gerona/glz235

Jollans, L., Boyle, R., Artiges, E., Banaschewski, T., Desrivières, S., Grigis, A., Martinot, J.-L., Paus, T. T., Smolka, M. N., Walter, H., Schumann, G., Garavan, H., Whelan, R., Desrivieres, S., Grigis, A., Martinot, J.-L., Paus, T. T., Smolka, M. N., Walter, H., … Whelan, R. (2019). Quantifying performance of machine learning methods for neuroimaging data. *NeuroImage*, *199*, 351–365. https://doi.org/10.1016/j.neuroimage.2019.05.082

Jollans, L., & Whelan, R. (2018). Neuromarkers for Mental Disorders: Harnessing Population Neuroscience. *Frontiers in Psychiatry*, *9*, 242. https://doi.org/10.3389/fpsyt.2018.00242

Jones, R. N., Manly, J., Glymour, M. M., Rentz, D. M., Jefferson, A. L., & Stern, Y. (2011). Conceptual and Measurement Challenges in Research on Cognitive Reserve. *Journal of the International Neuropsychological Society*, *17*(04), 593–601. https://doi.org/10.1017/S1355617710001748

Joshi, A., Scheinost, D., Okuda, H., Belhachemi, D., Murphy, I., Staib, L. H., & Papademetris, X. (2011). Unified framework for development, deployment and robust testing of neuroimaging algorithms. *Neuroinformatics*, *9*(1), 69–84. https://doi.org/10.1007/s12021-010-9092-8

Jurica, P. J., Leitten, C. L., & Mattis, S. (2001). *DRS-2: Dementia Rating Scale-2: Professional Manual*. Psychological Assessment Resources.

Jutkowitz, E., Kane, R. L., Gaugler, J. E., MacLehose, R. F., Dowd, B., & Kuntz, K. M. (2017). Societal and Family Lifetime Cost of Dementia: Implications for Policy. *Journal of the American Geriatrics Society*, *65*(10), 2169–2175. https://doi.org/10.1111/jgs.15043

Kalpouzos, G., Chételat, G., Baron, J.-C. C., Landeau, B., Mevel, K., Godeau, C., Barré,

L., Constans, J.-M. M., Viader, F., Eustache, F., & Desgranges, B. (2009). Voxel-based mapping of brain gray matter volume and glucose metabolism profiles in normal aging. *Neurobiology of Aging*, *30*(1), 112–124. https://doi.org/10.1016/j.neurobiolaging.2007.05.019

Kartschmit, N., Mikolajczyk, R., Schubert, T., & Lacruz, M. E. (2019). Measuring Cognitive Reserve (CR) - A systematic review of measurement properties of CR questionnaires for the adult population. *PloS One*, *14*(8), e0219851–e0219851. https://doi.org/10.1371/journal.pone.0219851

Katzman, R., Terry, R., DeTeresa, R., Brown, T., Davies, P., Fuld, P., Renbing, X., & Peck, A. (1988). Clinical, pathological, and neurochemical changes in dementia: A subgroup with preserved mental status and numerous neocortical plaques. *Annals of Neurology*, *23*(2), 138–144. https://doi.org/10.1002/ana.410230206

Kaufmann, T., van der Meer, D., Doan, N. T., Schwarz, E., Lund, M. J., Agartz, I., Alnæs, D., Barch, D. M., Baur-Streubel, R., Bertolino, A., Bettella, F., Beyer, M. K., Bøen, E., Borgwardt, S., Brandt, C. L., Buitelaar, J., Celius, E. G., Cervenka, S., Conzelmann, A., … (KaSP), K. S. P. (2019). Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nature Neuroscience*, *22*(10), 1617–1623. https://doi.org/10.1038/s41593-019-0471-7

Kaup, A. R., Simonsick, E. M., Harris, T. B., Satterfield, S., Metti, A. L., Ayonayon, H. N., Rubin, S. M., & Yaffe, K. (2013). Older Adults With Limited Literacy Are at Increased Risk for Likely Dementia. *The Journals of Gerontology: Series A*, *69*(7), 900–906. https://doi.org/10.1093/gerona/glt176

Kearney, P. M., Cronin, H., O'Regan, C., Kamiya, Y., Savva, G. M., Whelan, B., & Kenny, R. (2011). Cohort Profile: The Irish Longitudinal Study on Ageing. *International Journal of Epidemiology*, *40*(4), 877–884. https://doi.org/10.1093/ije/dyr116

Kim, T. D., Hong, G., Kim, J., & Yoon, S. (2019). Cognitive enhancement in neurological and psychiatric disorders using transcranial magnetic stimulation (TMS): A review of modalities, potential mechanisms and future implications. *Experimental Neurobiology*, *28*(1), 1–16. https://doi.org/10.5607/en.2019.28.1.1

Kivipelto, M., Mangialasche, F., & Ngandu, T. (2018). Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. *Nature Reviews Neurology*, *14*(11), 653–666. https://doi.org/10.1038/s41582-018-0070-3

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial*

Intelligence, 1137–1143.

Koutsouleris, N., Davatzikos, C., Borgwardt, S., Gaser, C., Bottlender, R., Frodl, T., Falkai, P., Riecher-Rossler, A., Moller, H.-J., Reiser, M., Pantelis, C., & Meisenzahl, E. (2014). Accelerated Brain Aging in Schizophrenia and Beyond: A Neuroanatomical Marker of Psychiatric Disorders. *Schizophrenia Bulletin*, *40*(5), 1140–1153. https://doi.org/10.1093/schbul/sbt142

Kröger, E., Andel, R., Lindsay, J., Benounissa, Z., Verreault, R., & Laurin, D. (2008). Is Complexity of Work Associated with Risk of Dementia?: The Canadian Study of Health and Aging. *American Journal of Epidemiology*, *167*(7), 820–830. https://doi.org/10.1093/aje/kwm382

Kwak, S., Shin, M., Kim, H., Cho, B., Ha, J., Han, G., Kim, H., Koo, Y., Kwon, S., Lee, C., Lee, H. J., Park, K., & Chey, J. (2020). Moderating effect of cognitive reserve on the association between grey matter atrophy and memory varies with age in older adults. *Psychogeriatrics*, *20*(1), 87–95. https://doi.org/10.1111/psyg.12460

Kwok, J. T.-., & Tsang, I. W.-. (2004). The pre-image problem in kernel methods. *IEEE Transactions on Neural Networks*, *15*(6), 1517–1525. https://doi.org/10.1109/TNN.2004.837781

Lancaster, J., Lorenz, R., Leech, R., & Cole, J. H. (2018). Bayesian optimization for neuroimaging pre-processing in brain age classification and prediction. *Frontiers in Aging Neuroscience*, *10*(FEB), 1–10. https://doi.org/10.3389/fnagi.2018.00028

Laubach, M., Lammers, F., Zacharias, N., Feinkohl, I., Pischon, T., Borchers, F., Slooter, A. J. C., Kühn, S., Spies, C., & Winterer, G. (2018). Size matters: Grey matter brain reserve predicts executive functioning in the elderly. *Neuropsychologia*, *119*, 172–181. https://doi.org/10.1016/j.neuropsychologia.2018.08.008

Lavrencic, L. M., Churches, O. F., & Keage, H. A. D. D. (2018). Cognitive reserve is not associated with improved performance in all cognitive domains. *Applied Neuropsychology: Adult*, *25*(5), 473–485. https://doi.org/10.1080/23279095.2017.1329146

Le, T. T., Kuplicki, R. T., McKinney, B. A., Yeh, H.-W., Thompson, W. K., Paulus, M. P., & Tulsa 1000 Investigators, T. 1000. (2018). A Nonlinear Simulation Framework Supports Adjusting for Age When Analyzing BrainAGE. *Frontiers in Aging Neuroscience*, *10*, 317. https://doi.org/10.3389/fnagi.2018.00317

Lee, D. H., Lee, P., Seo, S. W., Roh, J. H., Oh, M., Oh, J. S., Oh, S. J., Kim, J. S., & Jeong, Y. (2019). Neural substrates of cognitive reserve in Alzheimer's disease

spectrum and normal aging. *NeuroImage*, *186*, 690–702. https://doi.org/10.1016/j.neuroimage.2018.11.053

Lee, P. H., Macfarlane, D. J., Lam, T. H., & Stewart, S. M. (2011). Validity of the international physical activity questionnaire short form (IPAQ-SF): A systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, *8*(1), 1–11. https://doi.org/10.1186/1479-5868-8-115

Lee, T., Cheung, C. C. Y., Chan, J. K. P., & Chan, C. C. H. (2000). Trail Making Across Languages. *Journal of Clinical and Experimental Neuropsychology*, *22*(6), 772–778. https://doi.org/10.1076/jcen.22.6.772.954

Leoń, I., Garciá-García, J., Roldań-Tapia, L., León, I., García-García, J., & Roldán-Tapia, L. (2014). Estimating cognitive reserve in healthy adults using the cognitive reserve scale. *PLoS ONE*, *9*(7), e102632. https://doi.org/10.1371/journal.pone.0102632

Lever, J., Krzywinski, M., & Altman, N. (2016). Points of Significance: Model selection and overfitting. *Nature Methods*, *13*(9), 703–704. https://doi.org/10.1038/nmeth.3968

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Liem, F., Varoquaux, G., Kynast, J., Beyer, F., Kharabian Masouleh, S., Huntenburg, J. M., Lampe, L., Rahim, M., Abraham, A., Craddock, R. C., Riedel-Heller, S., Luck, T., Loeffler, M., Schroeter, M. L., Witte, A. V., Villringer, A., & Margulies, D. S. (2017). Predicting brain-age from multimodal imaging data captures cognitive impairment. *NeuroImage*, *148*, 179–188. https://doi.org/10.1016/j.neuroimage.2016.11.005

Lin, Y. C., Shih, Y. C., Tseng, W. Y. I., Chu, Y. H., Wu, M. T., Chen, T. F., Tang, P. F., & Chiu, M. J. (2014). Cingulum Correlates of Cognitive Functions in Patients with Mild Cognitive Impairment and Early Alzheimer's Disease: A Diffusion Spectrum Imaging Study. *Brain Topography*, *27*(3), 393–402. https://doi.org/10.1007/s10548-013-0346-2

Liozidou, A., Potagas, C., Papageorgiou, S. G., & Zalonis, I. (2012). The role of working memory and information processing speed on wisconsin card sorting test performance in parkinson disease without dementia. *Journal of Geriatric Psychiatry and Neurology*, *25*(4), 215–221. https://doi.org/10.1177/0891988712466456

Lipnicki, D. M., Makkar, S. R., Crawford, J. D., Thalamuthu, A., Kochan, N. A., Lima-Costa, M. F., Castro-Costa, E., Ferri, C. P., Brayne, C., Stephan, B., Llibre-

Rodriguez, J. J., Llibre-Guerra, J. J., Valhuerdi-Cepero, A. J., Lipton, R. B., Katz, M. J., Derby, C. A., Ritchie, K., Ancelin, M. L., Carrière, I., … Sachdev, P. S. (2019). Determinants of cognitive performance and decline in 20 diverse ethno-regional groups: A COSMIC collaboration cohort study. *PLoS Medicine*, *16*(7), 12. https://doi.org/10.1371/journal.pmed.1002853

Lipton, Z. C. (2018). The Mythos of Model Interpretability. *ACM Queue*, *16*(3), 30:31--30:57. https://doi.org/10.1145/3236386.3241340

Liu, W., & Li, Q. (2017). An Efficient Elastic Net with Regression Coefficients Method for Variable Selection of Spectrum Data. *PLOS ONE*, *12*(2), e0171122. https://doi.org/10.1371/journal.pone.0171122

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C., Costafreda, S. G., Dias, A., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Kivimäki, M., Larson, E. B., Ogunniyi, A., … Mukadam, N. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet*, *396*(10248), 413–446. https://doi.org/10.1016/S0140-6736(20)30367-6

Liyanage, S. I., Santos, C., & Weaver, D. F. (2018). The hidden variables problem in Alzheimer's disease clinical trial design. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *4*(1), 628–635. https://doi.org/10.1016/j.trci.2018.09.003

Logue, S. F., & Gould, T. J. (2014). The neural and genetic basis of executive function: Attention, cognitive flexibility, and response inhibition. *Pharmacology Biochemistry and Behavior*, *123*, 45–54. https://doi.org/10.1016/j.pbb.2013.08.007

López, M. E., Turrero, A., Cuesta, P., López-Sanz, D., Bruña, R., Marcos, A., Gil, P., Yus, M., Barabash, A., Cabranes, J. A., Maestú, F., & Fernández, A. (2016). Searching for Primary Predictors of Conversion from Mild Cognitive Impairment to Alzheimer's Disease: A Multivariate Follow-Up Study. *Journal of Alzheimer's Disease*, *52*(1), 133–143. https://doi.org/10.3233/JAD-151034

Lou, Y., Caruana, R., & Gehrke, J. (2012). Intelligible Models for Classification and Regression. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 150–158. https://doi.org/10.1145/2339530.2339556

Löwe, L. C., Gaser, C., & Franke, K. (2016). The Effect of the APOE Genotype on Individual BrainAGE in Normal Aging, Mild Cognitive Impairment, and Alzheimer's

Disease. *PloS One*, *11*(7), e0157514. https://doi.org/10.1371/journal.pone.0157514

Luders, E., Cherbuin, N., & Gaser, C. (2016). Estimating brain age using high-resolution pattern recognition: Younger brains in long-term meditation practitioners. *NeuroImage*, *134*, 508–513. https://doi.org/10.1016/j.neuroimage.2016.04.007

Luo, Y., Tseng, H.-H., Cui, S., Wei, L., Ten Haken, R. K., & El Naqa, I. (2019). Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling. *BJR Open*. https://doi.org/10.1259/bjro.20190021

Madan, C. R., & Kensinger, E. A. (2018). Predicting age from cortical structure across the lifespan. *European Journal of Neuroscience*, *47*(5), 399–416. https://doi.org/10.1111/ejn.13835

Malek-Ahmadi, M., Lu, S., Chan, Y., Perez, S. E., Chen, K., & Mufson, E. J. (2017). Static and Dynamic Cognitive Reserve Proxy Measures: Interactions with Alzheimer's Disease Neuropathology and Cognition. *Journal of Alzheimer's Disease & Parkinsonism*, *07*(06). https://doi.org/10.4172/2161-0460.1000390

Malpetti, M., Ballarini, T., Presotto, L., Garibotto, V., Tettamanti, M., & Perani, D. (2017). Gender differences in healthy aging and Alzheimer's Dementia: A ^{18}F-FDG-PET study of brain and cognitive reserve. *Human Brain Mapping*, *38*(8), 4212–4227. https://doi.org/10.1002/hbm.23659

Manly, J. J., Byrd, D. A., Touradji, P., & Stern, Y. (2004). Acculturation, reading level, and neuropsychological test performance among African American elders. *Applied Neuropsychology*, *11*(1), 37–46. https://doi.org/10.1207/s15324826an1101_5

Manly, J. J., Jacobs, D. M., Touradji, P., Small, S. A., & Stern, Y. (2002). Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society*, *8*(3), 341–348. https://doi.org/10.1017/S1355617702813157

Manly, J. J., Schupf, N., Tang, M.-X., & Stern, Y. (2005). Cognitive decline and literacy among ethnically diverse elders. *Journal of Geriatric Psychiatry and Neurology*, *18*(4), 213–217. https://doi.org/10.1177/0891988705281868

Manly, J. J., Touradji, P., Tang, M. X., & Stern, Y. (2003). Literacy and memory decline among ethnically diverse elders. *Journal of Clinical and Experimental Neuropsychology*, *25*(5), 680–690. https://doi.org/10.1076/jcen.25.5.680.14579

Marden, J. R., Tchetgen Tchetgen, E. J., Kawachi, I., & Glymour, M. M. (2017). Contribution of Socioeconomic Status at 3 Life-Course Periods to Late-Life Memory Function and Decline: Early and Late Predictors of Dementia Risk. *American Journal*

of Epidemiology, *186*(7), 805–814. https://doi.org/10.1093/aje/kwx155

Marinescu, R. V., Bron, E. E., Oxtoby, N. P., Young, A. L., Toga, A. W., Weiner, M. W., Barkhof, F., Fox, N. C., Golland, P., Klein, S., & Alexander, D. C. (2020). Predicting Alzheimer's disease progression: Results from the TADPOLE Challenge. *Alzheimer's & Dementia*, *16*(S4). https://doi.org/10.1002/alz.039538

Marinescu, R. V, Oxtoby, N. P., Young, A. L., Bron, E. E., Toga, A. W., Weiner, M. W., Barkhof, F., Fox, N. C., Eshaghi, A., Toni, T., Salaterski, M., Lunina, V., Ansart, M., Durrleman, S., Lu, P., Iddi, S., Li, D., Thompson, W. K., Donohue, M. C., … Alexander, D. C. (2020). The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) Challenge: Results after 1 Year Follow-up. *Mostafa M. Ghazi*, *17*, 23.

