Lifestyle, Brain Structure and Cognitive Ageing

Gesa Sophia Borgeest

MRC Cognition and Brain Sciences Unit

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Pembroke College February 2022

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. This dissertation contains less than 65,000 words including appendices, bibliography, footnotes, tables, and equations and has less than 150 figures.

Gesa Sophia Borgeest

February 2022

Acknowledgements

This PhD would not exist without the hundreds of volunteers who came to the lab, answered pages and pages of questionnaires, completed cognitive and behavioural tests, and lay still while a noisy machine scanned their brains. Thank you, dear participants, for making (this) science possible. Thank you also to all the research assistants, students, interns, technicians, administrative staff, and scholars who worked hard to collect, score, and prepare the large amounts of data I had the privilege of analysing in this thesis. Of course, it is funding bodies who ultimately make this research possible. I'd like to thank the Medical Research Council, Horizon 2020, The Wellcome Trust, and all the grant-writing scientists who convinced these institutions to fund the innovative and collaborative projects I am lucky to have benefitted from.

It is an honour to address the astonishing individuals who supported this work. Professor Rogier Kievit, a supervisor with more energy and dedication than I thought was humanly possible. Thank you for your faith, guidance, and freedom – and for being a true role model, both in how you do science and how you do life. Professor Rik Henson, I wonder if you know how deeply you impress your PhD students, and how well they speak of you years after leaving your lab. I am honoured to count myself among their ranks. Dr. Delia Fuhrmann – a former post-doc who now runs her own research group – the world is a better place because you are (doing science) in it. I am so grateful to be your friend! Thank you to my collaborators, Professor Tim Kietzmann, Dr. Christopher Madan, Theresa Fox, Dr. Maura Malpetti, Dr. Ethan Knights, Dr. Meredith Shafto, and Dr. David Samu.

My fellow (PhD) students, especially Dr. Joe Rennie, Dr. Hannes Mehrer, Dr. Alex Kaula, Dr. Marie Deserno, Dr. Ina Zieleke, Dr. Jonnie Penn, Lizzie Marx, Eric Rogers, George Britten-Neish, Isabel Thornton, Ray Siems – thank you for the ideas, friendship, the food, and the good times. To the "adults" at the CBU and at Pembroke College, Sally Butterfield, Dr. Fionnuala Murphy, Professor Duncan Astle, Dr. Becky Coombs, Tina

Emery, my mentors Dr. Sanne Cottaar and Dr. Rebecca Lämmle – thank you for asking, for listening, for your time and your advice.

To my husband Marcel, a true *Gefährte*. To my parents Gunda and Bernhard, for your lifelong *Freiheit und Geborgenheit*. To my little brother Emil. *Mein großes Vorbild*.

Part of what made these past five years so memorable were the three times I paused my PhD. First, to intern with Professor Charlotte Watts at the Department of International Development – thank you and Dr. Jasdeep Sandhu for opening my eyes to the complex importance of high-quality science policy. Second, to move to Geneva and support the Healthy Aging Team at the World Health Organization. Dr. Vânia de la Fuente-Núñez and Alana Officer, working with you was as fun as it was inspiring. The third intermission brought my life's greatest gift: my son Aaron. This thesis is written in dedication to him.

Abstract

This thesis explores the lifestyle and brain structure correlates of cognitive ageing. Using large datasets and a range of multivariate statistical approaches, it is divided into three empirical sections, which are framed by a General Introduction and Conclusion.

The first Chapter, the General Introduction, discusses this thesis within its wider context: ageing populations pose significant issues for individuals and societies, while the rise of Big Data offers promising opportunities to study the effects of brain health and lifestyles on cognitive ageing.

Chapter 2 investigates the relationship between lifestyle and cognitive abilities across the adult lifespan. It shows that, in a sample of cognitively healthy participants (Cam-CAN; N=708), higher education, better physical and mental health, more social engagement and a greater degree of intellectual engagement are each correlated with better fluid and crystallized cognitive abilities.

Chapter 3 focuses on brain structure. It explores how different aspects of morphology – such as cortical thickness, curvature, sulcal depth or surface area – change with age, and relate to cognitive outcomes, including fluid intelligence. This chapter's main finding is a cross-sectional and longitudinal double dissociation: while cortical thickness declines rapidly with age, it does not relate strongly to cognition, particularly after adjusting cognition for age. Surface area, on the other hand, has only moderate age-effects, but captures cognitive difference and change well. These findings replicated in a second, independent dataset (LCBC; N = 1236), suggesting that they are robust and generalizable across cohorts. It is plausible that this hitherto largely overlooked dissociation reflects two distinct neural features, which I discuss at the end of this section.

The final empirical section, Chapter 4, brings together the first two sections by assessing lifestyle, brain structure and cognition simultaneously. Specifically, it investigates whether brain structure mediates the relationship between lifestyle and cognitive abilities. After initial cross-sectional analyses in Cam-CAN using the lifestyle factors

created in Chapter 1, it then focuses on the mechanistically more plausible and (because of its longitudinal nature) statistically more robust relationship between cardiovascular health and cognitive performance in the LCBC data.

Finally, in the conclusion, I address the wider societal and policy context in which this PhD is, ultimately, embedded. This thesis – and indeed the science of healthy minds and brains in general – finds its urgency not least in the steadily increasing speed at which the older segment of many nations' populations is growing in size and proportion. Understanding (healthy) ageing amidst this demographic shift is of vital importance to individuals, society, and governments alike. I therefore reflect on my experiences at the World Health Organization (where an internship enabled me to contribute to the WHO's report on Ageism) and on the kind and quality of science which should (and, perhaps, should not) inform policies designed to ensure that living a longer life also means living a healthier one.

Contents

Declaration 3
Acknowledgements 5
Abstract 8
Contents 11
Scientific papers included in this thesis 17
Chapter 1: General Introduction 19
1.1 The Costs and Consequences of Population Ageing 19
1.2 The Rise of "Big Data" in Cognitive Neuroscience 22
1.3 Data analysis approaches 24
1.4 Cognitive and brain ageing: brief literature review 38
1.5 Summary and outlook on subsequent chapters 50
Chapter 2: Greater lifestyle engagement is associated with better age-adjusted cognitive abilities 52
2.1 Chapter Summary 52
2.2 Introduction 53
2.3 Methods 56
2.4 Results 69
2.5 Discussion 79
2.6 Conclusion 87
Chapter 3: A morphometric double dissociation: cortical thickness is more related to ageing; surface area is more related to cognition 89
3.1 Chapter Summary 89
3.2 Introduction 91
3.3 Methods 92
3.4 Results 97
3.5 Discussion 115
3.6 Conclusion 119
Chapter 4: Morphometry as a mediator in the lifestyle – cognition relationship 120
4.1 Chapter Summary 120
4.2 Introduction 121
4.3 Methods and Results 134
4.4 Discussion 149
4.5 Conclusion 156
Chapter 5: General Discussion 157

5.1 Chapter summary 157
5.2 Summary of findings 157
5.3 Strengths and limitations 159
5.4 Evidence-based policy 164
References and Appendices 181

Scientific papers included in this thesis

1. Borgeest, G. S., Henson, R. N., Shafto, M., Samu, D., Cam-CAN, & Kievit, R. A. (2020). Greater lifestyle engagement is associated with better age-adjusted cognitive abilities. *Plos One*.
2. Borgeest, G. S., Henson, R.N., Kietzmann, T.C., Madan, C.R., Fox, T., Malpetti, M., Fuhrmann, D., Knights, E., Carlin, J.D. & Kievit R.A. (2021). A morphometric double dissociation: cortical thickness is more related to aging; surface area is more related to cognition. *Under Review*

Chapter 1: General Introduction

At its broadest level, this thesis draws together two developments – one societal, the other scientific – which profoundly shape the way we live, but which are rarely discussed together: a rapidly ageing population, and the rise of "Big Data". The overarching question spanning the next five Chapters is: how can we harness the power of the latter to respond to the increasingly pressing concerns arising from the former?

The goal of this introductory chapter is fourfold: first, to review the individual and societal consequences of population ageing; second, to discuss the opportunities and challenges that come with large, multimodal, and shared datasets in Cognitive Neuroscience; third, to explain the statistical methodologies used in this thesis; and fourth, to provide a brief literature review summarizing relevant questions and findings in the field of cognitive and brain ageing.

1.1 The Costs and Consequences of Population Ageing

Advances in medicine and public health, rising standards of living, and improvements in education and nutrition have lengthened the human life span. Consequently, the older segment of the global adult population is increasing in size and proportion: according to estimates by the World Health Organization, by 2030, one in six people in the world will be over the age of 60, up from 1 in 11 in 2019 (World Health Organization, 2015). While this demographic shift known as population ageing started in high-income countries, it is now low- and middle-income countries that are experiencing the greatest change: by 2050, two-thirds of the world's population over 60 years will live in low- and

middle-income nations (Kämpfen et al., 2018). This has already and will continue to profoundly alter the way people across the globe work, live, vote, retire, fall sick and receive care. It will change how economies grow, nations trade, populations migrate and banks loan (Lee & Mason, 2017).

The Global Burden of Disease study predicts a steep rise in disability caused by increases in age-related chronic disease (Chang et al., 2019). For the first time, the loss of health and life worldwide will be greater from noncommunicable diseases (e.g., cardiovascular disease, dementia and Alzheimer's disease, cancer, arthritis, and diabetes) than from infectious diseases, childhood diseases, and accidents. This shift in the healthcare landscape will be expensive: if left unmitigated, the financial cost of longevity could be so high that the IMF predicts many nations will cease to be fiscally sustainable – that is, governments' abilities to collect tax revenue will decrease due to a smaller base of taxpayers while governments' expenditures, particularly on healthcare spending, will continue to increase (Honda & Miyamoto, 2020). Of the ailments affecting older people, none are more expensive than dementia: in the US, care expenses over the last five years of life were 57 percent higher for dementia patients than for patients who died of any other cause such as heart disease or cancer (Kelley et al., 2015). In 2015, the total global societal cost of dementia was estimated to be US$ 818 billion, equivalent to 1.1% of global gross domestic product (GDP; Hurd et al., 2013).

Scientists around the globe are racing to find ways to mitigate this looming (public) health crisis. They are searching for cures for dementia, ways of preventing or delaying its onset, as well as for a better understanding of risk profiles. This thesis contributes to the latter two bodies of research, asking what kind of lifestyle choices reduce the risk of cognitive decline and whether there are things people can do, as adults, to age cognitively healthier, and how we can use structural magnetic resonance imaging to describe and explain the neural mechanisms underlying the relationship between lifestyle and cognition. Preventative methods are predicted to be highly cost-effective: for instance, according to recent models in The Lancet, implementing three measures (treating hypertension, reducing smoking, providing hearing aids) would improve

health-related quality of life, reduce dementia prevalence by 8.5 percent and lower annual costs associated with dementia in England by £1·86 billion, accounting for intervention costs (Mukadam et al., 2020).

It is important, at this point, to distinguish between dementia and cognitive decline. Both refer to the age-related worsening of cognitive functions such as memory or processing speed, but dementia is a clinical diagnosis used when the decline in mental ability is deemed rapid enough to interfere with independence and daily life (World Health Organization, 2019). Dementia can arise from various aetiologies, such as Alzheimer's disease, vascular dementia, and frontotemporal dementia[1]. Cognitive decline, on the other hand, is a normal part of ageing: a person can be cognitively healthy even when showing some signs of decline. There is also *mild cognitive impairment* (MCI), which is a clinical diagnosis that refers to a transitional state between "typical" cognitive decline and dementia (Pandya et al., 2016). A meta-analysis identified the annual progression rate from MCI to dementia to be between 5 and 10 percent, with 60 to 80 percent of people with MCI not progressing to dementia even after 10 years of follow-up (Mitchell & Shiri-Feshki, 2009). However, as life expectancy increases, this number is likely to go up, further straining the healthcare system. Interestingly, there is growing evidence that exercise and cognitive interventions for people diagnosed with MCI help to slow down their rate of cognitive decline, suggesting that there is room for malleability and plasticity at this stage of neurodegeneration (Kinsella et al., 2009; Reijnders et al., 2013; Teixeira et al., 2012). This thesis focuses on healthy cognitive ageing, looking at people who show typical (i.e., non-clinical)

[1] In 2014, term dementia (which derives from the Latin root *demens*, which means 'out of one's mind') was abandoned and replaced in the DSM-5 with *major neurocognitive disorder* to acknowledge the biological and phenotypical heterogeneity of the syndrome (Sachdev et al., 2014). However, because the term dementia is still used by researchers and policy makers, I will continue to refer to it in this thesis.

cognitive decline which has not been diagnosed as MCI or dementia, asking what can be learned from observational, large-sample datasets about the lifestyle and morphometric correlates of healthy cognitive ageing. I describe this Big Data approach in the next two sections. Focusing on non-clinical populations heeds the philosophical argument that any understanding of disease should follow that of health, where health is considered not just the 'absence of disease' but the presence of something more, which Hausmann calls "functional efficiency" (Hausman, 2015). Understanding and describing the processes that allow a system to function well (e.g., non-clinical cognitive decline) provides insight into not just these functions, but, at a later stage, the reasons for which it can malfunction, e.g., in dementia (Millikan, 1989).

1.2 The Rise of "Big Data" in Cognitive Neuroscience

For most of the history of human neuroscience, scientists have qualitatively described individual or at most a small number of brains, relying, amongst other methods, on case and lesion studies (Harlow, 1848; Scoville & Milner, 1957) or post-mortem examinations (Brodman, 1909). Then, in the 1980s, magnetic resonance imaging (MRI) emerged, heralding decades of transformative research aimed at better understanding the MRI-derived neural correlates of a wide range of cognitive phenotypes in (typically relatively small) groups of people (for reviews see Cabeza & Nyberg, 2000; Smith et al., 2004). However, as part of the "reproducibility crisis" (Maxwell et al., 2015; Open Science Collaboration, 2015), it emerged that large segments of cognitive neuroscience were guilty of many of the factors that Ioannidis (2005) has argued would contribute to increased levels of spurious results in any scientific field: i) small sample sizes, ii) small effect sizes, iii) large number of tested effects, iv), flexibility in designs, definitions, outcomes, and analysis methods and v) being a "hot" scientific field (Ioannidis, 2005).

One way to improve replicability across studies is to increase sample size. This is because low statistical power not only reduces the chance of detecting a true effect, but also increases the likelihood that a statistically significant result reflects a false effect (Button et al., 2013; although see also Gelman & Carlin, 2014). In cognitive neuroscience, calls for larger sample sizes can be met only by reshaping how the field conducts its science. Because MRI studies are expensive to run (typically $700 per hour), it would be difficult for an individual research group to collect the large amounts of data needed (Poldrack, 2012). Therefore, researchers have started aggregating neuroimaging data and making them available to each other (Bzdok & Yeo, 2017; Choudhury et al., 2014; Poldrack & Gorgolewski, 2014). Sharing large amounts of complex MRI data comes with its own challenges, including ethical concerns related to subject privacy, computational storage capacity issues and the ever-changing nature of neuroimaging data (Nichols et al., 2017). However, as many of the practical, ethical, and statistical hurdles of sharing and analysing large neuroimaging datasets are being overcome, the benefits of this endeavour are undeniable. In recent years, investigators were able to address and answer research questions with a level of scientific confidence that was previously difficult to attain. For instance, a longitudinal structural MRI study with over 4400 observations from 2000 individuals refuted the influential hypothesis that higher education slows brain ageing (Nyberg et al., 2021). Similarly, a team of researchers analysing the UK Biobank data (which includes brain MRI scans from N=9722 participants) found that vascular risk factors like smoking, diabetes, and obesity were associated with greater brain atrophy, lower grey matter volume, and poorer white matter, providing strong observational evidence for a relationship between physical health and brain health (Cox et al., 2019). This shows that high-powered, well-conducted large sample studies can offer strong naturalistic support for (or against) existing hypotheses.

Large data consortia will not provide answers to all of cognitive neuroscience's questions – important insight has been and will continue to be derived from hypothesis-driven, well-controlled studies of small laboratory samples (Anderson, 2008). However,

I (and others) would argue that the more brain data become available, the more can potentially be learned about the mind and the brain – so long as one applies adequate statistical models (Bzdok & Yeo, 2017). Section 1.3 will explain some of the core statistical concepts and methodologies necessary to analyse large datasets and demonstrate how they have been employed in this thesis, while the following section summarizes the two cohorts assessed in this thesis.

1.3 Data analysis approaches

1.3.1 Structural Equation Modelling

In the following chapters, I study how cognition relates to other variables, such as lifestyle engagement and structural brain measures. Statistically, inter-individual differences (in, say, cognition) are estimated with a variance parameter, which captures the range of observed scores or performance on an outcome of interest. The associations *between* variables (e.g., age and memory) can be captured quantitatively using a *covariance* or correlation (the standardized covariance). Thus, if I am interested in the relationship between cognitive performance and morphometry, really what I study is how these two variables covary. Psychometric methods allow researchers to assess such variance-covariance structures in the form of hypothesized, simpler representations of the data (called models) which, in turn, function best in large samples (e.g. Wolf et al., 2013). In fact, with increasing N, the number of unique (co)variance terms increases (with N squared), which means that more and more data are needed to reliably estimate each term (as a rule of thumb, 10 participants per covariance term).

One set of psychometric tools employed in this thesis is *Structural Equation Modelling* (SEM). SEM offers a general framework in which hypotheses can be formulated at the construct (latent) level and explicit measurement models link the observed variables to the latent constructs (Little, 2013; MacCallum & Austin, 2000). More precisely, SEM estimates parameters that capture the strength of relationships (paths) between

variables in a model so as to minimise the difference between the observed data variance-covariance matrix and the one implied by the model. It does so for a model that is generally simpler than the original data, as estimating every path would merely replicate the observed covariance matrix. Latent variable models account for measurement error, assess reliability and validity, and often have greater generalizability and statistical power than methods based on observed variables (Jacobucci et al., 2019). In Chapter 2, for instance, I create the latent variable *Intellectual Engagement* from several questionnaire responses (observed variables). This abstracts away from individual variables, while reducing measurement error associated with simple sum scores. It also widens the interpretability of the construct (and its association to other latent variables), allowing me to refer to an activity type ("Intellectual Engagement") instead of individual activities (e.g., "doing Sudoku").

SEM also allows researchers to handle missing data in a relatively straightforward fashion. Full information maximum likelihood (FIML) has been shown to produce unbiased parameter estimates and standard errors (Enders & Bandalos, 2001). It estimates a likelihood function for each individual based on the present variables, thereby using all available data. Though the underlying mathematical principles exceed the scope of this thesis, FIMLs conceptual framework is relatively straightforward and can be explained with the analogy of broken pixels on a monitor. One can still process the image on a screen, even when some pixels are black, so long as there are enough functional pixels remaining, by using information of, for example, adjacent pixels as well as knowledge of the overall image. Similarly, FIML estimation uses what are known as case-wise log-likelihoods to fit a statistical model to incomplete data. By using only what is known from the observed data, FIML can infer what the whole model should look like without needing to know what the missing responses would truly be (Little et al., 2014). FIML generally requires missing values to be "missing at random" (MAR), a missingness category where the propensity for a data point to be missing is not related to the missing variable, but can be related to some of the present (non-missing) data (Enders, 2001).

However, violating the MAR assumption does not seriously distort parameter estimates (Collins et al., 2001).

Another advantage of SEM is that it is easy to compare different models, in order to contrast competing theories or hypotheses (Ullman & Bentler, 2012). An example of this is in Chapter 4, where I compare models to assess whether surface area and cortical thickness play dissociable roles in mediating the relationship between cardiovascular health and cognitive abilities. Constraining different parts (paths) of the models and then comparing their fit to each other allows me to investigate the strength of specific relationships between variables.

The ability to statistically compare different models (that is, mathematically described concepts and theories) is a powerful extension of null-hypothesis significance testing (NHST). NHST, by definition, assesses the probability that an effect equal or larger than the observed effect (such as the relationship between two variables) would occur by chance if the null hypothesis were true (that there is no effect). If that probability is deemed sufficiently small, the null hypothesis is rejected, and an effect is considered 'statistically significant'. NHST has been criticised on multiple fronts. Of relevance to this thesis is that rejecting the null does not provide logical or strong support for the alternative. The endeavour to build, assess and compare mathematical models does not reject the core logic NHST: whether we favour one model over another can still be based on the probabilistic inference provided by the same mathematics that guide general linear model analyses (such as the *t*-test or ANOVA). What modelling does do is shift the focal point from the null hypothesis to model content, asking: does the fit of the more complex model increase enough, compared with that of the less complex model, that the cost of additional complexity is worth it (Rodgers, 2010)? In this way, we are not comparing "nothing" (that is, the null) to "something" (the alternative hypothesis), but two plausible scientific hypotheses, theories, or models to one-another.

1.3.2 Longitudinal data

1.3.2.1 Cross-sectional versus longitudinal approaches

To study cognitive and brain ageing, researchers can (within the framework of observational data analysis) investigate cross-sectional data, longitudinal data, or a combination of both. Each approach comes with its own challenges and opportunities. Cross-sectional studies of ageing are generally (relatively) quick, easy, and cost-effective to conduct. Their main disadvantage is known as the "cohort effect": any difference between younger and older participants may not necessarily be due to age, but can be caused by the two cohorts being exposed to different experiences (Alwin, 2009). One well-documented cohort effect is the Flynn Effect (Flynn, 1996, 2007; Schaie et al., 2005), which posits that the steady increase in fluid intelligence between generations observed in the second half of the 20th century are (in part) explained by improved educational levels. The cohort effect also seems to refute what is often known as the "stability despite loss paradox"; a (largely cross-sectional) finding suggesting that older adults are generally more satisfied with their lives than younger adults, despite experiencing more "loss" such as loss of health or loss of loved ones (Kunzmann et al., 2000). Several studies have since shown that there is, in fact, an age-related decline in life satisfaction, but that this had previously been masked by cohort effects: the participants of many papers pointing to a wellbeing-paradox were born in Europe in the first half of the twentieth century. To *them*, their lives in older age were more stable and prosperous than when they were young and affected by the traumas of war (Kunzmann et al., 2000; López Ulloa et al., 2013). It has therefore been suggested that, as the experiential gap between generations narrows in Europe (which has now seen an unprecedentedly long period of peace), the frequency and magnitude of certain types of cohort effects should decline in European samples (Ganguli, 2017). Of particular relevance to this thesis is a cohort effect of brain volume: studies have documented generational drifts in body weight and brain weight (Resnick et al., 2000), making cross-sectional investigations of morphometry prone to obscuring the true rate of age-related brain changes (Sigurdsson et al., 2012).