Marioni, R. E., Proust-Lima, C., Amieva, H., Brayne, C., Matthews, F. E., Dartigues, J. F., & Jacqmin-Gadda, H. (2015). Social activity, cognitive decline and dementia risk: A 20-year prospective cohort study Chronic Disease epidemiology. *BMC Public Health*, *15*(1), 1089. https://doi.org/10.1186/s12889-015-2426-6

Marques, P., Moreira, P., Magalhães, R., Costa, P., Santos, N., Zihl, J., Soares, J., & Sousa, N. (2016). The functional connectome of cognitive reserve. *Human Brain Mapping*, *37*(9), 3310–3322. https://doi.org/10.1002/hbm.23242

Marques, P., Soares, J. M., Magalhães, R., Santos, N. C., & Sousa, N. (2015). The Bounds of Education in the Human Brain Connectome. *Scientific Reports*, *5*(1), 12812. https://doi.org/10.1038/srep12812

Mateos-Pérez, J. M., Dadar, M., Lacalle-Aurioles, M., Iturria-Medina, Y., Zeighami, Y., & Evans, A. C. (2018). Structural neuroimaging as clinical predictor: A review of machine learning applications. *NeuroImage: Clinical*, *20*, 506–522. https://doi.org/10.1016/j.nicl.2018.08.019

McCaffrey, R. J., & Westervelt, H. J. (1995). Issues associated with repeated neuropsychological assessments. *Neuropsychology Review*, *5*(3), 203–221. https://doi.org/10.1007/BF02214762

McCarrey, A. C., An, Y., Kitner-Triolo, M. H., Ferrucci, L., & Resnick, S. M. (2016). Sex differences in cognitive trajectories in clinically normal older adults. *Psychology and Aging*, *31*(2), 166–175. https://doi.org/10.1037/pag0000070

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*(2), 376–390. https://doi.org/10.1037/0033-2909.114.2.376

McCrory, C., Finucane, C., O'Hare, C., Frewen, J., Nolan, H., Layte, R., Kearney, P. M., & Kenny, R. A. (2016). Social Disadvantage and Social Isolation Are Associated With a Higher Resting Heart Rate: Evidence From The Irish Longitudinal Study on Ageing. *The Journals of Gerontology: Series B*, *71*(3), 463–473. https://doi.org/10.1093/geronb/gbu163

McCrory, C., & Kenny, R. A. (2018). Rebuking the concept of ageing as a disease. *The Lancet. Diabetes & Endocrinology*, *6*(10), 768. https://doi.org/10.1016/S2213-8587(18)30266-3

McDade, E., Bednar, M. M., Brashear, H. R., Miller, D. S., Maruff, P., Randolph, C., Ismail, Z., Carrillo, M. C., Weber, C. J., Bain, L. J., & Hake, A. M. (2020). The pathway to secondary prevention of Alzheimer's disease. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *6*(1), e12069. https://doi.org/10.1002/trc2.12069

McHutchison, C. A., Chappell, F. M., Makin, S., Shuler, K., Wardlaw, J. M., & Cvoro, V. (2019). Stability of Estimated Premorbid Cognitive Ability over Time after Minor Stroke and Its Relationship with Post-Stroke Cognitive Ability. *Brain Sciences*, *9*(5), 117. https://doi.org/10.3390/brainsci9050117

McIntyre, R. S., Cha, D. S., Soczynska, J. K., Woldeyohannes, H. O., Gallaugher, L. A., Kudlow, P., Alsuwaidan, M., & Baskaran, A. (2013). Cognitive deficits and functional outcomes in major depressive disorder: Determinants, substrates, and treatment interventions. *Depression and Anxiety*, *30*(6), 515–527. https://doi.org/10.1002/da.22063

Medaglia, J. D., Pasqualetti, F., Hamilton, R. H., Thompson-Schill, S. L., & Bassett, D. S. (2017). Brain and cognitive reserve: Translation via network control theory. *Neuroscience & Biobehavioral Reviews*, *75*, 53–64. https://doi.org/10.1016/J.NEUBIOREV.2017.01.016

Mendez, M. F., Chavez, D., & Akhlaghipour, G. (2019). Bilingualism Delays Expression of Alzheimer's Clinical Syndrome. *Dementia and Geriatric Cognitive Disorders*, *48*(5–6), 281–289. https://doi.org/10.1159/000505872

Mendoza, J. L., & Mumford, M. (1987). Corrections for Attenuation and Range Restriction on the Predictor. *Journal of Educational Statistics*, *12*(3), 282. https://doi.org/10.2307/1164688

Mennes, M., Biswal, B. B., Castellanos, F. X., & Milham, M. P. (2013). Making data sharing work: The FCP/INDI experience. *NeuroImage*, *82*, 683–691. https://doi.org/10.1016/j.neuroimage.2012.10.064

Messinis, L., Malegiannaki, A.-C., Christodoulou, T., Panagiotopoulos, V., & Papathanasopoulos, P. (2011). Color Trails Test: Normative Data and Criterion Validity for the Greek Adult Population. *Archives of Clinical Neuropsychology*, *26*(4), 322–330. https://doi.org/10.1093/arclin/acr027

Miciak, J., Taylor, W. P., Stuebing, K. K., Fletcher, J. M., & Vaughn, S. (2016). Designing Intervention Studies: Selected Populations, Range Restrictions, and Statistical Power. *Journal of Research on Educational Effectiveness*, *9*(4), 556–569. https://doi.org/10.1080/19345747.2015.1086916

Mitrushina, M. N., Boone, K. B., Razani, J. L., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment* (Second). Oxford University Press.

Moezzi, B., Lavrencic, L. M., Goldsworthy, M. R., Coussens, S., & Keage, H. A. D. (2019). Associations between EEG functional brain connectivity and a cognitive reserve proxy in healthy older adults. In *bioRxiv* (p. 625608). bioRxiv. https://doi.org/10.1101/625608

Moga, D. C., Beech, B. F., Abner, E. L., Schmitt, F. A., El Khouli, R. H., Martinez, A. I., Eckmann, L., Huffmyer, M., George, R., & Jicha, G. A. (2019). INtervention for Cognitive Reserve Enhancement in delaying the onset of Alzheimer's Symptomatic Expression (INCREASE), a randomized controlled trial: Rationale, study design, and protocol. *Trials*, *20*(1), 806. https://doi.org/10.1186/s13063-019-3993-0

Mondini, S., Madella, I., Zangrossi, A., Bigolin, A., Tomasi, C., Michieletto, M., Villani, D., Di Giovanni, G., & Mapelli, D. (2016). Cognitive Reserve in Dementia: Implications for Cognitive Training. *Frontiers in Aging Neuroscience*, *8*, 84. https://dx.doi.org/10.3389%2Ffnagi.2016.00084.

Monsch, A. U., Bondi, M. W., Salmon, D. P., Butters, N., Thal, L. J., Hansen, L. A., Wiederholt, W. C., Cahn, D. A., & Klauber, M. R. (1995). Clinical Validity of the Mattis Dementia Rating Scale in Detecting Dementia of the Alzheimer Type. *Archives of Neurology*, *52*(9), 899. https://doi.org/10.1001/archneur.1995.00540330081018

Morris, J. H., Sherman, J. D., & Mansfield, E. R. (1986). Failures to Detect Moderating Effects With Ordinary Least Squares-Moderated Multiple Regression. Some Reasons and a Remedy. *Psychological Bulletin*, *99*(2), 282–288. https://doi.org/10.1037/0033-2909.99.2.282

Mortimer, J. A., Borenstein, A. R., Gosche, K. M., & Snowdon, D. A. (2005). Very early detection of Alzheimer neuropathology and the role of brain reserve in modifying its

clinical expression. *Journal of Geriatric Psychiatry and Neurology*, *18*(4), 218–223. https://doi.org/10.1177/0891988705281869

Mosca, I., & Wright, R. E. (2019). The Long-Term Consequences of the Irish Marriage Bar. *Institute of Labor Economics (IZA) Discussion Papers*, 12301. http://hdl.handle.net/10419/196799

Mungas, D., Gavett, B., Fletcher, E., Farias, S. T., DeCarli, C., & Reed, B. (2018). Education amplifies brain atrophy effect on cognitive decline: implications for cognitive reserve. *Neurobiology of Aging*, *68*, 142–150. https://doi.org/10.1016/j.neurobiolaging.2018.04.002

Mwangi, B., Tian, T. S., & Soares, J. C. (2014). A review of feature reduction techniques in neuroimaging. *Neuroinformatics*, *12*(2), 229–244. https://doi.org/10.1007/s12021-013-9204-3

Negash, S., Wilson, R., Leurgans, S., Wolk, D., Schneider, J., Buchman, A., Bennett, D., & Arnold, S. (2013). Resilient Brain Aging: Characterization of Discordance between Alzheimer's Disease Pathology and Cognition. *Current Alzheimer Research*, *10*(8), 844–851. https://doi.org/10.2174/15672050113109990157

Nelson, H. E., & Willinson, J. (1982). *The National Adult Reading Test (NART): Test Manual.* Nelson.

Nenadić, I., Dietzek, M., Langbein, K., Sauer, H., & Gaser, C. (2017). BrainAGE score indicates accelerated brain aging in schizophrenia, but not bipolar disorder. *Psychiatry Research. Neuroimaging*, *266*, 86–89. https://doi.org/10.1016/j.pscychresns.2017.05.006

Neth, B. J., Graff-Radford, J., Mielke, M. M., Przybelski, S. A., Lesnick, T. G., Schwarz, C. G., Reid, R. I., Senjem, M. L., Lowe, V. J., Machulda, M. M., Petersen, R. C., Jr, C. R. J., Knopman, D. S., & Vemuri, P. (2020). Relationship Between Risk Factors and Brain Reserve in Late Middle Age: Implications for Cognitive Aging. *Frontiers in Aging Neuroscience*, *11*, 355. https://doi.org/10.3389/fnagi.2019.00355

Nichols, E., Szoeke, C. E. I., Vollset, S. E., Abbasi, N., Abd-Allah, F., Abdela, J., Aichour, M. T. E., Akinyemi, R. O., Alahdab, F., Asgedom, S. W., Awasthi, A., Barker-Collo, S. L., Baune, B. T., Béjot, Y., Belachew, A. B., Bennett, D. A., Biadgo, B., Bijani, A., Bin Sayeed, M. S., … Murray, C. J. L. (2019). Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*, *18*(1), 88–106. https://doi.org/10.1016/S1474-4422(18)30403-4

Nilsson, J., & Lövdén, M. (2018). Naming is not explaining: future directions for the cognitive reserve and brain maintenance theories. *Alzheimer's Research & Therapy*, *10*(1), 34. https://doi.org/10.1186/s13195-018-0365-z

Noble, S., Spann, M. N., Tokoglu, F., Shen, X., Constable, R. T., & Scheinost, D. (2017). Influences on the Test-Retest Reliability of Functional Connectivity MRI and its Relationship with Behavioral Utility. *Cerebral Cortex*, *27*(11), 5415–5429. https://doi.org/10.1093/cercor/bhx230

Nucci, M., Mapelli, D., & Mondini, S. (2012). Cognitive Reserve Index questionnaire (CRIq): A new instrument for measuring cognitive reserve. *Aging Clinical and Experimental Research, 24*(3), 218-226. https://doi.org/10.3275/7800

Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U., & Bäckman, L. (2012). Memory aging and brain maintenance. *Trends in Cognitive Sciences*, *16*(5), 292–305. https://doi.org/10.1016/j.tics.2012.04.005

O'Connor, D., Lake, E. M. R., Scheinost, D., Constable, R. T., & Todd Constable, R. (2020). Bootstrap aggregating improves the generalizability of Connectome Predictive Modelling. *BioRxiv*, 2020.07.08.193664. https://doi.org/10.1101/2020.07.08.193664

O'Leary, E. (1987). The Irish National Teachers' Organisation And The Marriage Bar For Women National Teachers, 1933-1958. *Saothar*, *12*, 47–52.

Ogbole, G. I., Adeyomoye, A. O., Badu-Peprah, A., Mensah, Y., & Nzeh, D. A. (2018). Survey of magnetic resonance imaging availability in West Africa. *Pan African Medical Journal*, *30*, 240. https://doi.org/10.11604/pamj.2018.30.240.14000

Ogino, E., Manly, J. J., Schupf, N., Mayeux, R., & Gu, Y. (2019). Current and past leisure time physical activity in relation to risk of Alzheimer's disease in older adults. *Alzheimer's and Dementia*, *15*(12), 1603–1611. https://doi.org/10.1016/j.jalz.2019.07.013

Oh, H., Razlighi, Q. R., & Stern, Y. (2018). Multiple pathways of reserve simultaneously present in cognitively normal older adults. *Neurology*, *90*(3), e197–e205. https://doi.org/10.1212/WNL.0000000000004829

Öktem, O. (1992). A verbal test of memory processes: A preliminary study. *Archives of Neuropsychiatry*, *29*(4), 196–206.