Cross-sectional results therefore need to be interpreted with reference to the likelihood that the associations found could be due to cohort effects. Questions such as "Is the younger generation less prone to smoking but more affected by obesity?" or "Has the older generation been more exposed to certain toxins?" can help contextualize cross-sectional research findings and guide generalizability.

In longitudinal samples, ageing-related differences can be measured within the same cohort, allowing researchers to differentiate between age effects and cohort effects. However, the quality of longitudinal results can be confounded by test-retest effects and selective attrition. The former refers to the learning that takes place when participants complete the same task more than once. It can look as though cognitive abilities have increased between two measurements occasions because re-doing the same test leads to improvements in test performance, masking cognitive decline that may, in fact, have occurred. Retest effects can be partially mitigated by using different tests to study the same construct over time (Lo et al., 2012) or by employing models which account for this issue (McCormick, 2021) . Selective attrition describes the problem that the kinds of participants who return to the lab for multiple test appointments differ from those who drop out in ways relevant to the research study (Burke et al., 2019). For instance, in ageing studies, attrition was found to most likely occur in older, male participants who were more socially isolated and showed more cognitive decline than their peers (Jacobsen et al., 2021). This is problematic (particularly when not using FIML or other methods to handle missing data) because it skews the remaining, analysable sample towards more high-functioning participants, potentially biasing the interpretability and generalizability of the findings. Oversampling and statistical approaches such as propensity score modelling, FIML (see above) and inverse probability weighting can be used to adjust the original dataset via matching or weighting in accordance with non-response or attrition bias, thus allowing results to be generalised back to the original cohort (Austin & Stuart, 2015; Eerola et al., 2005; Seaman & White, 2013; Wooldridge, 2007). Despite these challenges, longitudinal data have many advantages. When

analysed appropriately (see below) they can shed light on the temporal (and, under certain conditions, causal) dynamics underlying brain and cognitive ageing.

1.3.2.2 Analysing longitudinal data

To investigate brain and cognitive ageing it is, in principle, possible to just perform separate multiple regression analyses (one per measurement occasion), and to calculate the expected rate-of-change score from the respective differences (see Sullivan et al., 1995 as an example). However, many researchers have discussed the limitations using observed rates of change as outcome variables (Finkel et al., 2003; McArdle, 2012). *Latent Change Score Models* are better suited to address hypotheses about temporal, interactive dynamics over time.

Latent Change Score Models (LCSMs) are a powerful and flexible class of Structural Equation Models. They have been used to describe a large variety of temporal effects in cognitive neuroscience, such as the transfer of cognitive training beyond item-level cognitive ability (Schmiedek et al., 2010) or that white matter changes are associated with declines in fluid intelligence in older adults (Köhncke et al., 2016; Ritchie et al., 2015). Several tutorials are available explaining LCSMs in detail (Ghisletta & McArdle, 2012; Kievit et al., 2018; Zhang et al., 2015); below I will provide a brief description of the core aspects of these models and how LCSMs can be used to study cognitive and morphometric change.

Defining differences between measurement occasions (each comprised of observed scores or latent scores at each time point) as a latent change score factor (Δ) and then adding a regression parameter (β) to the change score yields two interesting parameters, which are of interest when investigating neurocognitive ageing or development. First, one can detect whether there is a reliable average change between time-point 1 and time-point 2 (captured by the mean of the latent change factor; $\Delta\mu$ – although note that this necessitates using a covariance instead of a regression to ensure the mean change is not conditional on the autoregressive effects). Second, the variance of the change

factor ($\Delta\sigma^2$) captures the extent to which individuals differ in the change they manifest over time. Once these parameters are in place, models can be constrained in different ways to test specific hypotheses (as described in 1.3.1 above). To properly investigate brain and cognitive ageing, one would want to simultaneously model longitudinal cognitive changes, longitudinal morphometric changes, and their interrelationship over time. This is easily achievable using LCSMs by including more than one longitudinal variable in the model. Investigating this cross-domain coupling captures the extent to which change in one domain (e.g., cognition) is a function of the starting level of or changes in the other (e.g., grey-matter volume). What is so compelling about this modelling technique is that it provides a powerful analytic framework for testing a wide range of hypotheses in a principled and rigorous manner. The following section will explore the extent to which this process allows researchers to not just detect and describe observational effects but to make inferences regarding their chain of causation.

1.3.3 Causal inference

According to the philosopher Thomas Kuhn, the birth of the Scientific Method can be traced back to the first half of the nineteenth century, when two developments – separated for over two hundred years by opposing schools of thought – finally merged (Kuhn, 2011). One, defined by figures such as Galileo Galilei and Thomas Hobbes, was the axiomatic-deductive style of the geometric tradition; the classical sciences of 16th and 17th century astronomers and mathematicians. The second, initially advocated for in the 1640s by Robert Boyle and then elevated to fame by Francis Bacon's thermodynamics in the 1800s was that of experimentation (Radder, 2009). The marriage of mathematics and experiments gave rise to statistics, and in doing so provided the world with the tools to not just *describe* natural phenomena, but to understand how one thing *affects* another – in other words, to infer causality.

Until today, the best way to detect causation is through well-controlled, randomized experiments. In ageing or developmental neuroscience, such studies are, however, not only expensive, and time-consuming to run – manipulating the independent variable is

simply not always possible. It would be unethical to assign children to different levels of adversity, unlikely to change multiple aspects of people's lifestyle and impossible to increase the size of a person's pre-frontal cortex. And yet, researchers are interested in how adversity affects development, how social, intellectual, and physical engagement affect mental health, and whether brain volume explains cognition.

What framework and principles should we apply to determine the likelihood that the "associations" found in observational studies do, in fact, reflect a causal effect? What are the necessary and sufficient conditions to infer causality and, consequently, to translate scientific findings into tangible recommendations and policies?

As discussed in Sections 1.3.1. and 1.3.3., Structural Equation Modelling provides researchers with the tools to derive model-based predictions of causal hypotheses, and examine the extent to which the data (dis)confirms these hypotheses (Bollen & Diamantopoulos, 2017). This approach, however, comes with its own challenges, most fundamentally that of *model equivalence*, also known as the "underdetermination of theory by data". The problem is that even a well-fitting model does not provide conclusive evidence for the causal hypotheses posited by the model (Kievit et al., 2018). This is because any observed data pattern is compatible with many different data generating mechanisms (Raykov & Penev, 1999). In other words, whereas *alternative* models almost always lead to differences in model fit, *equivalent* models are different representations of the model structure that result in exactly the same model fit. Such models can, therefore, not be distinguished by model fit indices alone. Another issue is that modelling choices (such as how to deal with missing data or which variables are allowed to covary in the model) can affect the results, and therefore the interpretation of causal direction (Usami et al., 2016). Thus, I would argue that although SEM-based techniques offer tools to infer causality from observational data under specific and sometimes implausible assumptions, such an *approximation* is not sufficient for drawing causal conclusions. Instead, high-quality, large-sample models should be combined with and integrated into *highly plausible theoretical mechanisms*. This requires a solid understanding of such mechanisms, served ideally by collaborations

that span different (sub)fields. In other words, solid models are necessary, not sufficient, to infer causality, a convincing integration of observational and experimental/mechanistic science, however, can be.

A successful example of such integrative work is the impairment of working memory (the temporal holding of information in anticipation of further processing) in schizophrenia. First, neuroimaging studies (several of which used SEM, e.g., Schlösser et al., 2003) found that activity in the pre-frontal cortex predicted working memory abilities in patients with schizophrenia. Then, EEG recordings in humans and monkeys revealed that low gamma oscillations were associated with impaired working memory (Haenschel et al., 2009). Moreover, histological studies reported that structural deficiencies in inhibitory GABA neurons underlie this dysfunction (Lewis et al., 2012). Finally, administering GABA receptor antagonists to reduce GABA uptake in the pre-frontal cortex of rats impaired their spatial reference and short-term memory (Auger & Floresco, 2015)

Such trans-disciplinary results provide sufficient evidence that it is indeed a GABA neuron deficiency in the frontal cortex that causes working memory problems in schizophrenia patients, opening avenues for future research into both the understanding of the disease (and, arguably non-impaired working memory functions) and possible treatment options. For example, a recent paper showed that modulating frontal gamma oscillations improved working memory in schizophrenic patients (Singh et al., 2020). Similarly strong converging sources of evidence stem from research suggesting that physical activity slows down age-related cognitive decline, a topic I discuss in detail in Chapter 4. I then return to the questions of causal inference, the promises of integrative neuroscience and what suffices as strong-enough evidence to inform policy making in Chapter 5 (the General Discussion).

1.3.4 Obtaining brain structure data: pipelines and pitfalls

As mentioned above, obtaining, organizing, and sharing large-sample neuroimaging data comes with a host of challenges. One issue lies in the extraction of meaningful and reliable information from the raw brain images. In human structural neuroimaging, there are generally two ways to achieve this: voxel-based- techniques and surfaced-based techniques.

Voxel-based morphometry (VBM) analyses first match MR images in a common space (a process called *spatial normalization* to a *group template*) to establish voxel-for-voxel correspondence across participants. This process involves warping the images in a non-linear manner, whereby voxels (MR's units of brain tissue, typically 1mm x 1mm x 1mm in size for T1-weighted images) are stretched or compressed to align participants' brains into the same space. This process creates a *deformation field* which is a map of how each voxel in the input (native) image must move to land at the matching point in the template image. This deformation is applied to the native image to create an image that is in voxel-for-voxel registration with the template. The deformed image is then *segmented* into tissue classes (such as grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF)) based upon the intensity in the image (which depends on the MRI contrast, e.g., T1-weighting), as well as tissue class priors, which indicate the likelihood of finding a given tissue class at a given location in the template space.[2] Thus, VBM estimates the proportion (or density) of each type of tissue in each voxel. By modulating (inverse-scaling) these proportions by the amount of stretching or compression needed to warp to the template space, one can estimate the volume of each tissue type in the original, native image. Normally (and in this thesis), one only investigates the GM tissue class. The modulated, segmented GM image is then *spatially smoothed* (e.g., with a Gaussian filter of full-width-at-half-maximum of 8mm) to

[2] In fact, the stages of normalisation and segmentation are intertwined, and for the Cam-CAN data analysed in this thesis, they are implemented in a single, iterative stage, rather than being sequential, using "unified segmentation" (Ashburner & Friston, 2005)

increase overlap between participants (since the above spatial normalisation is not perfect) and to facilitate parametric statistics (e.g., by central limit theorem).

Although VBM has been instrumental to our understanding of brain structure, it comes with some limitations (Davatzikos, 2004). Most importantly, the estimation of local GM volume does not always relate directly to the macrostructure of the biological brain. Furthermore, imperfections in the registration of native images to the same template can influence the volume estimates, i.e., unaccounted individual differences in brain shape can be confounded with individual differences in volume of the corresponding anatomical structures (Bookstein, 2001) . Finally, while VBM is an efficient way to search a whole image (across all voxels), statistics on smoothed images ignore discrete boundaries between distinct anatomical structures. On the other hand, while some discrete brain structures (e.g., subcortical nuclei) can be delineated on T1-weighted images of individual brains (traditional volumetry, whether manual or automated), this does not apply to T1 imaging of the cortex, which has few anatomical features with which to subdivide (i.e., features that unambiguously define functionally separate sub-regions).

Furthermore, the cortex is a convoluted 2D surface that happens to be embedded within a 3D space, which is ignored in estimates of total volume of GM within each voxel. Other computational approaches use the MRI-contrast between GM and WM, and between the GM and CSF, to estimate a surface directly, from which one can extract metrics like cortical thickness and cortical surface area, taking into account the precise folding of the cortex (Lövdén et al., 2013a). Because VBM's estimate of GM volume reflects a mixture of morphological properties (e.g., thickness, surface area and degree of cortical folding), it may miss important biological (morphological) features. Indeed, as will be discussed further in Chapter 3, cortical thickness and surface area appear to have dissociable relationships with ageing and cognition. This is consistent with other evidence that cortical thickness and surface area are genetically independent (Panizzon et al., 2009, Winkler et al., 2010).

There have been recent improvements in such **surface-based analyses (SBA)**, particularly in automated extraction of the cortical surface. The surface boundary between cortical white matter and cortical grey matter known as the *white surface*; this represents the inner boundary of cortex. The boundary between the grey matter and dura and/or CSF; this is referred to as the *pial surface*. The cortex is modelled as a tessellated surface comprised of triangles. Each triangle is known as a *face*. The point where the corners of the triangles meet is called a *vertex*. The parameters of the model are the coordinates (i.e., the X, Y, and Z) at each vertex. Once the coordinates of each vertex are known, the surface can be rendered as a surface embedded in 3D. From here, one can compute a variety of morphometric measures.

There are several software packages available for both VBM and SBA analyses. The most common ones – and the ones used in this thesis – are Statistical Parametric Mapping (SPM, www.fil.ion.ucl.ac.uk/spm) for VBM and FreeSurfer (surfer.nmr.mgh.harvard.edu) for SBA. One important advantage of these tried-and-tested, widely-used pipelines is the comparability of high-quality imaging metrics across studies: two entirely different labs can extract very similar information from their brain scans so long as they both employ the same software, making data sharing much easier. This (unlike, say, manual segmentation) also makes results reproducible, as the same input data will yield the same output metrics. The downside is that the field then focuses on the few readily available measures, mainly grey-matter volume, and cortical thickness. However, not only is there much more to the brain than its volume and thickness, but it is, in fact, possible to extract a host of additional metrics from T1-scans, especially using SBA: different groups have assessed, for instance, the depth of the brain's sulci or the degree of cortical curvature. The challenge is obvious: although such metrics are likely to contain interesting and novel information regarding the (ageing) brain, as they are not part of the standard pipelines' output, they are less readily available. One pipeline recently developed to this end is *Mindboggle*, a FreeSurfer-based tool which extracts a total of 12 imaging metrics (Klein et al., 2017). As will be discussed in Chapter 3, these metrics correlate differently with age and cognitive abilities, and

explain additional variance in both phenotypes, making them useful tools to investigate the ageing brain.

1.3.5 Introduction to cohorts

This thesis analyses two large-sample, publicly available neuroscientific datasets. While both cohorts are discussed in more detail in the relevant methods sections of the empirical chapters, I will briefly outline the samples' key features here.

The Cambridge Centre for Ageing and Neuroscience (Cam-CAN; https://www.cam-can.org/) project, supported by UK and EU funding bodies[3], was launched in 2010 with the goal to better understand the epidemiological, cognitive and neuroimaging correlates of healthy ageing. In Stage 1 of data collection, 3000 participants aged 18 and over were recruited via local GP surgeries in Cambridge city, and then contacted in person in an opt-out procedure. If agreeable, they then undertook a 2-hour home interview about demographics and lifestyle (Shafto et al., 2014). A subset of 700 cognitively and psychiatrically healthy adults aged 18-87 (100 people per age decline) continued to Stage 2 to undergo cognitive testing and functional and structural neuroimaging (Taylor et al., 2017). Cognition was assessed across multiple cognitive domains including attention and executive control, language, memory, emotion, action control and learning. Structural MRI data was processed with SPM and FreeSurfer (described in section 1.3.5), yielding canonical brain structure metrics including volume, surface area and cortical thickness. Stage 3 of the Cam-CAN data collection asked 280 adults from Stage 2 to return for further cognitive and neuroimaging assessments. These participants (N=261) thus have repeated measures (there was a time interval between an average of 1.3 years (sd = 0.66 years) between the two waves on average) and represent the longitudinal subsample within this cohort.

[3] The Cam-CAN project cost approximately £6 million ($8.2 million), resources no one lab would be able to afford. As of Dec 2021, the dataset has been requested by over 1000 research groups worldwide, making this a good example of the importance and usefulness of large-sample, openly available neuroimaging samples.

The Lifespan Changes in Brain and Cognition (LCBC; https://www.oslobrains.no/) consists of a series of studies investigating age-related changes in brain and cognition in relation to risk and protective factors of cognitive decline and dementia. The dataset contains rich cognitive, multimodal neuroimaging, physical and epidemiological information from approximately 2700 adults aged 20-89. Participants were recruited using advertisements in local newspapers and webpage adverts (de Lange et al., 2016). LCBC was collected with a focus on longitudinal data, which is available (in multiple waves) for approximately 1800 adults. I analysed a subset of these data (N=1236) for whom FreeSurfer metrics and fluid cognition measures were available.

Both Cam-CAN and LCBC are part of Lifebrain (https://www.lifebrain.uio.no/), an EU Horizon 2020 funded project which seeks to bring together and harmonize data from 6000 European research participants collected in seven countries. Access to Lifebrain data can be applied for by external researchers: https://www.lifebrain.uio.no/about/access-to-data/.

1.4 Cognitive and brain ageing: brief literature review

Having discussed the methodological background of this thesis, the following sections will provide a summary of some of the psychological and neuroscientific concepts underlying the subsequent empirical chapters. This brief literature review will tackle four key questions in the field of cognitive ageing:

i. How does cognition change with age?
ii. How does brain structure change with age?
iii. What theories might explain the inter-individual differences in cognitive and brain ageing?
iv. What is the relationship between (modifiable) lifestyle activities and cognitive ageing?

1.4.1 Age-related changes in cognitive abilities

According to the two-component model of general intelligence, human cognition is comprised of fluid cognitive abilities and crystallized cognitive abilities (Baltes et al., 1980; Cattell, 1943; Horn & Cattell, 1967). Crystallized ability denotes the ability to perform cognitive tasks based on "skilled judgment habits", that is, habits that have been learned, and hence "crystallized". Examples include the acquired knowledge about the world, vocabulary, or the acquired skill to perform mental arithmetic. In contrast, fluid ability is relevant "in tests requiring adaptation to new situations, where crystallized skills are of no particular advantage" (Cattell, 1963). Typical examples of fluid abilities are knowledge-independent reasoning skills, perceptual speed, working memory, and episodic memory. Crystallized and fluid cognitive abilities have long been recognized to follow different development trajectories across the lifespan (Horn & Cattell, 1967; Wang & Kaufman, 1993). As shown in Figure 1-1, crystallized intelligence

tends to increase throughout most of a person's lifetime, often plateauing in old age. Fluid intelligence peaks relatively early (mid to late twenties; McArdle et al., 2002) and then declines steadily.

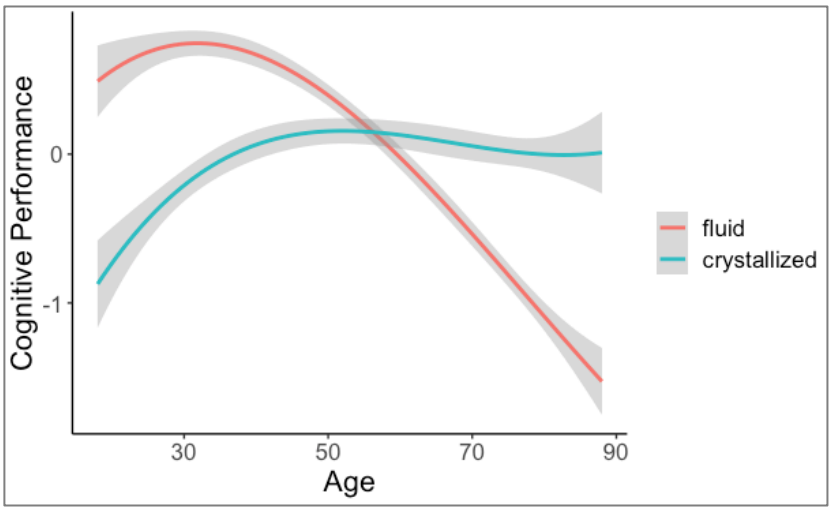


Figure 1-1: Fluid and crystallized ageing. Crystallized and fluid cognitive abilities follow different developmental trajectories across the adult lifespan. Figure shows data from cognitively healthy participants in Cam-CAN (N=708) aged 18-88, described further in Chapter 2.

Although this pattern of cognitive decline is well supported in the literature of healthy ageing – that is, by studies investigating people not diagnosed with dementia or MCI (Finkel et al., 2003; Ghisletta et al., 2012; Hedden & Gabrieli, 2004), a more recent meta-analysis suggest that fluid and crystallized abilities, in fact, decline in a correlated manner (Tucker-Drob et al., 2019). Moreover, longitudinal analyses have shown that for participants who later go on to develop dementia, however, fluid *and* crystallized abilities decline (Grober et al., 2008; Howieson et al., 2008). In one study, the onset of accelerated decline for fluid and crystallized abilities occurred approximately 10 and 5 years before diagnosis, respectively (Thorvaldsson et al., 2011). Thus, although fluid abilities begin to decline before crystallized abilities, these changes are difficult to tease apart from the "normal" changes in fluid abilities occurring in healthy adults. Conversely, the detection of crystallized decline has been proposed as a cognitive marker for the early (prodromal) stages of dementia (Thorvaldsson et al., 2011).

It is worth noting that the two-factor model described above is unlikely to fully capture human cognitive abilities, but that cognition can be usefully divided along more than two dimensions. For example, there is evidence that aspects of memory ability remain after accounting for fluid abilities, suggesting that individual differences in memory exist beyond those along fluid and crystallized dimensions (Henson et al., 2016). Nonetheless, for this thesis, the two-factor fluid-crystallised model suffices as a broad summary of individual differences in cognitive abilities.