Oosterman, J. M., Jansen, M. G., Scherder, E. J. A., & Kessels, R. P. C. (2020). Cognitive reserve relates to executive functioning in the old–old. *Aging Clinical and Experimental Research*, 1–6. https://doi.org/10.1007/s40520-020-01758-y

Oosterman, J. M., Vogels, R. L. C. C., Van Harten, B., Gouw, A. A., Poggesi, A., Scheltens, P., Kessels, R. P. C. C., & Scherder, E. J. A. A. (2010). Assessing mental flexibility: neuroanatomical and neuropsychological correlates of the trail making test in elderly people. *The Clinical Neuropsychologist*, *24*(2), 203–219. https://doi.org/10.1080/13854040903482848

Opdebeeck, C., Martyr, A., & Clare, L. (2016). Cognitive reserve and cognitive function in healthy older people: a meta-analysis. *Aging, Neuropsychology, and Cognition*, *23*(1), 40–60. https://doi.org/10.1080/13825585.2015.1041450

Paillard-Borg, S., Fratiglioni, L., Winblad, B., & Wang, H. X. (2009). Leisure activities in late life in relation to dementia risk: Principal component analysis. *Dementia and Geriatric Cognitive Disorders*, *28*(2), 136–144. https://doi.org/10.1159/000235576

Park, S., Choi, B., Choi, C., Kang, J. M., & Lee, J. Y. (2019). Relationship between education, leisure activities, and cognitive functions in older adults. *Aging and Mental Health*, *23*(12), 1651–1660. https://doi.org/10.1080/13607863.2018.1512083

Pavlik, V. N., Doody, R. S., Massman, P. J., & Chan, W. (2006). Influence of Premorbid IQ and Education on Progression of Alzheimer's Disease. *Dementia and Geriatric Cognitive Disorders*, *22*(4), 367–377. https://doi.org/10.1159/000095640

Peeters, G., Kenny, R. A., & Lawlor, B. (2020). Late life education and cognitive function in older adults. *International Journal of Geriatric Psychiatry*, gps.5281. https://doi.org/10.1002/gps.5281

Pellegrini, E., Ballerini, L., Hernandez, M. del C. V., Chappell, F. M., González-Castro, V., Anblagan, D., Danso, S., Muñoz-Maniega, S., Job, D., Pernet, C., Mair, G., MacGillivray, T. J., Trucco, E., & Wardlaw, J. M. (2018). Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, *10*, 519–535. https://doi.org/10.1016/j.dadm.2018.07.004

Perneczky, R., Kempermann, G., Korczyn, A. D., Matthews, F. E., Ikram, M. A., Scarmeas, N., Chetelat, G., Stern, Y., & Ewers, M. (2019). Translational research on reserve against neurodegenerative disease: Consensus report of the International Conference on Cognitive Reserve in the Dementias and the Alzheimer's Association Reserve, Resilience and Protective Factors Professional Interest Area. *BMC Medicine*. https://doi.org/10.1186/s12916-019-1283-z

Pettigrew, C., Soldan, A., Li, S., Lu, Y., Wang, M. C., Selnes, O. A., Moghekar, A., O'Brien, R., & Albert, M. (2013). Relationship of cognitive reserve and APOE status

to the emergence of clinical symptoms in preclinical Alzheimer's disease. *Cognitive Neuroscience*, *4*(3–4), 136–142. https://doi.org/10.1080/17588928.2013.831820

Pettigrew, C., Soldan, A., Zhu, Y., Wang, M. C., Brown, T., Miller, M., & Albert, M. (2017). Cognitive reserve and cortical thickness in preclinical Alzheimer's disease. *Brain Imaging and Behavior*, *11*(2), 357–367. https://doi.org/10.1007/s11682-016-9581-y

Pfeffer, R. I., Kurosaki, T. T., Chance, J. M., Filos, S., & Bates, D. (1984). Use of the mental function index in older adults: Reliability, validity, and measurement of change over time. *American Journal of Epidemiology*, *120*(6), 922–935. https://doi.org/10.1093/oxfordjournals.aje.a113963

Poldrack, R. A., Huckins, G., & Varoquaux, G. (2020). Establishment of Best Practices for Evidence for Prediction: A Review. *JAMA Psychiatry*, *77*(5), 534–540. https://doi.org/10.1001/jamapsychiatry.2019.3671

Potter, G. G., Helms, M. J., Burke, J. R., Steffens, D. C., & Plassman, B. L. (2007). Job demands and dementia risk among male twin pairs. *Alzheimer's and Dementia*, *3*(3), 192–199. https://doi.org/10.1016/j.jalz.2007.04.377

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, *59*(3), 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 320-341. https://doi.org/10.1016/j.neuroimage.2013.08.048

Ranganathan, P., Pramesh, C. S., & Buyse, M. (2016). Common pitfalls in statistical analysis: The perils of multiple testing. *Perspectives in Clinical Research*, *7*(2), 106–107. https://doi.org/10.4103/2229-3485.179436

Reed, B. R., Dowling, M., Tomaszewski Farias, S., Sonnen, J., Strauss, M., Schneider, J. A., Bennett, D. A., & Mungas, D. (2011). Cognitive activities during adulthood are more important than education in building reserve. *Journal of the International Neuropsychological Society*, *17*(4), 615–624. https://doi.org/10.1017/S1355617711000014

Reed, B. R., Mungas, D., Farias, S. T., Harvey, D., Beckett, L., Widaman, K., Hinton, L., & DeCarli, C. (2010). Measuring cognitive reserve based on the decomposition of episodic memory variance. *Brain*. https://doi.org/10.1093/brain/awq154

Reitan, R. M. (1955). The relation of the Trail Making Test to organic brain damage.

Journal of Consulting Psychology. https://doi.org/10.1037/h0044509

Ren, Z., Daker, R. J., Shi, L., Sun, J., Beaty, R. E., Wu, X., Chen, Q., Yang, W., Lyons, I. M., Green, A. E., & Qiu, J. (2021). Connectome-Based Predictive Modeling of Creativity Anxiety. *NeuroImage*, *225*, 117469. https://doi.org/https://doi.org/10.1016/j.neuroimage.2020.117469

Rentz, D. M., Mormino, E. C., Papp, K. V., Betensky, R. A., Sperling, R. A., & Johnson, K. A. (2017). Cognitive resilience in clinical and preclinical Alzheimer's disease: the Association of Amyloid and Tau Burden on cognitive performance. *Brain Imaging and Behavior*, *11*(2), 383–390. https://doi.org/10.1007/s11682-016-9640-4

Resende, E. de P. F., Rosen, H. J., Chiang, K., Staffaroni, A. M., Allen, I., Grinberg, L. T., Carmona, K. C., Guimarães, H. C., Carvalho, V. A., Barbosa, M. T., de Souza, L. C., & Caramelli, P. (2018). Primary School Education May Be Sufficient to Moderate a Memory-Hippocampal Relationship. *Frontiers in Aging Neuroscience*, *10*, 381. https://doi.org/10.3389/fnagi.2018.00381

Richard, G., Kolskår, K., Sanders, A.-M., Kaufmann, T., Petersen, A., Doan, N. T., Monereo Sánchez, J., Alnæs, D., Ulrichsen, K. M., Dørum, E. S., Andreassen, O. A., Nordvik, J. E., & Westlye, L. T. (2018). Assessing distinct patterns of cognitive aging using tissue-specific brain age prediction based on diffusion tensor imaging and brain morphometry. *PeerJ*, *6*, e5908. https://doi.org/10.7717/peerj.5908

Richardson, T. J., Lee, S. J., Berg-Weger, M., & Grossberg, G. T. (2013). Caregiver health: Health of caregivers of Alzheimer's and other dementia patients topical collection on geriatric disorders. *Current Psychiatry Reports*, *15*(7), 1–7. https://doi.org/10.1007/s11920-013-0367-2

Ridgway, G. R., Omar, R., Ourselin, S., Hill, D. L. G., Warren, J. D., & Fox, N. C. (2009). Issues with threshold masking in voxel-based morphometry of atrophied brains. *NeuroImage*, *44*(1), 99–111. https://doi.org/10.1016/J.NEUROIMAGE.2008.08.045

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). `Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747–758. https://doi.org/10.1016/S0028-3932(97)00015-8

Rogers, W. M. (2002). Theoretical and Mathematical Constraints of Interactive Regression Models. *Organizational Research Methods*, *5*(3), 212–230. https://doi.org/10.1177/1094428102005003002

Roldán-Tapia, L., García, J., Cánovas, R., & León, I. (2012). Cognitive Reserve, Age, and

Their Relation to Attentional and Executive Functions. *Applied Neuropsychology: Adult*, *19*(1), 2–8. https://doi.org/10.1080/09084282.2011.595458

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*, *19*(1), 165–171. https://doi.org/10.1038/nn.4179

Rovio, S., Kåreholt, I., Helkala, E. L., Viitanen, M., Winblad, B., Tuomilehto, J., Soininen, H., Nissinen, A., & Kivipelto, M. (2005). Leisure-time physical activity at midlife and the risk of dementia and Alzheimer's disease. *Lancet Neurology*, *4*(11), 705–711. https://doi.org/10.1016/S1474-4422(05)70198-8

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology*, *85*(1), 112–118. https://doi.org/10.1037/0021-9010.85.1.112

Saeb, S., Lonini, L., Jayaraman, A., Mohr, D. C., & Kording, K. P. (2016). Voodoo Machine Learning for Clinical Predictions. *BioRxiv*, 059774. https://doi.org/10.1101/059774

Salehi, M., Greene, A. S., Karbasi, A., Shen, X., Scheinost, D., & Constable, R. T. (2020). There is no single functional atlas even for a single individual: Functional parcel definitions change with task. *NeuroImage*, *208*, 116366. https://doi.org/10.1016/j.neuroimage.2019.116366

Sallim, A. Bin, Sayampanathan, A. A., Cuttilan, A., & Chun-Man Ho, R. (2015). Prevalence of Mental Health Disorders Among Caregivers of Patients With Alzheimer Disease. *Journal of the American Medical Directors Association*, *16*(12), 1034–1041. https://doi.org/10.1016/j.jamda.2015.09.007

Saloner, R., Casaletto, K. B., Marx, G., Dutt, S., Vanden Bussche, A. B., You, M., Fox, E., Stiver, J., & Kramer, J. H. (2018). Performance on a 1-week delayed recall task is associated with medial temporal lobe structures in neurologically normal older adults. *Clinical Neuropsychologist*, *32*(3), 456–467. https://doi.org/10.1080/13854046.2017.1370134

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, *30*(4), 507–514. https://doi.org/https://doi.org/10.1016/j.neurobiolaging.2008.09.023

Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society : JINS*, *16*(5), 754–760. https://doi.org/10.1017/S1355617710000706

Santos Nogueira, D., Azevedo Reis, E., & Vieira, A. (2016). Verbal Fluency Tasks: Effects of Age, Gender, and Education. *Folia Phoniatrica et Logopaedica*, *68*(3), 124–133. https://doi.org/10.1159/000450640

Sarracanie, M., Lapierre, C. D., Salameh, N., Waddington, D. E. J., Witzel, T., & Rosen, M. S. (2015). Low-Cost High-Performance MRI. *Scientific Reports*, *5*(1), 1–9. https://doi.org/10.1038/srep15177

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*(1), 240–256. https://doi.org/10.1016/j.neuroimage.2012.08.052

Savva, G. M., Maty, S. C., Setti, A., & Feeney, J. (2013). Cognitive and Physical Health of the Older Populations of England, the United States, and Ireland: International Comparability of The Irish Longitudinal Study on Ageing. *Journal of the American Geriatrics Society*, *61*, S291–S298. https://doi.org/10.1111/jgs.12196

Scarmeas, N., Luchsinger, J. A., Schupf, N., Brickman, A. M., Cosentino, S., Tang, M. X., & Stern, Y. (2009). Physical activity, diet, and risk of Alzheimer disease. *JAMA - Journal of the American Medical Association*, *302*(6), 627–637. https://doi.org/10.1001/jama.2009.1144

Scarmeas, N., Zarahn, E., Anderson, K. E., Habeck, C. G., Hilton, J., Flynn, J., Marder, K. S., Bell, K. L., Sackeim, H. A., Van Heertum, R. L., Moeller, J. R., & Stern, Y. (2003). Association of life activities with cerebral blood flow in Alzheimer disease: Implications for the cognitive reserve hypothesis. *Archives of Neurology*, *60*(3), 359–365. https://doi.org/10.1001/archneur.60.3.359

Scheinost, D., Hsu, T. W., Avery, E. W., Hampson, M., Constable, R. T., Chun, M. M., & Rosenberg, M. D. (2020). Connectome-based neurofeedback: A pilot study to improve sustained attention. *NeuroImage*, *212*, 116684. https://doi.org/10.1016/j.neuroimage.2020.116684

Scheinost, D., Noble, S., Horien, C., Greene, A. S., Lake, E. M., Salehi, M., Gao, S., Shen, X., O'Connor, D., Barron, D. S., Yip, S. W., Rosenberg, M. D., & Constable, R. T. (2019). Ten simple rules for predictive modeling of individual differences in neuroimaging. *NeuroImage*, *193*, 35–45. https://doi.org/10.1016/j.neuroimage.2019.02.057

Scheller, E., Schumacher, L. V, Peter, J., Lahr, J., Wehrle, J., Kaller, C. P., Gaser, C., &

Klöppel, S. (2018). Brain Aging and APOE ε4 Interact to Reveal Potential Neuronal Compensation in Healthy Older Adults. *Frontiers in Aging Neuroscience*, *10*, 74. https://doi.org/10.3389/fnagi.2018.00074

Schnack, H. G., & Kahn, R. S. (2016). Detecting Neuroimaging Biomarkers for Psychiatric Disorders: Sample Size Matters. *Frontiers in Psychiatry*, *7*, 50. https://doi.org/10.3389/fpsyt.2016.00050

Schnack, H. G., van Haren, N. E. M., Nieuwenhuis, M., Hulshoff Pol, H. E., Cahn, W., & Kahn, R. S. (2016). Accelerated Brain Aging in Schizophrenia: A Longitudinal Pattern Recognition Study. *American Journal of Psychiatry*, *173*(6), 607–616. https://doi.org/10.1176/appi.ajp.2015.15070922

Schwartz, C. E., Rapkin, B. D., & Healy, B. C. (2016). Reserve and Reserve-building activities research: Key challenges and future directions. *BMC Neuroscience*, *17*(1), 62. https://doi.org/10.1186/s12868-016-0297-0

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. *Proceedings of the 9th Python in Science Conference*.

Seignourel, P. J., Kunik, M. E., Snow, L., Wilson, N., & Stanley, M. (2008). Anxiety in dementia: A critical review. *Clinical Psychology Review*, 28*(7)*, 1071–1082. https://doi.org/10.1016/j.cpr.2008.02.008

Serra, L., Mancini, M., Cercignani, M., Di Domenico, C., Spanò, B., Giulietti, G., Koch, G., Marra, C., & Bozzali, M. (2016). Network-based substrate of cognitive reserve in Alzheimer's disease. *Journal of Alzheimer's Disease*, *55*(1), 421–430. https://doi.org/10.3233/JAD-160735

Seubert, J., Freiherr, J., Frasnelli, J., Hummel, T., & Lundstrom, J. N. (2013). Orbitofrontal Cortex and Olfactory Bulb Volume Predict Distinct Aspects of Olfactory Performance in Healthy Subjects. *Cerebral Cortex*, *23*(10), 2448–2456. https://doi.org/10.1093/cercor/bhs230

Shehzad, Z., Kelly, A. M. C., Reiss, P. T., Gee, D. G., Gotimer, K., Uddin, L. Q., Lee, S. H., Margulies, D. S., Roy, A. K., Biswal, B. B., Petkova, E., Castellanos, F. X., & Milham, M. P. (2009). The Resting Brain: Unconstrained yet Reliable. *Cerebral Cortex*, *19*(10), 2209–2229. https://doi.org/10.1093/cercor/bhn256

Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., & Constable, R. T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, *12*(3), 506–518. https://doi.org/10.1038/nprot.2016.178

Shen, X., Tokoglu, F., Papademetris, X., & Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage*, *82*, 403–415. https://doi.org/10.1016/j.neuroimage.2013.05.081

Simon, S. S., Hampstead, B. M., Nucci, M. P., Ferreira, L. K., Duran, F. L. S., Fonseca, L. M., Martin, M. da G. M., Ávila, R., Porto, F. H. G., Brucki, S. M. D., Martins, C. B., Tascone, L. S., Jr., E. A., Busatto, G. F., & Bottino, C. M. C. (2020). Mnemonic strategy training modulates functional connectivity at rest in mild cognitive impairment: Results from a randomized controlled trial. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *6*(1), e12075. https://doi.org/10.1002/trc2.12075

Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, *75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Skocik, M., Collins, J., Callahan-Flintoft, C., Bowman, H., & Wyble, B. (2016). I tried a bunch of things: The dangers of unexpected overfitting in classification. *BioRxiv*, 078816. https://doi.org/10.1101/078816