1.4.2 Age-related changes in brain morphology

1.4.2.1 MRI-detectable brain ageing

When assessed post-mortem, the most striking difference between a younger person's and an older person's brain is that the latter is much smaller. This wide-spread age-related brain shrinkage is caused by losses in both white matter and grey matter (Sigurdsson et al., 2012). Grey-matter atrophy is most pronounced in the hippocampus, caudate nucleus, association cortex, cerebellum, and the medial temporal lobe, while less atrophy is seen in other cortical regions such as the entorhinal cortex and the primary visual cortex (Jiang et al., 2014; Raz, 2005; Salat, 2011; Salat et al., 2004). Cross-sectional studies likely underestimate the rate of change: in one large-sample study, the yearly difference with age in grey- and white-matter volume in cross-sectional data (N=4303) was approximately 40 percent less than the annual change in that study's longitudinal data (N=367; Sigurdsson et al., 2012); a pattern also found elsewhere (Raz, 2005; Raz et al., 2003). Longitudinal estimates suggest that whole brain grey matter declines 0.6 percent per year in normal (i.e., nonclinical) ageing individuals aged 55-90 (Barnes et al., 2009; Sigurdsson et al., 2012). However, the shape of this decline is still subject of debate, i.e., whether it accelerates (Driscoll et al., 2009; McDonald et al., 2009; Scahill et al., 2003; Walhovd et al., 2005), accelerates to a plateau (Schuff et al., 2012) or remains relatively constant (Allen et al., 2005; Fjell et al., 2009; Jack et al., 2008).

Alongside shrinkage, the ageing brain undergoes a series of other MR-detectable structural transformations: Sulci widen while also becoming shallower (Kochunov et al., 2005); the cortex becomes more curved (Wang et al., 2019); and the ventricular system expands (Jiang et al., 2014; Sigurdsson et al., 2012). As will be discussed in more detail in Chapter 3, how we measure MR-derived morphology impacts how well we can describe and understand this ageing brain metamorphosis, and the extent to which these structural changes explain cognitive decline.

1.4.2.2 Mechanisms of brain ageing

What neural mechanisms might underlie these structural brain changes? While I do not, in this thesis, investigate the biological substrates of brain ageing directly, I think it is crucial to use mechanistically plausible hypotheses to guide the interpretation of brain imaging studies. Chapters 3 and 4 show how MR-derived imaging metrics can be harnessed to this end, discussing the neural mechanisms of cognitive decline and exercise-induced cognitive improvements, respectively. The General Discussion (Chapter 5) explores how *integrative* neuroscience might provide answers to some of the largest open questions of how the mind and the brain age. The subsequent paragraph is meant to provide a very brief overview of the basic molecular and cellular mechanisms underlying brain ageing.

Until the latter half of the 20th century, neuronal death was thought to be the principal driver of brain ageing (and, for that matter, of brain shrinkage (Bishop et al., 2010)). Today, scientists believe that the main reason for brain atrophy is the loss, or reduced arborization, of dendrites and axons (Yankner et al., 2008), causing neurons to shrink in volume, not necessarily in number. This, in turn, is hypothesized to be triggered in part by wide-spread chronic, age-related inflammation (a phenomena termed "inflammaging"; Frank-Cannon et al., 2009). An important source of neuroinflammation in the ageing brain are the proliferative glial cells (astrocytes, oligodendrocytes and microglia). These cells normally provide structural, metabolic and trophic support to neurons (Allen & Barres, 2009). However, they can also have

detrimental effects on neighbouring neurons due to the chronic production of pro-inflammatory factors, including reactive oxygen species (ROS) and leukocyte-attracting cytokines, a process that occurs with increasing frequency during ageing (Chinta et al., 2015).

1.4.3 Theories of cognitive ageing

The following sections describe three main theories which have been proposed to explain inter-individual differences and cognitive and brain ageing: i) cognitive reserve, ii) brain maintenance and iii) scaffolding. Note that these theories are not mutually exclusive; it has been argued that these (and other) theories should draw sharper conceptual boundaries in order to more reliably and usefully explain differences in cognitive ageing (Barulli & Stern, 2013). This section concludes with a summary of a more recently proposed neural mechanism of (healthy) cognitive ageing: the release of noradrenaline by the locus coeruleus.

1.4.3.1 Cognitive reserve

In 1988, a group of researchers at the University of California San Diego first described what became one of the greatest puzzles in cognitive neuroscience. They inspected 137 post-mortem brains of residents of a nursing facility, whose cognitive abilities had been evaluated during life. To the researchers' surprise, ten subjects who had maintained excellent (top quintile) cognitive performance until the very end of their lives had the same pathological features of some of their peers with Alzheimer's disease (Katzman et al., 1988). The observation that a person can present with considerable brain pathology which point to symptoms of dementia, without, in fact, showing any such symptoms, has since been made by other post-mortem (Mortimer, 1997; MRC CFAS, 2001) and brain imaging (Bartrés-Faz & Arenaza-Urquijo, 2011; Giovacchini et al., 2019; Groot et al., 2018; Laubach et al., 2018; Neth et al., 2020) studies.

How can this apparent brain-cognition paradox, and individual differences in cognitive ability more generally, be explained? According to the theory of *cognitive reserve* (Stern, 2002, 2009), two possible neurobiological mechanisms are at play. The first is the efficiency of brain networks: more efficient, capable or flexible brain networks might allow for better cognitive function even in the presence of brain atrophy or pathology (e.g., Martínez et al., 2018). The second mechanism is compensation, whereby the brain accesses brain structures or networks beyond those normally used for a given task, when the latter are disrupted by pathology (Cabeza & Dennis, 2013). Cognitive reserve has been used to explain the epidemiological findings whereby education, occupational exposure and/or leisure activities are associated with reduced risk of dementia and slower rates of cognitive decline (see Robertson, 2014 for a review). However, others have argued that the concept of cognitive reserve has no operational definition, making it hard to directly investigate whether it does, or does not, explain individual differences in cognitive ageing (Nilsson & Lövdén, 2018).

1.4.3.2 Brain maintenance

Regardless of the role of cognitive reserve, another reason that some older individuals show no, or little cognitive decline could simply be that they have relatively intact brains – that is, that there a match between (maintained) brain capacity and cognitive functioning. In other words, even if some people can maintain cognition despite dramatic brain changes (as in previous section), it is still generally good for cognition to keep the brain healthy. Older adults differ in the degree of cellular damage to brain structure, and according to the "brain maintenance" theory, these differences are reflected in an age-related increase in variability in cognitive function (Nyberg et al., 2012). Whereas cognitive reserve denotes ways of coping with brain pathology, the maintenance theory focuses on the relative lack of brain changes or decline as key to preserved cognition in older age. Robertson argues that brain maintenance and cognitive reserve are closely related and complementary theories, as healthy ageing likely requires compensatory adjustments to small neural declines (Robertson, 2014).

One important question is whether brain maintenance is genetically pre-determined or whether individual choices during one's lifetime affect the degree to which the brain can be "maintained". Convincing evidence for the latter notion would come from observed improvements in a neural measure (such as cellular or neurochemical markers or MR-derived brain structure) being positively associated with changes in cognitive performance. Such change-change associations are still relatively rare and are discussed in more detail in Chapter 4. Briefly, there is evidence from structural imaging studies of "neural restoration" after both cognitive (Lövdén et al., 2010) and physical (Kramer & Erickson, 2007) intervention programs.

1.4.3.3 Scaffolding

The Scaffolding Theory of Aging and Cognition (STAC) has been developed to show how the combined effects of adverse and compensatory neural processes lead to inter-individual differences in cognitive function (Park & Reuter-Lorenz, 2009). The theory posits that increased prefrontal cortex activation in older adults performing a cognitively challenging task (observed in some functional imaging studies, e.g., Gutchess et al., 2005) is a marker of an adaptive brain that engages in compensatory scaffolding in response to its structural decline. However, others have shown that increases in prefrontal activation do not reflect compensation, but rather reduced neural efficiency or specificity (Morcom & Henson, 2018). STAC would predict stronger levels of (white matter) connectivity and neurogenesis in brains which successfully "scaffold" (Reuter-Lorenz & Park, 2014). Compelling evidence for STAC would therefore stem from change-change studies, whereby increases in connectivity (in the presence of other structural decline) are associated with improvements of cognitive functioning. Perhaps because it is difficult to reliably measure and test this effect, no such longitudinal papers exist to my knowledge. *If* neural scaffolding does take place in the ageing brain, then just like for brain maintenance described above, the extent to which such a phenomenon can be influenced with medical, lifestyle or cognitive interventions

is an important question in the endeavour to slow down or prevent cognitive decline and dementia.

1.4.3.4 The noradrenergic theory of cognitive ageing

It can be difficult to operationally (let alone mechanistically) define, and therefore empirically test, the above theories. The noradrenergic theory of cognitive ageing is based on a plausible biological mechanism, rather than simply an observed phenomenon (e.g., that some people seem to have intact cognitive abilities despite signs of atrophy; see above), making it a potentially useful lens through which to study cognitive and brain ageing. Note, this theory is not separate from the concepts described above – instead it has been proposed as a candidate neuro-mechanism of cognitive reserve (Clewett et al., 2016).

The Locus Coeruleus (LC) is a small, elongated nucleus inside the brain stem. It is the brain's sole source of noradrenaline (NA; also called norepinephrine): a neuromodulator which was thought to be mainly involved in the control of heart rate, sleep-wake cycles, and blood pressure, but which has since been understood to play an important role in regulating attention, memory, and other cognitive abilities (Sara, 2009). Noradrenaline is also neuroprotective: it lowers toxicity in neurons (Counts & Mufson, 2010) and can buffer brain cells from oxidative stress (Troadec et al., 2001). What motivates the LC to release this shield-like, cognition-enhancing chemical? Studies in rodents show that the LC-NA system is mobilized to face environmental challenges (see Sara & Bouret, 2012 for a review): when confronted with, for example, an environmental stressor (such as a gate rapidly closing in a rat's maze), the LC releases noradrenaline into the forebrain, activating an efficient and approportionate cognitive response to the stimulus (e.g., the rat changes direction). Some researchers have argued that *novelty* and *environmental stimulation* are the most reliable predictors of LC activation and, consequently, NA release (Duszkiewicz et al., 2019; Mather & Harley, 2016). For example, in rats, long-term environmental enrichment increased the presence of NA by 68 percent.

What might such noradrenaline-triggering stimuli look like in humans? A pioneering neuroimaging study (Murphy et al., 2014) found a LC BOLD response to an oddball task, a classic (fMRI) attentional paradigm where participants are infrequently confronted with an unusual (oddball) stimulus and have to respond using via finger-pressing. Moreover, Murphy and colleagues (2014) showed that the LC BOLD response covaried with increased pupil diameter, suggesting that pupil size can be used as a window into (LC-dependent) neural substrates of cognition (Joshi & Gold, 2020). An additional source of evidence that the LC modulates human cognition – and that this effect can be detected using MRI – comes from studies showing that LC integrity is associated with better cognitive abilities (e.g., Dahl et al., 2019).

Normal (i.e., nonclinical) ageing is known to negatively affect the LC-NA system, as shown by LC cell loss (Manaye et al., 1995) and changes in LC MRI signal intensity (Clewett et al., 2016). The LC is also often the first place where Alzheimer's-related pathology appears, with most people who later develop Alzheimer's showing at least some tau accumulation in their mid-20s (Braak et al., 2011). Finally, age-related reduction of LC structural integrity, has been associated with impaired cognitive and behavioural function in Cam-CAN (Liu et al., 2020), suggesting that MR-derived LC measures can be used to capture individual differences in cognitive ageing.

Based on these (and other) animal and human studies, the noradrenergic theory of cognitive ageing (Mather & Harley, 2016; Robertson, 2013) posits the following: the neuroprotective effects of noradrenaline released by the LC are partially responsible for preserving cognitive abilities in old age. Because novel, complex situations trigger the human LC-NA system, the more people engage in intellectually, physically and socially stimulating activities (see following section), the more NA will be released into their brains, protecting their minds from cognitive decline. Although enticing, this theory is based on several as-of-yet unconfirmed assumptions. Many open questions will need to be addressed, including: What, if any, real-life conditions trigger the LC-NA response? Do such triggers work throughout the lifespan or is NA more readily released in early- or mid-life? Does increasing the frequency of NA-release protect against, or even

prevent, age-related damage to the LC? Even if the LC-NA system partially accounts for individual differences in cognitive aging, how much of the variance does it explain? Even though the LC-NA system is not assessed in this thesis, I would argue that these (and other) questions provide specific, tangible avenues through which to test a mechanistically plausible theory of cognitive ageing, allowing the field to gain exciting and concrete insight into the ageing mind and brain.

1.4.4 Modifiable lifestyle activities

A theme already addressed in this introduction and investigated in Chapters 2 and 4 of this thesis is whether there is something one can do – ideally in early or mid-life – to slow down age-related cognitive decline. Might learning a musical instrument help, or joining a reading group, or volunteering for a local charity? In other words, does partaking in socially, intellectually, or physically stimulating activities help to maintain cognitive abilities in old age? Chapter 2 explores some of these questions by addressing a broad range of (possibly) *modifiable lifestyle activities*. Chapter 4 focuses on the association between physical health and cognition. The following sections are meant to provide a general backdrop against which to assess claims about the relationship between lifestyle and cognitive abilities.

The best evidence for or against the benefit of modifiable lifestyle activities comes from randomized controlled trials (RCTs). This is because such intervention studies, if done carefully, allow for causal inferences. RCTs should ideally be targeted at seniors or adults in mid-life to ensure that the benefit of a modified lifestyle activity does not soley apply to younger adults, or that the beneficial activity must be engaged in for most of a person's life.

A systematic review of 24 such mid-life RCTs (addressing 10 personally modifiable factors, two of which were classified as "lifestyle": mindfulness and social engagement; the other categories were nutritional supplements, prescription drugs, and dietary

factors) showed that most (lifestyle and other) interventions had no or very small effects on cognitive abilities (Lehert et al., 2015). Studies assessing the impact of intellectually challenging activities point to more consistent successes: a recent systematic review of cognitive training RCTs concluded that there is good evidence that mid-life cognitive training improves the trained cognitive task (especially processing speed) in healthy adults (Butler et al., 2018). However, these gains do not seem to generalise to domains not trained. As discussed in more detail in Chapter 4, the strongest evidence for lifestyle activity-induced improvements to cognitive abilities stems from studies investigating physical (especially aerobic) activity. According to a systematic review of 29 RCTs, aerobic exercise training is associated with improvements in attention, processing speed, executive function and memory (Smith et al., 2010).

In the 2015 systematic review mentioned above, of the three included mindfulness RCTs (which assessed the effect of six months of Hatha yoga, qi gong, and tai chi, respectively), only the tai chi study reported significant improvements to cognitive outcomes (Lehert et al., 2015). Should that lead us to conclude that tai chi helps, while qi gong does not? Probably not. These results are an example of a larger problem in the field of cognitive ageing: the difficulty of disentangling possible mechanisms of causation from a host of relatively unspecific interventions (let alone from observational data). While clinical drug trials are based on biologically plausible mechanisms of disease (and its cure), a lifestyle RCT is usually launched based on results from (often observational) studies pointing to possible benefits said lifestyle activity. Exactly how or why tai chi – as opposed to qi gong – may improve cognition remains unknown. This, conversely, leaves no scientific explanation for the presence – or absence – of the benefits of (related) activities, crowding the field with mechanistically untestable (null) results, and creating an unsatisfying "collective shrug" situation: even *if* mindfulness activities like tai chi improve cognition, the mechanisms underlying this effect remain speculative. This limits our understanding of the effect and the possible magnitude of any benefit to cognition, making it difficult to specify recommendations to clinicians or

policy makers. (A successful translation of scientific findings into policy is, surely, the ultimate goal of this area of research).

What would help is a third variable: rather than just assessing the effect of an independent variable (e.g., tai chi) on a measured outcome (e.g., cognition), the presence of a third (biologically meaningful) marker variable would make the possible benefits of a lifestyle activity measurable and tangible. Such *biomarkers* could be MR-derived (for instance, one study on tai chi found increases brain volume in the treatment compared the active control group) but are ideally one level further down in the chain of causation. Measures of the LC-NA system outlined in the previous section are an example of mechanistically plausible third variables. I discuss another such marker – brain derived neurotrophic factor – and its causal role in the relationship between aerobic exercise and cognitive abilities in Chapter 4. That chapter also tests the utility of MR-derived metrics as mediators in the lifestyle-cognition relationship. Generally, the presence of a well-understood biomarker would move the focus of investigation away from asking "Does modifiable Activity X improve cognition?" to instead exploring whether Activity X taps into (or activates) a well-understood neural mechanism of cognitive maintenance or improvement.

A final problem with RCTs is that they are expensive to run for longer periods of time, and thus it is possible that any trial-caused outcomes are only temporary and short-lived. Ideally, one would track the consequences of an intervention in mid-life, say, over the subsequent 2-3 decades, to see whether it really affects normal ageing. I am not aware of any such RCTs. In the absence of these prospective studies, the best one can do is investigate the cognitive or brain health of people in late life as a function of retrospective questionnaires about what lifestyle choices those people made 2-3 decades earlier (see Chapter 2, and Borgeest et al., 2020; Chan et al., 2018).

To summarize, even high-quality RCTs of potentially modifiable lifestyle activities come with important weaknesses. First, intervention studies often differ in their design and results, making it difficult to contextualize and generalise their findings. Second, RCTs

rarely offer mechanistic explanations of the effect under investigation, making it difficult to accurately interpret (null) results and translate their findings into tangible policies. Despite these issues, it is, I think, imperative to investigate if and how modifiable activities attenuate cognitive decline: Population ageing poses a real risk to individuals and societies alike – in the absence of a "cure for ageing", cognitive decline-prevention strategies are the best chance we have at ensuring that living a longer life means also living a healthy one.

1.5 Summary and outlook on subsequent chapters

This General Introduction described the societal, methodological, and scientific context in which this PhD is embedded: the rapidly ageing global population, the promises of "Big Data" in cognitive neuroscience, the statistical and neuroimaging approaches employed in this thesis, and the possible reasons for (individual differences in) age-related cognitive decline. The subsequent empirical chapters hope to contribute to this scientific landscape in the following ways: Chapter 2 explores the relationship between a broad range of lifestyle factors and people's cognitive abilities. While cross-sectional and observational in nature, it offers well-powered and well-modelled evidence for potentially accumulative benefits of engaging in multiple theoretically modifiable aspects of lifestyle. Chapter 3 focuses on brain structure, showing how large neuroimaging datasets can be harnessed to capture ageing in previously often overlooked ways. Chapter 4 assess lifestyle, brain structure and cognitive ageing simultaneously by exploring the cross-sectional and longitudinal mediating role of morphometry in the relationship between physical and cognitive health. Finally, the General Discussion (Chapter 5) discusses this thesis's findings, strengths, and limitations in the context of the existing literature, as well as with regards to the translatability of this research into policy recommendations.

Chapter 2: Greater lifestyle engagement is associated with better age-adjusted cognitive abilities

2.1 Chapter Summary

Previous evidence suggests that modifiable lifestyle factors, such as engagement in leisure activities, might slow the age-related decline of cognitive functions. Less is known, however, about which aspects of lifestyle might be particularly beneficial to healthy cognitive ageing, and whether they are differentially associated with distinct cognitive domains (e.g. fluid and crystallized abilities). I investigated these questions in the cross-sectional Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data (N = 708, age 18–88), using data-driven exploratory structural equation modelling, confirmatory factor analyses, and age-residualized measures of cognitive differences across the lifespan. Specifically, I assessed the relative associations of the following five lifestyle factors on age-related differences of fluid and crystallized age-adjusted abilities: education/SES, physical health, mental health, social engagement, and intellectual engagement. I found that higher education, better physical and mental health, more social engagement, and a greater degree of intellectual engagement were each individually correlated with better fluid and crystallized cognitive age-adjusted abilities. A joint path model of all lifestyle factors on crystallized and fluid abilities, which allowed a simultaneous assessment of the lifestyle domains, showed that physical health, social and intellectual engagement, and education/SES explained unique, complementary variance, but mental health did not make significant contributions above and beyond the other lifestyle factors and age. The total variance explained for fluid abilities was 14% and 16% for crystallized abilities. My results are compatible with the hypothesis that

intellectually and physically challenging as well as socially engaging activities are associated with better crystallized and fluid performance across the lifespan.

The work in this chapter was published in (Borgeest et al., 2020).

2.2 Introduction

Cognitive abilities are known to decline with age (Harada et al., 2013; Salthouse, 2009). The extent to which leading an active lifestyle can slow down this decline has been debated in the literature, with some studies associating physical, intellectual and social activities with cognitive and neural health while others did not find such relationships (Bielak et al., 2007; Bosma et al., 2003; Chan, Shafto, Kievit, Matthews, Spink, Valenzuela, & Henson, 2018; Clare et al., 2017; Gow et al., 2017; Lövdén et al., 2005; Newson & Kemps, 2005; Small et al., 2012a). In this chapter, I address three concerns regarding the possible associations between lifestyle and cognitive age-adjusted abilities.

First, the relationship between lifestyle and cognition has predominantly been studied by assessing lifestyle activities separately (e.g., by focusing on physical health or social engagement, but rarely both). Previous studies which have assessed various aspects of lifestyle have tended to rely on separate linear regressions (Clare et al., 2017; Crowe et al., 2003; Gow et al., 2012), mediation analyses (Brown et al., 2016) or sum scores (Karp et al., 2006; Newson & Kemps, 2005) for their analyses, limiting the extent to which the multidimensionality of people's lives can be captured, and possible complementary benefits of lifestyle detected. Thus, unless these factors are analysed mathematically simultaneously, it remains an open question as to whether individual lifestyle factors will 'sum up' to demonstrate incremental benefits, or rather be redundantly associated with better outcomes (see also Kremen et al., 2019). The structural equation modelling approach, outlined below, addresses this gap in the literature by offering several benefits compared to previous approaches. First, I model both cognitive and lifestyle factors as

latent variables, which abstracts away from individual variables whilst reducing measurement error associated with simple sum scores. Latent variables widen the interpretability of lifestyle-cognition associations to activity types (for instance 'social activity') instead of individual activities (e.g., 'attending church'). Moreover, I model multiple lifestyle factors within the same large healthy sample, allowing me to compare effect sizes. Most uniquely, the structural model captures the simultaneous effect of multiple latent lifestyle factors on cognitive lifespan differences, allowing me to investigate whether associations of specific lifestyle domains remain after taking into account distinct, but correlated, lifestyle factors.