Smart, E. L., Gow, A. J., & Deary, I. J. (2014). Occupational complexity and lifetime cognitive abilities. *Neurology*, *83*(24), 2285–2291. https://doi.org/10.1212/WNL.0000000000001075

Smith, S. M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T. E., & Miller, K. L. (2019). Estimation of Brain Age Delta from Brain Imaging. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2019.06.017

Snyder, J. C., Mika, S., Burke, K., & Müller, K.-R. (2013). Kernels, Pre-images and Optimization. In B. Schölkopf, Z. Luo, & V. Vovk (Eds.), *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik* (pp. 245–259). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41136-6_21

Soldan, A., Pettigrew, C., Cai, Q., Wang, J., Wang, M. C., Moghekar, A., Miller, M. I., & Albert, M. (2017). Cognitive reserve and long-term change in cognition in aging and preclinical Alzheimer's disease. *Neurobiology of Aging*, *60*, 164–172. https://doi.org/10.1016/j.neurobiolaging.2017.09.002

Solomon, A., Mangialasche, F., Richard, E., Andrieu, S., Bennett, D. A., Breteler, M., Fratiglioni, L., Hooshmand, B., Khachaturian, A. S., Schneider, L. S., Skoog, I., & Kivipelto, M. (2014). Advances in the prevention of Alzheimer's disease and dementia. *Journal of Internal Medicine*, *275*(3), 229–250.

https://doi.org/10.1111/joim.12178

Sommerlad, A., Sabia, S., Livingston, G., Kivimäki, M., Lewis, G., & Singh-Manoux, A. (2020). Leisure activity participation and risk of dementia: An 18-year follow-up of the Whitehall II Study. *Neurology*, *95*(20), e2803–e2815. https://doi.org/10.1212/WNL.0000000000010966

Sowman, P. F., Ryan, M., Johnson, B. W., Savage, G., Crain, S., Harrison, E., Martin, E., & Burianová, H. (2017). Grey matter volume differences in the left caudate nucleus of people who stutter. *Brain and Language*, *164*, 9–15. https://doi.org/10.1016/j.bandl.2016.08.009

Speer, M. E., & Soldan, A. (2015). Cognitive reserve modulates ERPs associated with verbal working memory in healthy younger and older adults. *Neurobiology of Aging*, *36*(3), 1424–1434. https://doi.org/10.1016/j.neurobiolaging.2014.12.025

Springer, M. V., McIntosh, A. R., Winocur, G., & Grady, C. L. (2005). The relation between brain activity during memory tasks and years of education in young and older adults. *Neuropsychology*, *19*(2), 181–192. https://doi.org/10.1037/0894-4105.19.2.181

Sprott, R. L. (2010). Biomarkers of aging and disease: Introduction and definitions. *Experimental Gerontology*, *45*(1), 2–4. https://doi.org/10.1016/j.exger.2009.07.008

Staff, R. T., Murray, A. D., Deary, I. J., & Whalley, L. J. (2004). What provides cerebral reserve? *Brain*, *127*(5), 1191–1199. https://doi.org/10.1093/brain/awh144

Steffener, J., Barulli, D., Habeck, C., O'Shea, D., Razlighi, Q., & Stern, Y. (2014). The Role of Education and Verbal Abilities in Altering the Effect of Age-Related Gray Matter Differences on Cognition. *PLoS ONE*, *9*(3), e91196. https://doi.org/10.1371/journal.pone.0091196

Steffener, J., Habeck, C., O'Shea, D., Razlighi, Q., Bherer, L., & Stern, Y. (2016). Differences between chronological and brain age are related to education and self-reported physical activity. *Neurobiology of Aging*, *40*, 138–144. https://doi.org/10.1016/j.neurobiolaging.2016.01.014

Steffener, J., Reuben, A., Rakitin, B. C., & Stern, Y. (2011). Supporting performance in the face of age-related neural changes: Testing mechanistic roles of cognitive reserve. *Brain Imaging and Behavior*, *5*(3), 212–221. https://doi.org/10.1007/s11682-011-9125-4

Steffener, J., & Stern, Y. (2012). Exploring the neural basis of cognitive reserve in aging. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, *1822*(3), 467–473. https://doi.org/10.1016/j.bbadis.2011.09.012

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society : JINS*, *8*(3), 448–460. https://doi.org/10.1017/S1355617702813248

Stern, Y. (2005). Brain Networks Associated with Cognitive Reserve in Healthy Young and Old Adults. *Cerebral Cortex*, *15*(4), 394–402. https://doi.org/10.1093/cercor/bhh142

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia*, *47*(10), 2015–2028.

Stern, Y. (2012). Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurology*, *11*(11), 1006–1012. https://doi.org/10.1016/S1474-4422(12)70191-6

Stern, Y., Arenaza-Urquijo, E. M., Bartrés-Faz, D., Belleville, S., Cantilon, M., Chetelat, G., Ewers, M., Franzmeier, N., Kempermann, G., Kremen, W. S., Okonkwo, O., Scarmeas, N., Soldan, A., Udeh-Momoh, C., Valenzuela, M., Vemuri, P., Vuoksimaa, E., Arenaza Urquiljo, E. M., Bartrés-Faz, D., … Vuoksimaa, E. (2020). Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimer's and Dementia*, *16*(9), 1305–1311. https://doi.org/10.1016/j.jalz.2018.07.219

Stern, Y., & Barulli, D. (2019). Cognitive reserve. In S. T. Dekosky & S. Asthana (Eds.), *Geriatric Neurology* (pp. 181–190). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-804766-8.00011-X

Stern, Y., Gazes, Y., Razlighi, Q., Steffener, J., & Habeck, C. (2018). A task-invariant cognitive reserve network. *NeuroImage*, *178*, 36–45. https://doi.org/10.1016/j.neuroimage.2018.05.033

Stern, Y., & Habeck, C. (2018). Deriving and testing the validity of cognitive reserve candidates. In G. Perneczky, Robert (Ed.), *Neuromethods* (Vol. 137, pp. 63–70). Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-7674-4_4

Stern, Y., Habeck, C., Steffener, J., Barulli, D., Gazes, Y., Razlighi, Q., Shaked, D., & Salthouse, T. (2014). The Reference Ability Neural Network Study: Motivation, design, and initial feasibility analyses. *NeuroImage*, *103*, 139–151. https://doi.org/10.1016/j.neuroimage.2014.09.029

Stern, Y., Mackay-Brandt, A., Lee, S., McKinley, P., McIntyre, K., Razlighi, Q., Agarunov, E., Bartels, M., & Sloan, R. P. (2019). Effect of aerobic exercise on cognition in younger adults: A randomized clinical trial. *Neurology*, *92*(9), E905–E916. https://doi.org/10.1212/WNL.0000000000007003

Strauss, E. H., Sherman, E. M. S., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests; Administration Norms And Commentary* (3rd ed.). Oxford

University Press. https://doi.org/10.1016/j.jvolgeores.2008.06.015

Subramaniapillai, S., Almey, A., Natasha Rajah, M., & Einstein, G. (2021). Sex and gender differences in cognitive and brain reserve: Implications for Alzheimer's disease in women. *Frontiers in Neuroendocrinology*, *60*, 100879. https://doi.org/10.1016/j.yfrne.2020.100879

Sundermann, E. E., Maki, P. M., Rubin, L. H., Lipton, R. B., Landau, S., & Biegon, A. (2016). Female advantage in verbal memory. *Neurology*, *87*(18), 1916–1924. https://doi.org/10.1212/WNL.0000000000003288

Tombaugh, T. N., & McIntyre, N. J. (1992). The mini-mental state examination: a comprehensive review. *Journal of the American Geriatrics Society*, *40*(9), 922–935. https://doi.org/10.1111/j.1532-5415.1992.tb01992.x

Topiwala, A., Suri, S., Allan, C., Valkanova, V., Filippini, N., Sexton, C. E., Heise, V., Zsoldos, E., Mahmood, A., Singh-Manoux, A., Mackay, C. E., Kivimäki, M., & Ebmeier, K. P. (2019). Predicting cognitive resilience from midlife lifestyle and multi-modal MRI: A 30-year prospective cohort study. *PLOS ONE*, *14*(2), e0211273. https://doi.org/10.1371/journal.pone.0211273

Tsai, H. J., & Chang, F. K. (2019). Associations of exercise, nutritional status, and smoking with cognitive decline among older adults in Taiwan: Results of a longitudinal population-based study. *Archives of Gerontology and Geriatrics*, *82*, 133–138. https://doi.org/10.1016/j.archger.2018.12.008

Tsapanou, A., Habeck, C., Gazes, Y., Razlighi, Q., Sakhardande, J., Stern, Y., & Salthouse, T. A. (2019). Brain biomarkers and cognition across adulthood. *Human Brain Mapping*, *40*(13), hbm.24634. https://doi.org/10.1002/hbm.24634

Tumac, A. (1997). *The effect of age and education on performance in some tests that are sensitive to frontal damage in normal subjects* [Master's thesis]. Istanbul University. http://nek.istanbul.edu.tr:4444/ekos/TEZ/31348.pdf

United States Employment Service. *Dictionary of Occupational Titles* (4th ed.). Superintendent of Documents, US Government Printing Office.

Uttl, B., Graf, P., & Richter, L. K. (2002). Verbal Paired Associates tests limits on validity and reliability. *Archives of Clinical Neuropsychology : The Official Journal of the National Academy of Neuropsychologists*, *17*(6), 567–581.

Valenzuela, M. J., & Sachdev, P. (2007). Assessment of complex mental activity across the lifespan: Development of the Lifetime of Experiences Questionnaire (LEQ). *Psychological Medicine*. https://doi.org/10.1017/S003329170600938X

Van Der Flier, W. M., & Scheltens, P. (2005). Epidemiology and risk factors of dementia. *Neurology in Practice*, *76*, (5), 2–7. https://doi.org/10.1136/jnnp.2005.082867

Van Dijk, K. R. A., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W., & Buckner, R. L. (2010). Intrinsic functional connectivity as a tool for human connectomics: Theory, properties, and optimization. *Journal of Neurophysiology*, *103*(1), 297–321. https://doi.org/10.1152/jn.00783.2009

van Dinteren, R., Arns, M., Jongsma, M. L. A., & Kessels, R. P. C. (2014). P300 Development across the Lifespan: A Systematic Review and Meta-Analysis. *PLoS ONE*, *9*(2), e87347. https://doi.org/10.1371/journal.pone.0087347

van Loenhoud, A. C., Groot, C., Vogel, J. W., Van Der Flier, W. M., & Ossenkoppele, R. (2018). Is intracranial volume a suitable proxy for brain reserve? *Alzheimer's Research and Therapy*, *10*(1). https://doi.org/10.1186/s13195-018-0408-5

van Loenhoud, A. C., Habeck, C., van der Flier, W. M., Ossenkoppele, R., & Stern, Y. (2020). Identifying a task-invariant cognitive reserve network using task potency. *NeuroImage*, *210*, 116593. https://doi.org/10.1016/j.neuroimage.2020.116593

van Loenhoud, A. C., Van Der Flier, W. M., Wink, A. M., Dicks, E., Groot, C., Twisk, J., Barkhof, F., Scheltens, P., & Ossenkoppele, R. (2019). Cognitive reserve and clinical progression in Alzheimer disease: A paradoxical relationship. *Neurology*, *93*(4), E334–E346. https://doi.org/10.1212/WNL.0000000000007821

van Loenhoud, A. C., Wink, A. M., Groot, C., Verfaillie, S. C. J., Twisk, J., Barkhof, F., van Berckel, B., Scheltens, P., van der Flier, W. M., & Ossenkoppele, R. (2017). A neuroimaging approach to capture cognitive reserve: Application to Alzheimer's disease. *Human Brain Mapping*, *38*(9), 4703–4715. https://doi.org/10.1002/hbm.23695

van Straaten, E. C. W., & Stam, C. J. (2013). Structure out of chaos: Functional brain network analysis with EEG, MEG, and functional MRI. *European Neuropsychopharmacology*, *23*(1), 7–18. https://doi.org/10.1016/j.euroneuro.2012.10.010

Varikuti, D. P., Genon, S., Sotiras, A., Schwender, H., Hoffstaedter, F., Patil, K. R., Jockwitz, C., Caspers, S., Moebus, S., Amunts, K., Davatzikos, C., & Eickhoff, S. B. (2018). Evaluation of non-negative matrix factorization of grey matter in age prediction. *NeuroImage*, *173*(March), 394–410. https://doi.org/10.1016/j.neuroimage.2018.03.007

Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., &

Thirion, B. (2017). Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage*, *145*, 166–179. https://doi.org/10.1016/J.NEUROIMAGE.2016.10.038

Vaucher, P., Herzig, D., Cardoso, I., Herzog, M. H., Mangin, P., & Favrat, B. (2014). The trail making test as a screening instrument for driving performance in older drivers; a translational research. *BMC Geriatrics*, *14*(1), 123. https://doi.org/10.1186/1471-2318-14-123

Vellas, B., Andrieu, S., Sampaio, C., Coley, N., & Wilcock, G. (2008). Endpoints for trials in Alzheimer's disease: a European task force consensus. *The Lancet Neurology*, *7*(5), 436–450). https://doi.org/10.1016/S1474-4422(08)70087-5

Vernooij, M. W., Ikram, M. A., Vrooman, H. A., Wielopolski, P. A., Krestin, G. P., Hofman, A., Niessen, W. J., Van Der Lugt, A., & Breteler, M. M. B. (2009). White Matter microstructural integrity and cognitive function in a general elderly population. *Archives of General Psychiatry*, *66*(5), 545–553. https://doi.org/10.1001/archgenpsychiatry.2009.5

Vuoksimaa, E., Panizzon, M. S., Chen, C. H., Eyler, L. T., Fennema-Notestine, C., Fiecas, M. J. A., Fischl, B., Franz, C. E., Grant, M. D., Jak, A. J., Lyons, M. J., Neale, M. C., Thompson, W. K., Tsuang, M. T., Xian, H., Dale, A. M., & Kremen, W. S. (2013). Cognitive reserve moderates the association between hippocampal volume and episodic memory in middle age. *Neuropsychologia*, *51*(6), 1124–1131. https://doi.org/10.1016/j.neuropsychologia.2013.02.022

Wallace, R. B., & Herzog, A. R. (1995). Overview of the Health Measures in the Health and Retirement Study. *The Journal of Human Resources*, *30*, S84. https://doi.org/10.2307/146279

Wang, H. X., Gustafson, D. R., Kivipelto, M., Pedersen, N. L., Skoog, I., Windblad, B., & Fratiglioni, L. (2012). Education halves the risk of dementia due to apolipoprotein ε4 allele: A collaborative study from the Swedish Brain Power initiative. *Neurobiology of Aging*, *33*(5), 1007.e1-1007.e7. https://doi.org/10.1016/j.neurobiolaging.2011.10.003

Wang, H. X., Karp, A., Winblad, B., & Fratiglioni, L. (2002). Late-Life Engagement in Social and Leisure Activities Is Associated with a Decreased Risk of Dementia: A Longitudinal Study from the Kungsholmen Project. *American Journal of Epidemiology*, *155*(12), 1081–1087. https://doi.org/10.1093/aje/155.12.1081

Wang, Y., Zou, Z., Song, H., Xu, X., Wang, H., d'Oleire Uquillas, F., & Huang, X. (2016). Altered Gray Matter Volume and White Matter Integrity in College Students with

Mobile Phone Dependence. *Frontiers in Psychology*, *7*, 597. https://doi.org/10.3389/fpsyg.2016.00597

Ward, D. D., Summers, M. J., Saunders, N. L., & Vickers, J. C. (2015). Modeling cognitive reserve in healthy middle-aged and older adults: The Tasmanian Healthy Brain Project. *International Psychogeriatrics*, *27*(4), 579–589. https://doi.org/10.1017/S1041610214002075

Washington, S. D., & VanMeter, J. W. (2015). Anterior-Posterior Connectivity within the Default Mode Network Increases During Maturation. *International Journal of Medical and Biological Frontiers*, *21*(2), 207–218.

Wechsler, D. (1987). *Wechsler Memory Scale - Revised manual.* The Psychological Corporation.

Wechsler, D. (1997). *WAIS-III administration and scoring manual.* The Psychological Corporation.

Weinstein, A. M., Voss, M. W., Prakash, R. S., Chaddock, L., Szabo, A., White, S. M., Wojcicki, T. R., Mailey, E., McAuley, E., Kramer, A. F., & Erickson, K. I. (2012). The association between aerobic fitness and executive function is mediated by prefrontal cortex volume. *Brain, Behavior, and Immunity*, *26*(5), 811–819. https://doi.org/10.1016/j.bbi.2011.11.008

Whelan, B. J., & Savva, G. M. (2013). Design and Methodology of The Irish Longitudinal Study on Ageing. *Journal of the American Geriatrics Society*, *61*, S265–S268. https://doi.org/10.1111/jgs.12199

Whelan, R., & Garavan, H. (2014). When Optimism Hurts: Inflated Predictions in Psychiatric Neuroimaging. *Biological Psychiatry*, *75*(9), 746–748. https://doi.org/10.1016/j.biopsych.2013.05.014

Whisman, M. A., & McClelland, G. H. (2005). Designing, testing, and interpreting interactions and moderator effects in family research. *Journal of Family Psychology*, *19*(1), 111–120. https://doi.org/10.1037/0893-3200.19.1.111

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2016). Verbal Fluency: Language or Executive Function Measure? *Applied Neuropsychology: Adult*, *23*(1), 29–34. https://doi.org/10.1080/23279095.2015.1004574

Whitty, E., Mansour, H., Aguirre, E., Palomo, M., Charlesworth, G., Ramjee, S., Poppe, M., Brodaty, H., Kales, H. C., Morgan-Trimmer, S., Nyman, S., Lang, I., Walters, K., Petersen, I., Wenborn, J., Minihane, A. M., Ritchie, K., Huntley, J., Walker, Z., &

Cooper, C. (2020). Efficacy of lifestyle and psychosocial interventions in reducing cognitive decline in older people: Systematic review. *Ageing Research Reviews*, *62*, p. 101113). https://doi.org/10.1016/j.arr.2020.101113

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. *Frontiers in Psychology*, *7*(NOV), 1832. https://doi.org/10.3389/fpsyg.2016.01832

Willer, L., Pedersen, P. M., Forchhammer, H. B., & Christensen, H. (2016). Cognitive assessment at bedside for iPad: A preliminary validation of a novel cognitive test for stroke patients. *European Stroke Journal*, *1*(4), 294–301. https://doi.org/10.1177/2396987316665233

Wilson, B. A., Watson, P. C., Baddeley, A. D., Emslie, H., & Evans, J. J. (2000). Improvement or simply practice? The effects of twenty repeated assessments on people with and without brain injury. *Journal of the International Neuropsychological Society : JINS*, *6*(4), 469–479.

Wilson, R. S., Beckett, L. A., Barnes, L. L., Schneider, J. A., Bach, J., Evans, D. A., & Bennett, D. A. (2002). Individual differences in rates of change in cognitive abilities of older persons. *Psychology and Aging*, *17*(2), 179–193. https://doi.org/10.1037/0882-7974.17.2.179

Woo, C. W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better biomarkers: brain models in translational neuroimaging. *Nature Neuroscience*, *20*(3), 365–377. https://doi.org/10.1038/nn.4478

Woods, D. L., Wyma, J. M., Herron, T. J., & Yund, E. W. (2016). Computerized Analysis of Verbal Fluency: Normative Data and the Effects of Repeated Testing, Simulated Malingering, and Traumatic Brain Injury. *PloS One*, *11*(12), e0166439. https://doi.org/10.1371/journal.pone.0166439

Woodward, N. D., & Cascio, C. J. (2015). Resting-state functional connectivity in psychiatric disorders. *JAMA Psychiatry*, *72*(8), 743–744. https://doi.org/10.1001/jamapsychiatry.2015.0484

Xu, W., Tan, L., Wang, H. F., Tan, M. S., Tan, L., Li, J. Q., Zhao, Q. F., & Yu, J. T. (2016). Education and Risk of Dementia: Dose-Response Meta-Analysis of Prospective Cohort Studies. *Molecular Neurobiology*, *53*(5), 3113–3123. https://doi.org/10.1007/s12035-015-9211-5

Yaffe, K., Falvey, C., Harris, T. B., Newman, A., Satterfield, S., Koster, A., Ayonayon, H.,

& Simonsick, E. (2013). Effect of socioeconomic disparities on incidence of dementia among biracial older adults: Prospective study. *BMJ (Online)*, *347*. https://doi.org/10.1136/bmj.f7051

Yan, C. G., Cheung, B., Kelly, C., Colcombe, S., Craddock, R. C., Di Martino, A., Li, Q., Zuo, X. N., Castellanos, F. X., & Milham, M. P. (2013). A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *NeuroImage*, *76*, 183–201. https://doi.org/10.1016/j.neuroimage.2013.03.004

Yang, C. Y., & Lin, C. P. (2020). Classification of cognitive reserve in healthy older adults based on brain activity using support vector machine. *Physiological Measurement*, *41*(6), 065009. https://doi.org/10.1088/1361-6579/ab979e

Yeung, C. M., St John, P. D., Menec, V., & Tyas, S. L. (2014). Is bilingualism associated with a lower risk of dementia in community-living older adults? cross-sectional and prospective analyses. *Alzheimer Disease and Associated Disorders*, *28*(4), 326–332. https://doi.org/10.1097/WAD.0000000000000019

Yip, S. W., Scheinost, D., Potenza, M. N., & Carroll, K. M. (2019). Connectome-based prediction of cocaine abstinence. *American Journal of Psychiatry*, *176*(2), 156–164. https://doi.org/10.1176/appi.ajp.2018.17101147

Ylikoski, R., Erkinjuntti, T., Sulkava, R., Juva, K., Tilvis, R., & Valvanne, J. (1992). Correction for age, education and other demographic variables in the use of the Mini Mental State Examination in Finland. *Acta Neurologica Scandinavica*, *85*(6), 391–396. https://doi.org/10.1111/j.1600-0404.1992.tb06034.x

Yoo, K., Rosenberg, M. D., Hsu, W. T., Zhang, S., Li, C. S. R., Scheinost, D., Constable, R. T., & Chun, M. M. (2018). Connectome-based predictive modeling of attention: Comparing different functional connectivity features and prediction methods across datasets. *NeuroImage*, *167*, 11–22. https://doi.org/10.1016/j.neuroimage.2017.11.010

Yoo, K., Rosenberg, M. D., Noble, S., Scheinost, D., Constable, R. T., & Chun, M. M. (2019). Multivariate approaches improve the reliability and validity of functional connectivity and prediction of individual behaviors. *NeuroImage*, *197*, 212–223. https://doi.org/10.1016/j.neuroimage.2019.04.060

Zahodne, L. B., Manly, J. J., Brickman, A. M., Narkhede, A., Griffith, E. Y., Guzman, V. A., Schupf, N., & Stern, Y. (2015). Is residual memory variance a valid method for quantifying cognitive reserve? A longitudinal application. *Neuropsychologia*, *77*, 260–266. https://doi.org/10.1016/j.neuropsychologia.2015.09.009

Zahodne, L. B., Manly, J. J., Brickman, A. M., Siedlecki, K. L., Decarli, C., & Stern, Y. (2013). Quantifying Cognitive Reserve in Older Adults by Decomposing Episodic Memory Variance: Replication and Extension. *Journal of the International Neuropsychological Society*, *19*, 1–9. https://doi.org/10.1017/S1355617713000738

Zahodne, L. B., Mayeda, E. R., Hohman, T. J., Fletcher, E., Racine, A. M., Gavett, B., Manly, J. J., Schupf, N., Mayeux, R., Brickman, A. M., & Mungas, D. (2019). The role of education in a vascular pathway to episodic memory: brain maintenance or cognitive reserve? *Neurobiology of Aging*, *84*, 109–118. https://doi.org/10.1016/j.neurobiolaging.2019.08.009

Zahodne, L. B., Schofield, P. W., Farrell, M. T., Stern, Y., & Manly, J. J. (2014). Bilingualism does not alter cognitive decline or dementia risk among spanish-speaking immigrants. *Neuropsychology*, *28*(2), 238–246. https://doi.org/10.1037/neu0000014

Zahodne, L. B., Stern, Y., & Manly, J. J. (2015). Differing effects of education on cognitive decline in diverse elders with low versus high educational attainment. *Neuropsychology*, *29*(4), 649–657. https://doi.org/10.1037/neu0000141

Zhao, W., Palmer, C. E., Thompson, W. K., Chaarani, B., Garavan, H. P., Casey, B. J., Jernigan, T. L., Dale, A. M., & Fan, C. C. (2021). Individual Differences in Cognitive Performance Are Better Predicted by Global Rather Than Localized BOLD Activity Patterns Across the Cortex. *Cerebral Cortex*, *31*(3), 1478–1488. https://doi.org/10.1093/cercor/bhaa290

Zhou, Z., Wang, P., & Fang, Y. (2018). Social Engagement and Its Change are Associated with Dementia Risk among Chinese Older Adults: A Longitudinal Study. *Scientific Reports*, *8*(1), 1–7. https://doi.org/10.1038/s41598-017-17879-w

Zhu, J., Li, Y., Fang, Q., Shen, Y., Qian, Y., Cai, H., & Yu, Y. (2021). Dynamic functional connectome predicts individual working memory performance across diagnostic categories. *NeuroImage: Clinical*, *30*, 102593. https://doi.org/10.1016/j.nicl.2021.102593

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*. https://doi.org/10.1111/j.1467-9868.2005.00503.x

7 Appendices

7.1 Appendix I: Supplemental Information for Chapter 2

7.1.1 Supplemental Methods

Table S1 List of open-access repositories used with information on exclusion during image processing, age and sex of dataset, and scanner information.

Dataset *(Country - site)*	**Eligible n (excluded)**	**Reason for exclusion**	**Final n**	**Mean age of final n** (SD, range)	**Male/ Female**	**Scanner** (field strength)	**Voxel dimensions** (mm)
Autism Brain Imaging Data Exchange (ABIDE)	102 (39)	Failed auto reorientation	63	28.70 (11.32, 18 - 64)	46/17		
USA – Barrow Neurological Institute						Philips Ingenia (3T)	1.11 x 1.11 x 1.2
USA – Indiana University						Siemens TriTim (3T)	0.7 x 0.7 x 0.7
France – Institut Pasteur and Robert Debré Hospital						Phillips Achieva (1.5T)	1.00 x 1.00 x 1.00
Belgium – Katholieke Universiteit Leuven						Phillips Achieva (3T)	1.20 x 1.20 x 1.20
SA – New York University Langone Medical Center						Siemens Allegra (3T)	1.30 x 1.00 x 1.30
Ireland – Trinity College Dublin						Philips Intera Achieva (3T)	0.89 x 0.89 x 0.89
The Neuro Bureau – Berlin: Mind & Brain *(Germany)*	49 (22)	Failed auto reorientation	27	32.62 (6.18, 22.24 – 49.37)	15/12	Siemens TrioTim (3T)	1 x 1 x 1
Beijing Normal University *(China)*	179 (10)	Failed auto reorientation	169	21.22 (1.88, 18 – 28)	66/103	Siemens TrioTim (3T)	1.33 x 1.0 x 1.0

Cleveland Clinic Foundation (CCF) *(USA – Cleveland Clinic Hospital)*	31 (31)	Failed auto reorientation (10); Failed QC (21)	0	n/a	n/a	Siemens Trio Tim (3T)	2 x 1 x 1.2
Center for Biomedical Research Excellence (COBRE) *(USA – The Mind Research Network)*	72 (13)	Failed auto reorientation	59	34.98 (11.61, 18 – 62)	38/21	Siemens Trio Tim (3T)	1 x 1 x 1
Dallas Lifespan Brain Study (DLBS) *(USA – University of Texas at Dallas)*	315 (138)	Failed auto reorientation	177	46.82 (18.39, 20.57 – 88.36)	45/132	Philips (3T)	
Information eXtraction from Images (IXI) ***UK – Hammersmith Hospital*** ***UK –Guy's Hospital*** ***UK – Institute of Psychiatry)***	565 (325)	Failed auto reorientation	240	47.20 (16.17, 19.98 – 80.17)	62/178	 Philips Medical Systems Intera (3T) Philips Gyroscan Intera (1.5T) General Electric Signa (1.5T)	0.9375 x 0.9375 x 1.2
Nathan Kline Institute - Rockland Sample (NKI) *(USA – Nathan Kline Institute)*	143 (28)	Failed auto reorientation (27); Failed QC (1)	115	42.83 (17.95, 18 - 83)	40/75	Siemens TrioTim (3T)	1 x 1 x 1
Southwest University Adult Lifespan Dataset (SALD) (*China – Southwest University)*	494 (46)	Failed auto reorientation (26); Failed QC (20)	448	44.73 (17.44, 19 - 80)	162/286	Siemens Trio Tim (3T)	1 x 1 x 1

Power et al (2014) *(USA – Washington University in St. Louis)*	83 (22)	Failed auto reorientation	61	24.56 (2.29, 19.69 – 37.73)	30/31	Siemens Trio Tim (3T)	1 x 1 x 1
Training set total	2033 (674)	Failed auto reorientation (632); Failed QC (42)	1359	40.04 (17.78, 18 – 88.36)	504/855	N/A	N/A

Note: Eligible n = Healthy controls >= 18 years old with age and sex data available. All datasets, except for the IXI dataset were downloaded from the 1000 Functional Connectomes Project via the NITRC repository http://fcon_1000.projects.nitrc.org/. The IXI dataset was downloaded from http://brain-development.org/ixi-dataset/.

Choice of voxel size

Larger voxel sizes were required to reduce the computational expense of the machine learning analysis as the original images, with 1 mm^3 voxels, were 27.7 MB per image. Preprocessed training set images were resized to images with $2mm^3$, $3mm^3$, and $5mm^3$ voxels. The accuracy of the different voxel sizes was then compared using an Elastic Net with 10-fold nested cross-validation in order to predict chronological age in the training set. The lowest MAE and highest Pearson's r was found for $2mm^3$ voxels (see Table S2). As such, this voxel size was selected for the full analysis. While it is possible that accuracy could have been further increased with a smaller voxel size again, using voxels smaller than $2mm^3$ becomes very computationally intensive.

Table S2 Comparison of model accuracy using full training set with different voxel sizes.

Voxel Size	File size as % of original image	MAE	r
2 mm^3	12.7%	5.99	0.908
3 mm^3	3.82%	6.44	0.895
5 mm^3	0.87%	6.55	0.889

Choice of GM threshold

The GM threshold of > 0.2 threshold has been described as an optimal threshold (see Fig. 10, Ridgway et al., 2009), and has been used in several publications (Agroskin, Klackl, & Jonas, 2014; Almairac, Duffau, & Herbet, 2018; Daniels, Gaebler, Lamke, & Walter, 2015; Hanssen et al., 2018; Seubert, Freiherr, Frasnelli, Hummel, & Lundstrom, 2013; Sowman et al., 2017; Y. Wang et al., 2016). However, a lower threshold of > 0.05 was also assessed to compare model accuracy and computational efficiency (see Table S3). The lower threshold resulted in more accurate predictions but was significantly less computationally efficient, taking 8.79 times longer to run one model (Elastic Net with 10-fold nested cross-validation). As such, for computational and theoretical reasons (i.e., less probability of including non GM information), the standard threshold of > 0.2 was selected.