Second, little is known about whether different aspects of cognition are associated differently with lifestyle engagement. Following a distinction first made by Cattell (Cattell, 1943), cognitive abilities can, at their broadest level, often be grouped into fluid and crystallized abilities (although newer, more detailed conceptualizations are available, Schneider & McGrew, 2012, I focus on fluid and crystallized for their importance in theories of cognitive ageing). Fluid intelligence refers to the ability to solve novel problems in the absence of task-specific knowledge or experience. It predicts important life outcomes such as expected income or work performance (Gottfredson & Deary, 2004). Age produces a marked impairment in fluid intelligence; a decline that begins in early adulthood (see Schaie, 1994 for a review). Moreover, recent findings have demonstrated that individual declines in fluid intelligence are highly correlated with individual declines in the ability to live and function independently (Tucker-Drob, 2011). Crystallized intelligence refers to acquired knowledge about the world (such as vocabulary) and shows more modest changes with age than fluid intelligence, typically declining only in old age (i.e. after the late sixties; Salthouse, 2000). One open question, addressed here, is whether crystallized and fluid abilities, known to differ in their lifespan trajectories, also benefit differently from measures associated with better cognitive ageing.

Third, it has been difficult to reliably identify those lifestyle activities that enhance cognitive reserve, as is demonstrated by the considerable heterogeneity of findings in

the literature (Kralj et al., 2018). This is likely to be due to at least two reasons. One concerns the large diversity of lifestyle variables that have been assessed, with studies differing on the types of activities that make up, say, social engagement. A second explanation is the variable and often imprecise definition of 'healthy ageing' in cross-sectional studies. For instance, many cross-sectional studies rely on classifying groups of people according to their absolute performance on cognitive tests (e.g., Folstein et al., 1975; Taylor et al., 2017). In such an approach, older individuals who score an arbitrary number of standard deviations above a task mean are labelled 'healthy', 'successful', or in some cases even 'super' agers (Gefen et al., 2014; Harrison et al., 2012; Lin et al., 2017; Rogalski et al., 2013; Sun et al., 2016), while those beneath this cut-off point are considered to age only 'normally' or 'poorly'. Here, I conceptualize 'healthy ageing' in terms of 'age-adjusted cognitive abilities', by using a simple continuous age-residualized measure, which I describe in more detail below. This measure avoids the drawbacks of arbitrary statistical cut-off points and dichotomisation (McClelland et al., 2015), and allows for a natural conceptualization of age-adjusted cognitive abilities, namely whether an individual is performing better or worse than would be expected at her age.

2.2.1 The present study

Although enhanced physical, mental and social lifestyle factors have all been associated with healthier cognition, these effects have predominately been investigated separately (e.g., by looking at physical health *or* social engagement, but rarely both). Simultaneous analysis of these associations would shed more light on the possible complementary benefits of various aspects of people's lives. Moreover, understanding if lifestyle is associated differently with crystallised and fluid cognition is important in order to shape effective interventions. I therefore investigated the simultaneous associations between various aspects of lifestyle and both fluid and crystallized resilience. I used a large (N = 708) age-heterogeneous population-based sample from the Cambridge Centre for Ageing and Neuroscience (Cam-CAN), employing age-residualized measures of

cognition, data-driven exploratory structural equation modelling and confirmatory factor analysis.

2.3 Methods

2.3.1 Participants

Participants were drawn from the Stage 2 sample of the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) dataset, described in more detail in the Chapter 1 of this thesis as well in in other papers (Shafto et al., 2014; Taylor et al., 2017). 708 people (359 women, 349 men) were recruited, including approximately 100 people in each decile (age range 18–88, M = 53.4, SD = 18.62). Participants provided a wide range of cognitive measures and questionnaire data, summarized below.

2.3.2 Cognitive variables

Thirteen cognitive tasks were used to assess five broad cognitive domains, which are described in Table 2-1. The cognitive domains assessed were executive functions, memory, language functions, motor and action function and emotional processing.

Table 2-1: description of cognitive behavioural tasks

Cognitive Domain	Cognitive Task	Task Description	Descriptive Statistics (Mean, SD, Range, Missingness)	References
Executive Function	Fluid Intelligence	Cattell Culture Fair Test, incl. nonverbal puzzles involving series completion, classification, matrices, and conditions.	M=31.8 SD=6.76 Range=11-44 Missing=6.8%	R.B. Cattell & Cattell, 1960
	Multitasking (Hotel Task)	Perform tasks in role of hotel manager: write customer bills, sort money, proofread advert, sort playing cards, alphabetise list of names. Total time must be allocated equally between tasks; there is not enough time to complete any one task.	M=3.07 SD=1.74 Range=0.2-9.6 Missing=7.1%	Shallice & Burgess, 1991
Language Functions	Spot the Word	Involves presenting an individual with pairs of items comprising one word and one non-word, for example, 'flonty – xylophone', the individual is required to point to the real word in the pair.	M=53.58 SD=5.39 Range=24-60 Missing=0.42%	Baddeley, Emslie & Nimmo-Smith, 1993
	Sentence Comprehension	Listen to and judge grammatical acceptability of partial sentences, beginning with an (ambiguous, unambiguous)	M=0.89 SD=0.07 Range=0.46-1 Missing=11.4%	Rodd, Longe, Randall, & Tyler, 2010

	sentence stem (e.g., "Tom noticed that landing planes…") followed by a disambiguating continuation word (e.g., "are") in a different voice. Ambiguity is either semantic or syntactic, with empirically determined dominant and subordinate interpretations		
Picture-Picture Priming	Name the pictured object presented alone (baseline), then when preceded by a prime object that is phonologically related (one, two initial phonemes), semantically related (low, high relatedness), or unrelated	M=0.78 SD=0.09 Range=0.5-0.94 Missing=8.3%	Clarke, Taylor, Devereux, Randall, & Tyler, 2013
Verbal Fluency	Mean of Letter (phonemic) fluency and animal (semantic) fluency task. For phonemic fluency task, participants have 1 min to generate as many words as possible beginning with the letter 'p'. For semantic fluency task, participants have 1 min to generate as many words as possible in the category 'animals'.	M=20.56, SD=5.34 Range=6-37.5 Missing=0.28%	Lezak, Muriel, & Deutsch, 1995

	Proverb Comprehension	Read and interpret three English proverbs.	M=4.53 SD=1.63 Range=0-6 Missing=7.5%	Hodges, 1994
Emotional Processing	Face Recognition	Given a target image of a face, identify same individual in an array of 6 face images (with possible changes in head orientation and lighting between target and same face in the test array)	M=22.88 SD=2.36 Range=14-27 Missing=7.2%	Benton, 1994
	Emotion Expression Recognition	View face and label emotion expressed (happy, sad, anger, fear, disgust, surprise) where faces are morphs along axes between emotional expressions.	M=8.66 SD=1.09 Range=3.33-10 Missing=7.1%	(Ekman & Friesen, 1976)
Memory	Visual Short-Term Memory	View (1–4) coloured discs briefly presented on a computer screen, then after a delay, attempt to remember the colour of the disc that was at a cued location.	M=2.43 SD=0.59 Range=0-3.5 Missing=7.3%	(W. Zhang & Luck, 2008)
	Story Recall	Listen to a short story, recall freely immediately after, then again after a delay, and finally answer recognition memory questions. Delayed recall measure used here.	M=12.88 SD=4.31 Range=0-24 Missing=0.14%	(Wechsler, 1999)
Motor and Action Function	Choice Motor Speed	Time-pressured movement of a cursor to a target by	M=0.19 SD=0.06 Range=0.05-0.85	

	moving an (occluded) stylus under veridical, perturbed (30°), and reset (veridical again) mappings between visual and real space.	Missing=7.34%
Choice Motor Coefficient of Variation	Standard deviation divided by mean of reaction time of choice motor speed. Reflects the relative measure of variability.	M=1.84 SD=0.38 Range=0.86-2.98 Missing=7.34%

2.3.3 Lifestyle variables

I included a broad set of 23 lifestyle measures from the Cam-CAN dataset, which were collected via a series of different questionnaires, summarized in Table 2. Eight lifestyle variables were obtained during the Home Interview, an extensive face-to-face interview conducted at Stage 1 of Cam-CAN data collection. The remaining variables were obtained during the main data collection period for the Cam-CAN cohort (Stage 2). Measures of physical activity, depression and sleep were assessed via the physical activity energy expenditure (PAEE) questionnaire, the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983) and the Pittsburgh Sleep Quality Index (PSQI; (Buysse et al., 1989), respectively. The remaining 12 lifestyle variables were taken from the Lifetime of Experiences Questionnaire (LEQ; Valenzuela & Sachdev, 2007), which measures a broad range of cognitively stimulating experiences and activities during three life phases: youth, 13-29 years; mid-life, 30-64 years; and late-life, 65 years onwards. Within each phase, further details about activities "specific" to that time of life (e.g., education in youth) were solicited, as well as "non-specific" activities applicable to any phase (e.g., socialising). The LEQ therefore provides information about current life experiences for all participants, as well as retrospective information about previous life experience for participants in their mid- and late-life phases. Usually, this information is reflected in one specific and one non-specific sum score for each stage of life. In this

study, however, I focused on a more fine-grained scoring procedure. First, I defined the measure of education as the young-age specific score, derived from the UK's National Career Service categories and multiplied by number of years at each category. Second, I included only current non-specific activities depending on the age of the individual, as I wanted to focus on contemporaneous activities, and allow consistent data across the full age range. Third, as the core goal of this study was to assess multiple lifestyle aspects and their relationships to cognition, I obtained separate scores for the seven non-specific questions, rather than calculating the usual sum-score. As these seven questions (see Table 2-2) cover a range of lifestyle activities, individual scores allowed me to more precisely determine their covariances to other lifestyle factors. Non-specific activities were assessed through the same seven questions during youth, mid-life and late-life, addressing participation in i) travel, ii) social outings, iii) playing a musical instrument, iv) artistic pastimes, v) physical activity (mild, moderate, vigorous), vi) reading, vii) learning or speaking a second language. In addition, participants were asked whether their typical day included any of the following activities: i) internet use, ii) strategic games (e.g., chess, bridge, cards), iii) prayer/religious activity, iv) brain training games. All non-specific scores were scaled to a score from 0-5.

Table 2-2: Description of lifestyle variables.