Table S3 Comparison of model accuracy using different GM thresholds.

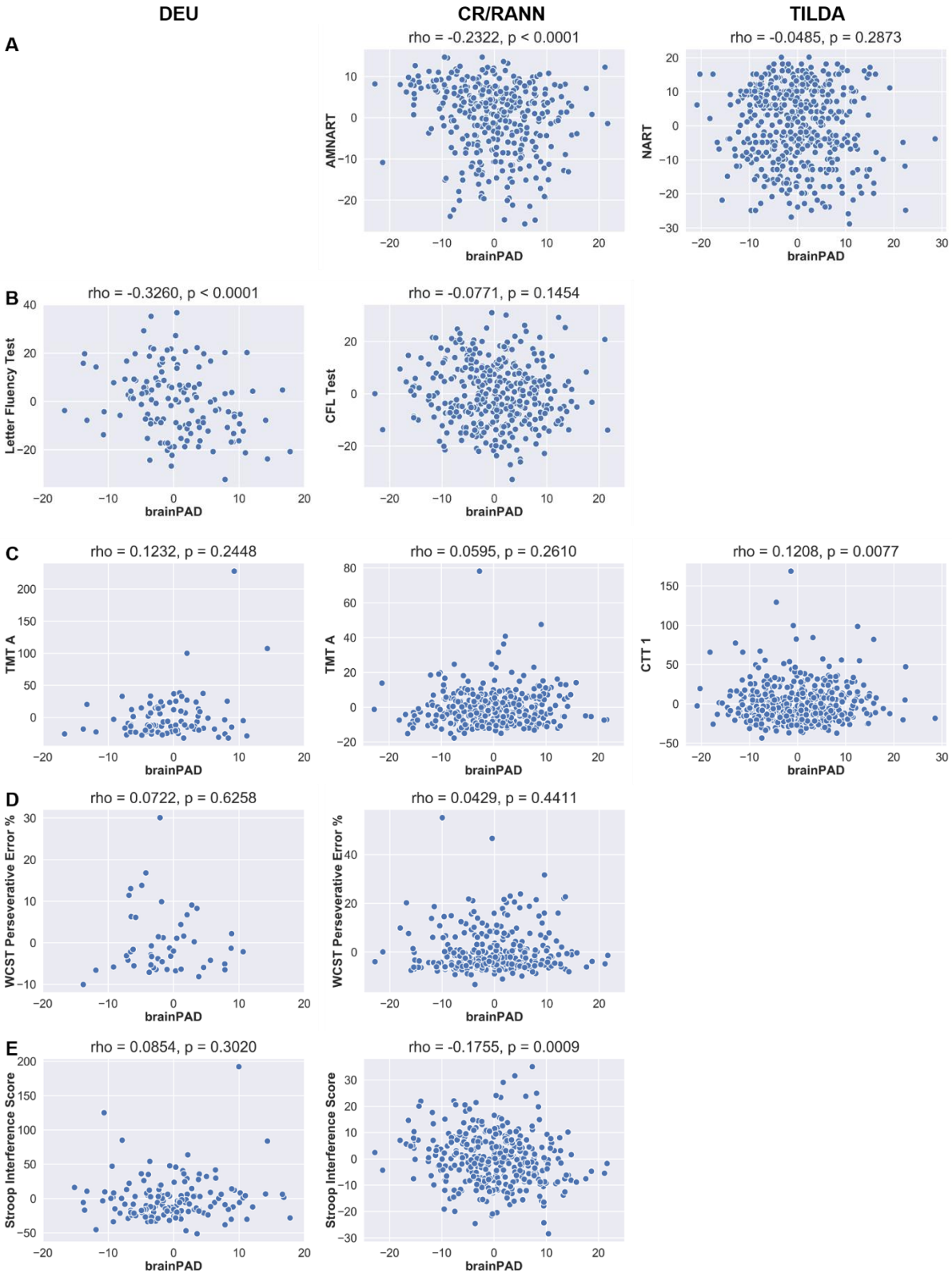
Density Threshold	Number of voxels	Runtime (hours)	MAE	r
>0.2	54,869	3.13	5.99	0.908
>0.05	148,762	27.5	5.81	0.915

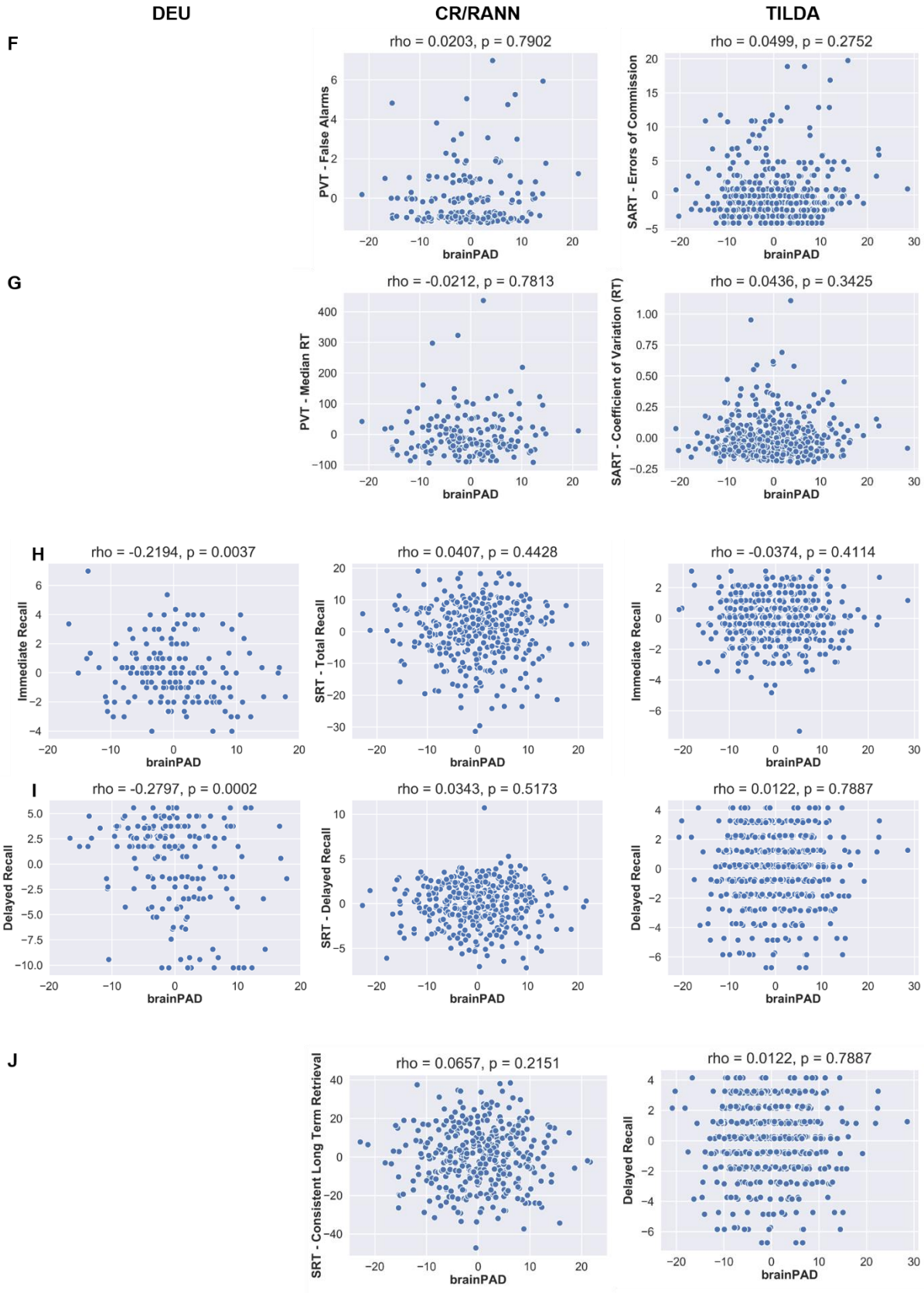
7.1.2 Supplementary Results

Table S4 Range, standard deviation (SD), and interquartile range (IQR) of variables used in partial correlations between brainPAD and cognitive functions.

Cognitive Domain	DEU			CR/RANN			TILDA		
	Range	SD	IQR	Range	SD	IQR	Range	SD	IQR
Age	45.95	8.59	13.08	61.00	17.09	30.25	38.00	7.21	8.00
BrainPAD	37.79	6.44	7.27	49.26	8.57	11.24	52.18	7.52	10.15
General Cognitive Status	16.00	2.98	4.00	16.00	2.85	4.00	9.00	1.41	2.00
Verbal Intelligence		n/a		48.00	9.31	14.00	49.00	11.22	17.00
Phonemic Verbal Fluency	70.00	13.86	17.00	64.00	11.97	16.00			
Semantic Verbal Fluency	34.00	5.79	8.00	45.00	5.52	6.00	36.00	5.43	8.00
Processing Speed	267.00	34.79	33.00	93.45	10.93	12.00	211.88	24.28	27.35
Executive Function	342.00	65.77	69.00	279.18	40.74	38.50	309.13	41.63	42.23
Executive Function (without Processing Speed)	290.00	54.73	62.00	271.47	37.21	28.78	223.77	27.95	29.72
Cognitive Flexibility	40.12	7.68	7.99	66.67	8.69	8.18		n/a	
Response Inhibition, Selective Attention	241.00	32.16	27.50	68.00	11.31	15.00		n/a	
Sustained Attention (Errors of Commission)		n/a		8.00	1.50	2.00	23.00	3.71	4.00
Sustained Attention (Reaction Time)		n/a		572.00	75.97	74.50	1.30	0.16	0.15
Verbal Episodic Memory (Immediate)	11.00	1.97	3.00	47.00	9.79	14.00	10.00	1.48	2.00
Verbal Episodic Memory (Delayed)	15.00	4.51	7.00	19.00	2.48	4.00	10.00	2.46	3.00
Verbal Episodic Memory (Learning)	67.00	29.33	45.00	71.00	17.54	26.25		n/a	
Working Memory	5.00†	1.21	1.00	16.00	3.13	4.00		n/a	
	7.00‡	1.17	2.00		n/a			n/a	
Visuospatial Ability	19.00	4.45	6.00	60.00	13.22	20.25		n/a	

Note: † Digit Span Forwards; ‡ Digit Span Backwards


DEU
CR/RANN
TILDA
A
rho = -0.2322, p < 0.0001
AMNART
brainPAD
rho = -0.0485, p = 0.2873
NART
brainPAD
B
rho = -0.3260, p < 0.0001
Letter Fluency Test
brainPAD
rho = -0.0771, p = 0.1454
CFL Test
brainPAD
C
rho = 0.1232, p = 0.2448
TMT A
brainPAD
rho = 0.0595, p = 0.2610
TMT A
brainPAD
rho = 0.1208, p = 0.0077
CTT 1
brainPAD
D
rho = 0.0722, p = 0.6258
WCST Perseverative Error %
brainPAD
rho = 0.0429, p = 0.4411
WCST Perseverative Error %
brainPAD
E
rho = 0.0854, p = 0.3020
Stroop Interference Score
brainPAD
rho = -0.1755, p = 0.0009
Stroop Interference Score
brainPAD


DEU
CR/RANN
TILDA
F
rho = 0.0203, p = 0.7902
PVT - False Alarms
rho = 0.0499, p = 0.2752
SART - Errors of Commission
G
rho = -0.0212, p = 0.7813
PVT - Median RT
rho = 0.0436, p = 0.3425
SART - Coefficient of Variation (RT)
H
rho = -0.2194, p = 0.0037
Immediate Recall
rho = 0.0407, p = 0.4428
SRT - Total Recall
rho = -0.0374, p = 0.4114
Immediate Recall
I
rho = -0.2797, p = 0.0002
Delayed Recall
rho = 0.0343, p = 0.5173
SRT - Delayed Recall
rho = 0.0122, p = 0.7887
Delayed Recall
J
rho = 0.0657, p = 0.2151
SRT - Consistent Long Term Retrieval
rho = 0.0122, p = 0.7887
Delayed Recall
brainPAD

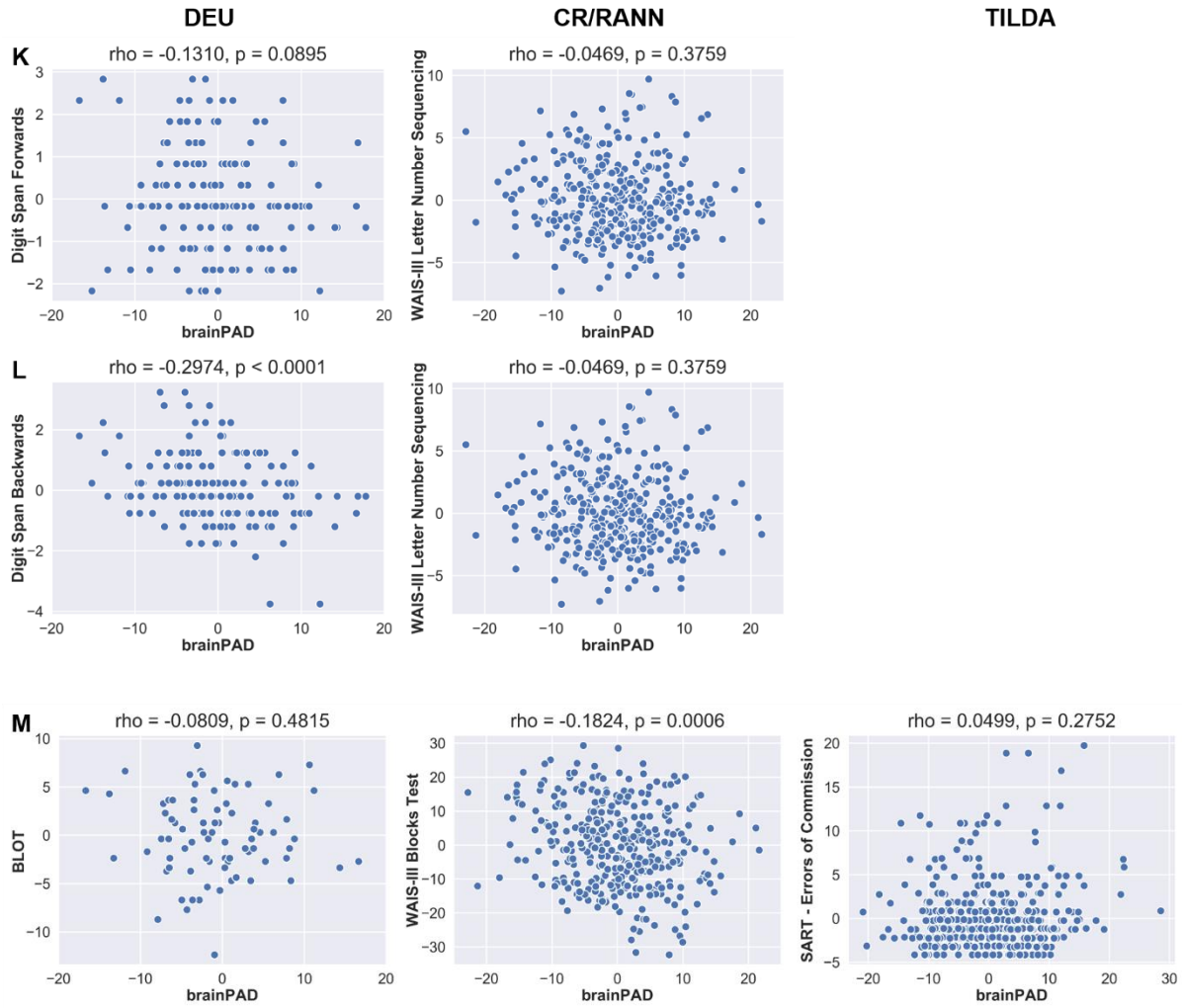


Figure S1. Scatterplots of non-replicated correlations between the residuals of brainPAD and cognitive measures after regressing brainPAD on age and sex, and regressing each cognitive measure on age and sex. *A: Verbal Intelligence; B: Phonemic Verbal Fluency; C: Processing Speed; D: Cognitive Flexibility; E: Response Inhibition and Selective Attention; F: Sustained Attention (Errors of Commission); G: Sustained Attention (Reaction Time); H: Verbal Episodic Memory (Immediate); I: Verbal Episodic Memory (Delayed); J: Verbal Episodic Memory (Learning); K: Working Memory; L: Working Memory; M: Visuospatial Ability.*

Differences in Education across samples

A Kruskal-Wallis H test was conducted to compare years of education across each of the three test datasets. There was a statistically significant difference in years of education between each group, $\chi^2(2) = 239.279$, $p < 0.001$, with a mean rank of 216.48 for DEU, 462.80 for CR/RANN and 593.12 for TILDA. Post-hoc tests using Dunn-Bonferroni pairwise comparisons revealed that TILDA had significantly higher years of education as compared to both CR/RANN and DEU, and that CR/RANN had significantly higher years of education as compared to DEU.

7.2 Appendix II: Supplemental Information for Chapter 3

7.2.1 Supplemental Results

Table S5 Negative moderation effects of cognitive reserve proxies within TILDA.

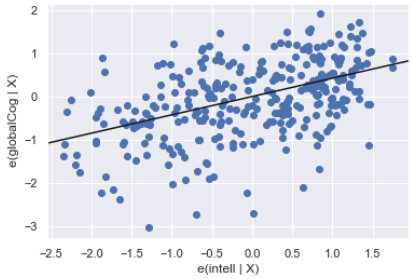
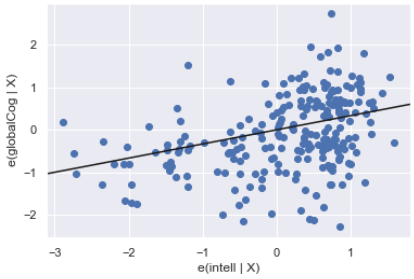
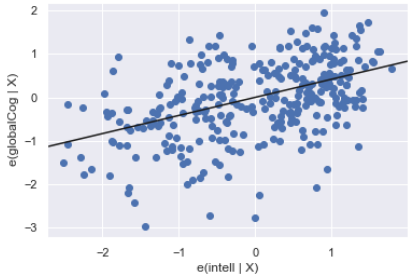
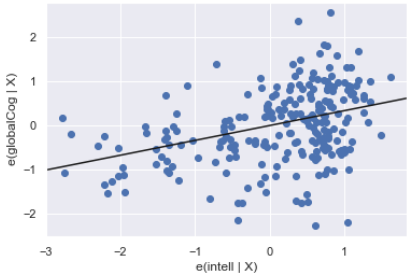
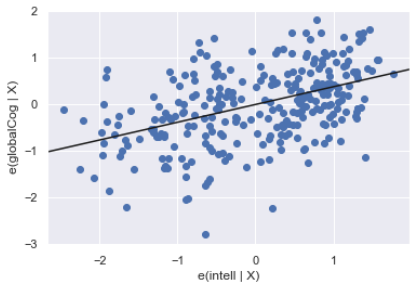
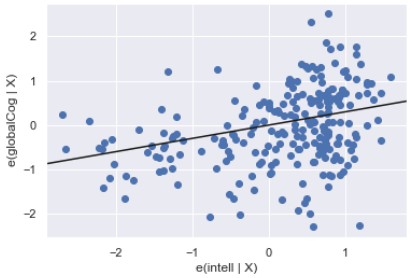
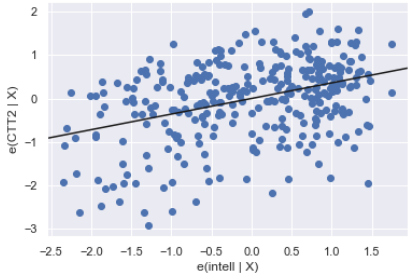
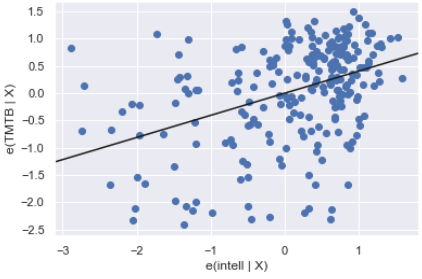
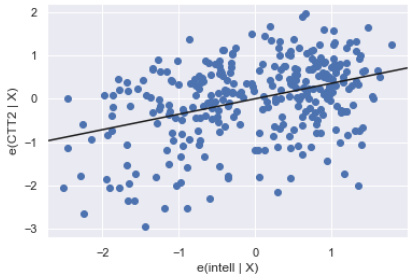
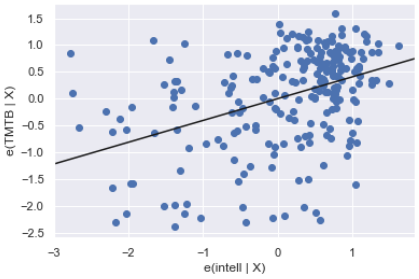
Brain Structure	Cognition	Cognitive Reserve Proxy	n	ΔR^2	β
Cx Thickness	Epi Mem	Occu + CogStim + Ex	279	.034	-.340**
Cx Thickness	Epi Mem	Occu + CogStim + Leisure + Ex	279	.032	-.335**
Cx Thickness	Epi Mem	Occu + Leisure + Ex	279	.030	-.310**
Cx Thickness	Epi Mem	CogStim + Ex	279	.029	-.286**
Cx Thickness	Epi Mem	Leisure + Ex	279	.026	-.245**
Cx Thickness	Epi Mem	CogStim + Leisure + Ex	279	.025	-.258**
Cx Thickness	Epi Mem	Occu + Ex	279	.022	-.242**
Cx Thickness	Epi Mem	Edu + Occu + CogStim + Ex	279	.021	-.270**
Cx Thickness	Epi Mem	Occu + CogStim	279	.021	-.214*
Cx Thickness	Epi Mem	Edu + Occu + CogStim + Leisure + Ex	279	.020	-.265*
Cx Thickness	Epi Mem	Edu + Occu + Leisure + Ex	279	.019	-.255*
Cx Thickness	Epi Mem	Ex	279	.018	-.200*
Cx Thickness	Epi Mem	Occu + CogStim + Leisure	279	.018	-.204*
Cx Thickness	Epi Mem	Occu + Social + CogStim + Leisure + Ex	279	.015	-.233*
GM Volume	Epi Mem	CogStim + Ex	313	.015	-.176*
Cx Thickness	Epi Mem	Edu + Occu + Ex	279	.015	-.205*
Cx Thickness	Glob Cog	Edu + Leisure + Ex	277	.015	-.198*
Cx Thickness	Glob Cog	CogStim + Leisure + Ex	277	.014	-.185*
Cx Thickness	Glob Cog	Leisure + Ex	277	.014	-.175*
Cx Thickness	Epi Mem	Occu + Social + CogStim + Ex	279	.014	-.227*
Cx Thickness	Epi Mem	Occu + Verbal IQ + CogStim + Leisure + Ex	279	.013	-.223*
Cx Thickness	Glob Cog	Edu	277	.013	-.113*
Cx Thickness	Glob Cog	Edu + CogStim + Leisure + Ex	277	.013	-.187*
Cx Thickness	Epi Mem	Occu + Verbal IQ + CogStim + Ex	279	.012	-.212*
Cx Thickness	Glob Cog	Edu + Occu + CogStim + Leisure + Ex	277	.012	-.195*
GM Volume	Exec Func	Edu + CogStim + Ex	311	.011	-.179*
Cx Thickness	Glob Cog	Edu + Leisure	277	.011	-.131*
Cx Thickness	Glob Cog	Leisure	277	.011	-.103*
Cx Thickness	Glob Cog	Edu + Occu + Leisure + Ex	277	.011	-.189*
Cx Thickness	Glob Cog	Occu + CogStim + Leisure + Ex	277	.011	-.191*
GM Volume	Glob Cog	Edu + CogStim + Ex	311	.010	-.161*

*Note: ΔR^2 = moderation effect size; β = standardised regression coefficient for moderation effect; * = p < .05, ** = p < .01 Cx Thickness = Mean Cortical Thickness, GM Volume = Grey Matter Volume, Epi Mem = Episodic Memory, Glob Cog = Global Cognition, Exec Func = Executive Function, Occu = Occupational Complexity, CogStim = Cognitively Stimulating Activities, Ex = Exercise, Leisure = Leisure Activities, Edu = Educational Attainment, Social = Social Engagement, Verbal IQ = Verbal Intelligence.*

Table S6 Positive moderation effects of cognitive reserve proxies within both datasets.

Dataset	Brain Structure	Cognition	Cognitive Reserve Proxy	n	ΔR^2	β
TILDA	HC Volume	Verb Flu	Occu + Ex	313	.019	.227*
	GM Volume	Proc Speed	Occu	313	.018	.148*
	GM Volume	Verb Flu	Occu + Ex	313	.017	.214*
	Cx Thickness	Verb Flu	Occu + Social + Ex	279	.015	.235*
	HC Volume	Verb Flu	Occu + Social + Ex	313	.015	.243*
	HC Volume	Verb Flu	Ex	313	.014	.192*
	GM Vol	Verb Flu	Occu + Social + Ex	313	.014	.231*
	HC Volume	Verb Flu	Social + Ex	313	.014	.208*
CR/RANN	HC Volume	Glob Cog	Occu + Verbal IQ	234	.030	.215**
	HC Volume	Epi Mem	Occu + Ex	234	.026	.232*
	HC Volume	Glob Cog	Occu + Verbal IQ + Ex	234	.026	.266**
	HC Volume	Glob Cog	Occu	234	.025	.155*
	HC Volume	Epi Mem	Occu + Verbal IQ + Ex	234	.025	.263*
	HC Volume	Epi Mem	Occu	234	.024	.151*
	HC Volume	Exec Func	Occu + Verbal IQ	234	.018	.168*

*Note: ΔR^2 = moderation effect size; β = standardised regression coefficient for moderation effect; p < .05, ** = p < .01. HC Volume = Hippocampal Volume, GM Volume = Grey Matter Volume, Cx Thickness = Mean Cortical Thickness, Verb Flu = Verbal Fluency, Proc Speed = Processing Speed, Glob Cog = Global Cognition, Epi Mem = Episodic Memory, Exec Func = Executive Function, Occu = Occupational Complexity, Ex = Exercise, Social = Social Engagement, Verbal IQ = Verbal Intelligence.*

	TILDA	CR/RANN
Glob Cog ~ Intell | Hipp Vol, Age, Sex Mean ΔR^2 = .168		
Glob Cog ~ Intell | GM Vol, Age, Sex Mean ΔR^2 = .159		
Glob Cog ~ Intell | Cx Th, Age, Sex Mean ΔR^2 = .147		
Exec Func ~ Intell | Hipp Vol, Age, Sex Mean ΔR^2 = .146		
Exec Func ~ Intell | GM Vol, Age, Sex Mean ΔR^2 = .142		

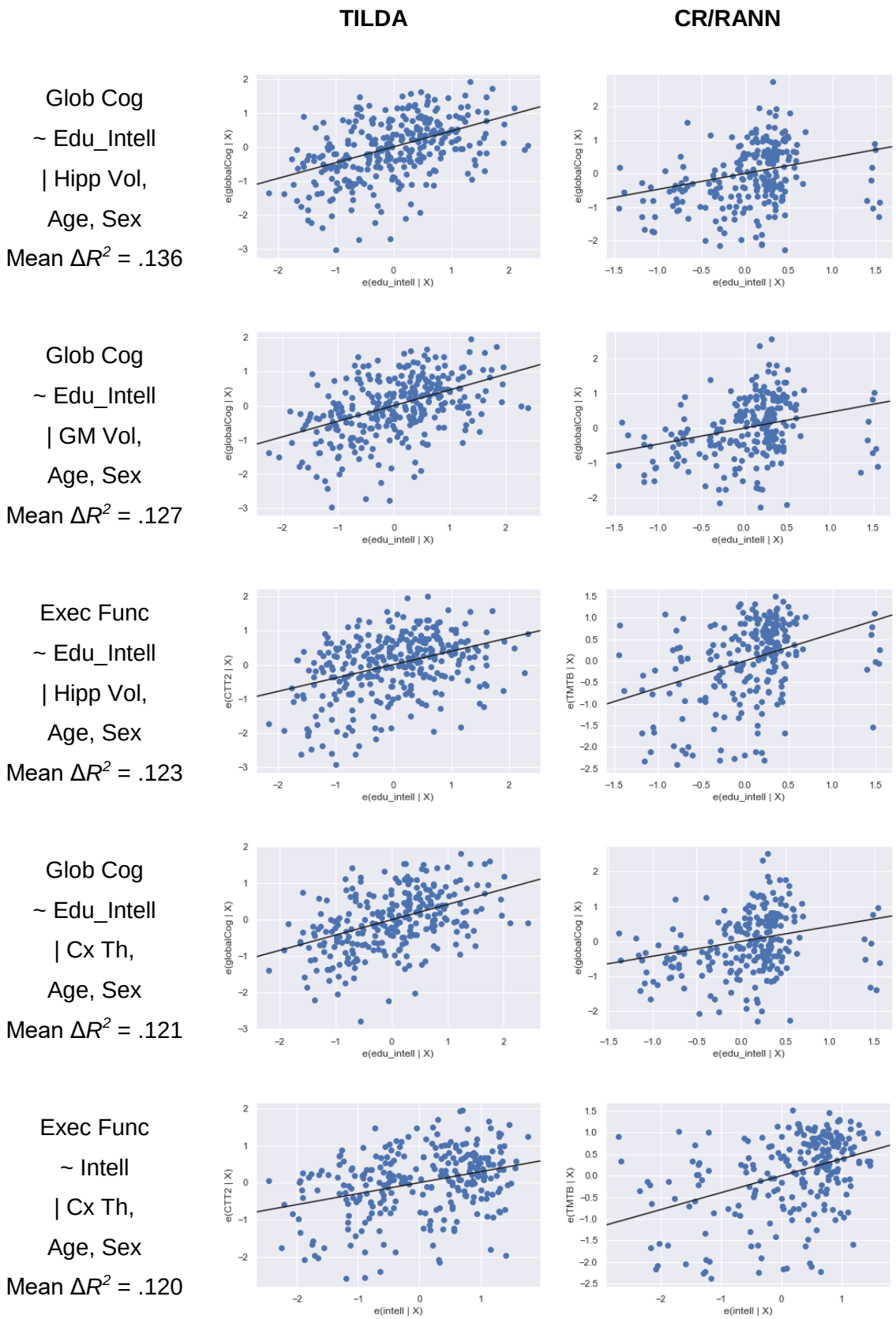


Figure S2. Association between proxies and cognition, adjusting for brain structure, age, and sex. *Plots are shown for the 10 largest mean R^2 change (ΔR^2) across datasets for proxies with significant effects. Glob Cog = Global Cognition, Exec Func = Executive Function, Intell = Verbal Intelligence, Edu_Intell = Composite of Education and Verbal*

Intelligence, Hipp Vol = Hippocampal Volume, GM Vol = Grey Matter Volume, Cx Th = Mean Cortical Thickness.

7.3 Appendix III: Supplemental Information for Chapter 4

7.3.1 Supplemental Methods

Table S7 Optimisation of feature selection threshold in the training set.

p-value for Feature Selection	Positive Network Strength		Negative Network Strength		Combined Network Strength	
	r	p	r	p	r	p
0.0001	0.1646	0.0145	0.2210	0.0010	0.2436	0.0003
0.0002	0.0560	0.4083	0.2022	0.0026	0.1895	0.0048
0.0003	0.0694	0.3058	0.1089	0.1072	0.1235	0.0675
0.0004	0.0335	0.6207	0.0915	0.1764	0.0846	0.2113
0.0005	0.0934	0.1676	0.1659	0.0137	0.1571	0.0197
0.0006	0.1304	0.0535	0.2116	0.0016	0.2022	0.0026
0.0007	0.1846	0.0060	0.2502	0.0002	0.2469	0.0002
0.0008	0.2324	0.0005	0.2687	0.0001	0.2783	0.0000
0.0009	0.2385	0.0004	0.2816	0.0000	0.2896	0.0000
0.0010	0.2152	0.0013	0.2849	0.0000	0.2818	0.0000
0.0011	0.1549	0.0215	0.2765	0.0000	0.2550	0.0001
0.0012	0.1337	0.0476	0.2744	0.0000	0.2464	0.0002
0.0013	0.0992	0.1423	0.2752	0.0000	0.2342	0.0005
0.0014	0.0728	0.2823	0.2751	0.0000	0.2238	0.0008
0.0015	0.0551	0.4161	0.2640	0.0001	0.2096	0.0018
0.0016	0.0445	0.5113	0.2577	0.0001	0.1990	0.0030
0.0017	0.0330	0.6265	0.2469	0.0002	0.1871	0.0054
0.0018	0.0344	0.6116	0.2504	0.0002	0.1879	0.0052
0.0019	0.0484	0.4751	0.2436	0.0003	0.1872	0.0053
0.0020	0.0606	0.3713	0.2559	0.0001	0.2006	0.0028
0.0021	0.0767	0.2575	0.2644	0.0001	0.2111	0.0016
0.0022	0.0931	0.1686	0.2694	0.0001	0.2196	0.0010
0.0023	0.1071	0.1130	0.2705	0.0000	0.2235	0.0008
0.0024	0.1098	0.1042	0.2716	0.0000	0.2235	0.0008
0.0025	0.1045	0.1223	0.2731	0.0000	0.2212	0.0010
0.0026	0.1045	0.1221	0.2768	0.0000	0.2224	0.0009
0.0027	0.1005	0.1372	0.2756	0.0000	0.2194	0.0011
0.0028	0.0967	0.1531	0.2707	0.0000	0.2143	0.0014
0.0029	0.0988	0.1442	0.2729	0.0000	0.2139	0.0014
0.0030	0.0996	0.1408	0.2693	0.0001	0.2112	0.0016
0.0031	0.0869	0.1993	0.2666	0.0001	0.2046	0.0023
0.0032	0.0881	0.1931	0.2652	0.0001	0.2035	0.0024
0.0033	0.0949	0.1608	0.2611	0.0001	0.2034	0.0024
0.0034	0.0868	0.1997	0.2584	0.0001	0.1975	0.0033
0.0035	0.0945	0.1625	0.2600	0.0001	0.2011	0.0027
0.0036	0.0955	0.1581	0.2608	0.0001	0.2014	0.0027
0.0037	0.0910	0.1785	0.2603	0.0001	0.1989	0.0030
0.0038	0.0791	0.2427	0.2597	0.0001	0.1933	0.0040
0.0039	0.0791	0.2429	0.2560	0.0001	0.1906	0.0046
0.0040	0.0766	0.2580	0.2562	0.0001	0.1893	0.0048
0.0041	0.0777	0.2513	0.2575	0.0001	0.1892	0.0049
0.0042	0.0741	0.2738	0.2586	0.0001	0.1882	0.0051
0.0043	0.0703	0.2991	0.2590	0.0001	0.1875	0.0053
0.0044	0.0640	0.3448	0.2575	0.0001	0.1840	0.0062
0.0045	0.0632	0.3511	0.2574	0.0001	0.1831	0.0065
0.0046	0.0698	0.3024	0.2501	0.0002	0.1819	0.0068
0.0047	0.0695	0.3047	0.2483	0.0002	0.1802	0.0074
0.0048	0.0730	0.2811	0.2454	0.0002	0.1791	0.0077
0.0049	0.0751	0.2674	0.2454	0.0002	0.1795	0.0076

0.0050	0.0761	0.2612	0.2385	0.0004	0.1758	0.0090
0.0051	0.0785	0.2461	0.2314	0.0005	0.1724	0.0104
0.0052	0.0798	0.2382	0.2219	0.0009	0.1669	0.0132
0.0053	0.0798	0.2384	0.2190	0.0011	0.1649	0.0143
0.0054	0.0812	0.2304	0.2183	0.0011	0.1654	0.0141
0.0055	0.0796	0.2398	0.2105	0.0017	0.1592	0.0182
0.0056	0.0800	0.2375	0.2070	0.0020	0.1572	0.0196
0.0057	0.0834	0.2180	0.2077	0.0020	0.1589	0.0183
0.0058	0.0871	0.1981	0.2057	0.0022	0.1594	0.0180
0.0059	0.0865	0.2011	0.2040	0.0024	0.1577	0.0193
0.0060	0.0871	0.1979	0.2014	0.0027	0.1563	0.0204
0.0061	0.0795	0.2403	0.1970	0.0033	0.1503	0.0258
0.0062	0.0852	0.2081	0.1927	0.0041	0.1504	0.0257
0.0063	0.0822	0.2244	0.1937	0.0039	0.1494	0.0267
0.0064	0.0812	0.2305	0.1923	0.0042	0.1476	0.0286
0.0065	0.0763	0.2600	0.1926	0.0041	0.1452	0.0313
0.0066	0.0728	0.2821	0.1843	0.0061	0.1390	0.0395
0.0067	0.0661	0.3289	0.1795	0.0076	0.1331	0.0486
0.0068	0.0629	0.3533	0.1806	0.0072	0.1324	0.0498
0.0069	0.0484	0.4749	0.1775	0.0083	0.1236	0.0673
0.0070	0.0412	0.5436	0.1787	0.0079	0.1210	0.0733
0.0071	0.0348	0.6075	0.1706	0.0113	0.1136	0.0927
0.0072	0.0272	0.6877	0.1667	0.0133	0.1075	0.1117
0.0073	0.0206	0.7615	0.1664	0.0135	0.1042	0.1233
0.0074	0.0166	0.8063	0.1672	0.0130	0.1027	0.1290
0.0075	0.0120	0.8597	0.1654	0.0140	0.0992	0.1426
0.0076	0.0046	0.9461	0.1682	0.0125	0.0971	0.1513
0.0077	-0.0028	0.9668	0.1658	0.0138	0.0917	0.1753
0.0078	-0.0103	0.8791	0.1662	0.0136	0.0884	0.1913
0.0079	-0.0122	0.8576	0.1633	0.0153	0.0859	0.2046
0.0080	-0.0140	0.8364	0.1628	0.0157	0.0844	0.2124
0.0081	-0.0213	0.7535	0.1608	0.0170	0.0799	0.2378
0.0082	-0.0232	0.7322	0.1575	0.0194	0.0767	0.2571
0.0083	-0.0290	0.6688	0.1593	0.0180	0.0745	0.2710
0.0084	-0.0278	0.6820	0.1612	0.0167	0.0761	0.2609
0.0085	-0.0272	0.6883	0.1599	0.0176	0.0756	0.2643
0.0086	-0.0316	0.6409	0.1593	0.0181	0.0729	0.2816
0.0087	-0.0332	0.6246	0.1611	0.0168	0.0728	0.2824
0.0088	-0.0352	0.6033	0.1594	0.0180	0.0713	0.2923
0.0089	-0.0334	0.6224	0.1615	0.0165	0.0733	0.2792
0.0090	-0.0326	0.6302	0.1636	0.0152	0.0751	0.2671
0.0091	-0.0288	0.6712	0.1657	0.0139	0.0786	0.2455
0.0092	-0.0265	0.6961	0.1673	0.0129	0.0810	0.2315
0.0093	-0.0256	0.7055	0.1675	0.0129	0.0812	0.2302
0.0094	-0.0198	0.7702	0.1672	0.0130	0.0840	0.2148
0.0095	-0.0182	0.7888	0.1675	0.0128	0.0849	0.2097
0.0096	-0.0163	0.8099	0.1661	0.0136	0.0850	0.2094
0.0097	-0.0130	0.8482	0.1632	0.0154	0.0849	0.2096
0.0098	-0.0074	0.9132	0.1634	0.0153	0.0876	0.1953
0.0099	-0.0004	0.9958	0.1654	0.0141	0.0922	0.1728
0.0100	0.0050	0.9407	0.1610	0.0168	0.0925	0.1717

7.3.2 Supplemental Results

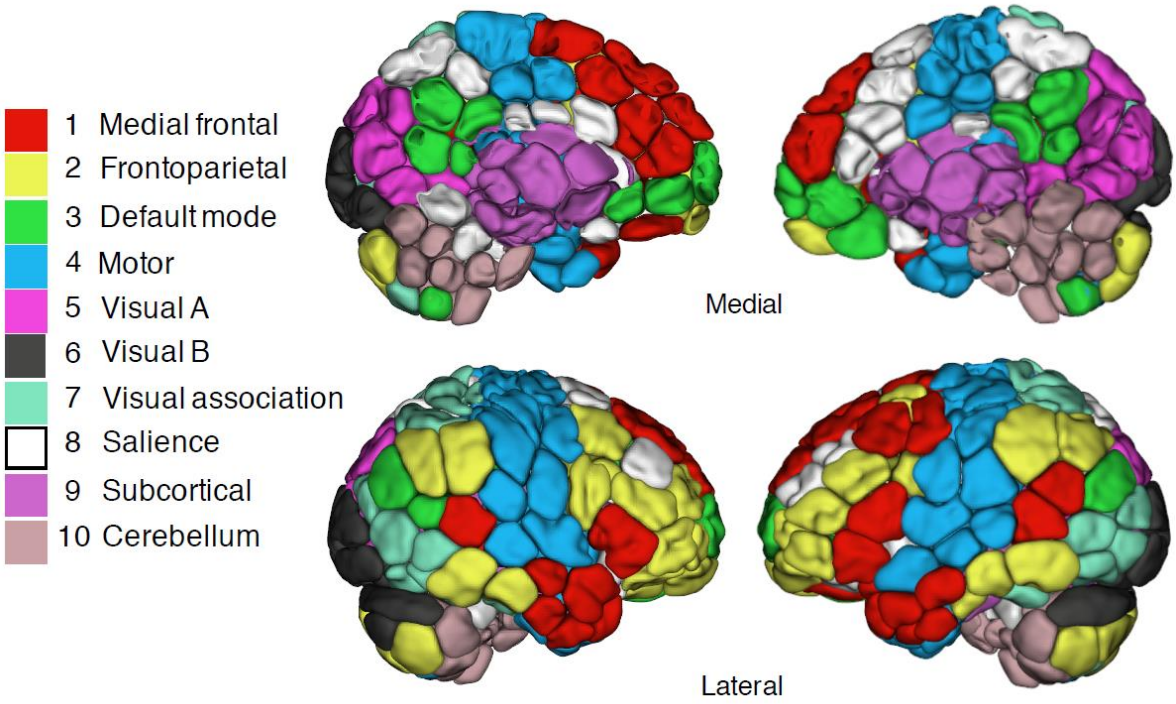


Figure S3. Ten canonical functional networks. *Figure reproduced with permission, under a Creative Commons Attribution 4.0 International License, from Supplementary Figure 6 in Greene et al. (2018).*

Table S8 Hierarchical regression results demonstrating average additional variance explained in global cognition across datasets.

Dataset	Step	R^2	ΔR^2	*Additional % variance explained*
CR/RANN	1	0.293*		
	2	0.988*	.696	69.6%
TILDA	1	0.188*		
	2	0.651*	.462	46.2%
Average across datasets			.579	57.9%

*Note: * = p < .001, Step 1 independent variables = age, sex, hippocampal volume;*
Step 2 independent variables = age, sex, hippocampal volume, CR residual

Cross-Validation Scheme	Dataset	Positive Network Strength			Negative Network Strength			Combined Network Strength		
		r	*p*	R^2	*r*	*p*	R^2	*r*	*p*	R^2
5-Fold	Training Set	.087	.281	.010	.197	.013	.041	.175	.033	.032
5-Fold	Test Set	-.021	.726	4.2e-4	-.139	.017	.019	-.130	.0253	.017
10-Fold	Training Set	.093	.249	.011	.212	.004	.046	.187	.012	.036
10-Fold	Test Set	-.079	.175	.006	-.142	.015	.020	-.164	.005	.027

Table S9 CPM performance for prediction of CR residuals in both datasets using k-fold cross-validation in the training set.

Table S10 CPM performance for prediction of CR residuals in both datasets using an age-restricted sample in the training set**.**

	Positive Network Strength				Negative Network Strength				Combined Network Strength			
	r	*p*	R^2	*MAE*	*r*	*p*	R^2	*MAE*	*r*	*p*	R^2	*MAE*
Training Set	.142	.109	.020	.563	.209	.018	.044	.527	.195	.028	.038	.536
Test Set	.079	.175	.006	.707	-.138	.018	.019	.609	-.076	.194	.006	.680

Table S11 Correlation of selected edges with CR residual in test set.

Network Strength Model	Node A	Node B	Networks (Node A - Node B)	Training Set *r*	Test Set *r*
Positive	59	14	VAs – FP	0.253**	0.105
Positive	49	19	DMN – FP	0.261**	-0.037
Positive	154	30	FP – FP	0.248**	0.022
Positive	154	49	FP – DMN	0.321***	-0.021
Positive	185	62	MF – Mot	0.251**	0.045
Positive	163	141	Mot – DMN	0.278***	-0.085
Positive	164	147	FP – FP	0.237**	0.016
Positive	166	147	FP – FP	0.241**	-0.075
Positive	164	154	FP – FP	0.237**	2e-4
Negative	185	43	FP – VAs	-0.277***	0.138*
Negative	185	45	FP – Mot	-0.315***	0.118*
Negative	166	46	Mot – Mot	-0.257**	0.020
Negative	89	49	Mot – DMN	-0.266**	0.048
Negative	161	49	Mot – DMN	-0.242**	0.007
Negative	218	49	Mot – DMN	-0.260**	0.018
Negative	182	55	FP – FP	-0.285***	0.071
Negative	166	61	Mot – Mot	-0.265**	0.034
Negative	211	164	Vis I – FP	-0.248**	0.106
Negative	185	177	FP – FP	-0.309***	0.102
Negative	185	178	FP – SAL	-0.269**	0.121*
Negative	185	179	FP – Mot	-0.281***	0.053

*Note: * < .05, ** < .0005, *** < .0001. VAs = Visual Association Network; FP = Frontoparietal Network; DMN = Default Mode Network; MF = Medial Frontal Network; Mot = Motor Network; Vis I = Visual I Network; SAL = Salience Network.*