Lifestyle Factor	**Variable**	**Description/Question**	**Descriptive Statistics**	**Reference**
Education /SES	Income	What is the average total income before tax received by your household? (1-5)	M=2.83 SD=1.49 Range=1-5 Missing=0.14%	HI
	Smoking habits	category of smoking based on self-report questions (1-3)	M=1.03 SD=0.97 Range=0-3 Missing=1.4%	HI
	TV watching**	How much TV do you watch per week?	M=2.2 SD=1.47 Range=0-7	HI

			Missing=61.9%	
	Body Mass Index (BMI)	Weight (kg) / Height^2 (m^2)	M=25.78 SD=4.59 Range=16.75-48.32 Missing=17.2%	HI
	Travel	Did you travel to any of the following continents between the ages of 13–30 years? (9 options available)	M=2.3 SD=1.25 Range=0-5 Missing=12.01%	LEQ
	Instrument	How often are you practising or playing a musical instrument?	M=1.97 SD=1.22 Range=0-5 Missing=12.01%	LEQ
	Language	How often do you practise speaking, reading, writing or learning a second language?	M=1.89 SD=1.26 Range=0-5 Missing=12.01%	LEQ
	Years of education	Sum score derived from the UK's National Career Service categories, multiplied by number of years at each category	M=3 SD=2.49 Range=0-13.29 Missing=12.3%	LEQ
Physical Health	Internet	Does your typical day include internet use?	M=3.39 SD=1.89 Range=0-5 Missing=12.01%	LEQ
	Exercise$^+$	Please give the typical number of hours per week you spend in sports and physical activities. Divided into mild, moderate and vigorous activities.	M=3.43 SD=1.02 Range=0-5 Missing=12.01%	LEQ
	Systolic Blood Pressure	Mean systolic blood pressure of three samples	M=120.08 SD=17 Range=78.5-186 Missing=18.1%	HI

	Physical activity	Total physical activity energy expenditure (PAEE) calculated from self-report ACTMETS (kJ/day/kg)	M=4.29 SD=2.19 Range=0-17.71 Missing=11.9%	HI
Mental Health	Depression	Hospital Anxiety and Depression Scale (HADS)	M=2.82 SD=2.58 Range=0-17 Missing=0.56%	(Zigmond & Snaith, 1983)
	Quality of sleep	Pittsburgh Sleep Quality Index (PSQI)	M=5.41 SD=3.68 Range=0-22 Missing=5.4%	(Buysse et al., 1989)
	Alcohol consumption	Amount of alcohol used weekly	M=3.29 SD=1.37 Range=0-5 Missing=3.9%	HI
	Self-Health	Self-reported health. 4-point scale; 1= excellent 4=poor	M=1.87 SD=0.69 Range=1-4 Missing=0.28%	HI
Social Engagement	Exercise^{+}	Please give the typical number of hours per week you spend in sports and physical activities. Divided into mild, moderate and vigorous activities.	M=3.43 SD=1.02 Range=0-5 Missing=12.01%	LEQ
	Social outings	How often might you make an outing to see a family member, friend or group of friends?	M=3.66 SD=1.08 Range=0-5 Missing=12.01%	LEQ
	Religious Activities	Does your typical day include prayer / religious activities?	M=2.2 SD=1.33 Range=0-5 Missing=12.01%	LEQ

	Social Mean Score	Derived from 13 question sub-section of Home interview	M=2.32 SD=0.6 Range=0-4.18 Missing=0%	HI
Intellectual Engagement	Reading	Does your typical day include reading?	M=4.68 SD=0.92 Range=0-5 Missing=12.01%	LEQ
	Brain Training Games	Does your typical day include brain training games (e.g., Computer or Nintendo)?	M=1.7 SD=1.2 Range=0-5 Missing=67.8%	LEQ*
	Strategic Games	Does your typical day include strategic games (e.g., Chess, Bridge, Cards)?	M=1.55 SD=0.98 Range=0-5 Missing=12.01%	LEQ
	Artistic Pastime	How often do you practise or develop an artistic pastime (e.g., drawing, painting, sculpture, creative writing, acting, etc.)?	M=2.09 SD=1.48 Range=0-5 Missing=12.01%	LEQ

The grouping into 'lifestyle factors' is the result of the factor analysis outlined in more detail below.

HI = Home Interview. LEQ = Lifetime of Experiences Questionnaire. * Only older participants were asked this question (N=228) ** This question was completed in a take-home questionnaire by a subset of the sample (N=270) + The LEQ exercise question cross-loaded onto Social Engagement and Physical Health in the CFA model and is thus included twice in this table

2.3.4 Exploratory Structural Equation Modelling

In order to obtain a data-driven categorization of the cognitive and lifestyle variables, I used a relatively novel technique called exploratory structural equation modelling (ESEM; Asparouhov & Muthén, 2009). ESEM integrates confirmatory factor analysis (CFA) and structural equation modelling (SEM) to provide confirmatory tests of a priori factor structures.

Based on theory, CFA measurement models specify a number of factor loadings fixed at zero to reflect a hypothesis that only certain factors influence certain factor indicators. The CFA approach of fixing many or all cross-loadings at zero may force a researcher to specify a more parsimonious model than is suitable for the data. Because of this, models often do not fit the data well and there is a tendency to rely on extensive model modification to find a well-fitting model (Tóth-Király et al., 2017). A commonly used alternative to CFA is Exploratory Factor Analysis (EFA), which solves some of CFA's challenges in situations of limited measurement knowledge of the researcher and / or a more complex measurement structure. However, EFA can be difficult to perform while allowing correlated residuals, and it assumes that all measured variables are related to every latent variable. For these reasons, researchers often opt for an ad-hoc procedure that mimics the EFA factor definitions in a SEM model with a CFA measurement specification (that is, they run an EFA first, followed by a CFA). This EFA-to-CFA conversion has been shown to be challenging, and can lead to mis-specified models (Marsh et al., 2014). The main advantage of the ESEM model over other modelling practices is that ESEM incorporates seamlessly the EFA and SEM models. ESEM integrates EFA into SEM (which otherwise relies on CFA measurement models) by estimating the measurement and structural model parts simultaneously. I used the package psych (version 1.7.8; 41) in R-Studio 1.0.153 (R version 3.4.2).

2.3.5 Age-residualized cognitive abilities

After computing the best age-related trajectories, I calculated indices of age-adjusted cognitive abilities in each domain. For this, I separately regressed fluid and crystallized factor scores on age, retaining the residual score for each participant and factor. As depicted in Figure 2-1, each residual score thus reflects the difference between the participants' observed and her age-predicted factor scores.

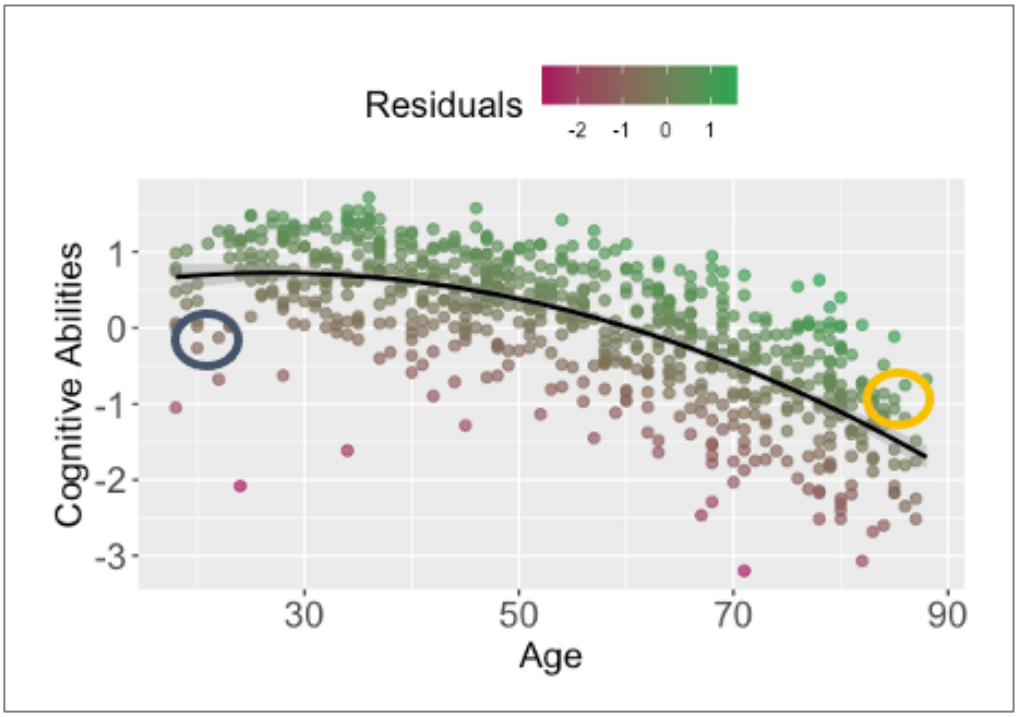


Figure 2-1: Depiction of residuals. By regressing cognitive abilities on (a second-order polynomial expansion of) age, one retains a residual score for each participant. This residual score reflects the difference between the participants' observed and her age-predicted scores. Here, pink dots refer to scores below the age-expected mean, green dots represent scores above the age-expected mean. This allows for an older person with a (compared to the full sample) relatively low score to be considered cognitively healthy (e.g., yellow circle). Likewise, a younger person with a high absolute score (blue circle) can be considered as being cognitively below his age-related peers.

Because the residuals were obtained from a curve reflecting age-related differences, they do not represent the difference between a participant's score and the overall mean, but rather of the mean expected for the participant's age (thus, the age-adjusted mean). Although these scores will still correlate with raw scores within each domain, these

residuals adjust for age-expected declines, allowing, for example, an 80-year-old person with a relatively low absolute score to be considered cognitively healthier than a younger individual with a higher score. Residualized fluid and crystallized cognition therefore serve as my measure of age-adjusted cognitive abilities in further analyses. Similar measures have been proposed to quantify brain structure adjusted for calendar age (Cole & Franke, 2017), and psychosocial functioning adjusted for the severity of adverse childhood experiences (van Harmelen et al., 2017). I tested for the assumption of homoscedastic residuals using the Breusch-Pagan test to check if the variability of the residuals increased across the lifespan. Where appropriate I also computed robust regressions to ensure heteroscedasticity did not affect my inferences.

2.3.6 Confirmatory Factor Analysis (CFA)

In the second step of my analyses, I used a set of simpler confirmatory factor analyses (CFA models) to a) achieve stable model estimation and b) facilitate detailed model comparisons. CFA is a multivariate statistical procedure that allows the researcher to specify the number of latent and observed constructs in order to test how well the former are captured by the latter. Translating my ESEM solutions to CFA models allowed me to formally test more parsimonious models that remove negligible cross-loadings, and to assess overall model fit using a more conventional range of model fit indices. Although such a two-step procedure is ideally performed on two independent subsamples of the data, this was not feasible given the necessity to balance between model complexity and sample size. While one-step, or factor score regression approaches (Devlieger & Rosseel, 2017), are generally considered preferable, challenges with convergence and model estimation precluded such approaches here. As such, I specified CFA's separately for each domain and used estimated factor scores in the second stage. All models were fit using Lavaan 0.6-1.1203 version (Rosseel, 2014). Prior to model fitting, one variable with very large variance (multitasking, measured in milliseconds) was rescaled by dividing by 100 to avoid convergence problems. All

models were fit using maximum likelihood estimation with robust (Huber-White) standard errors and a scaled chi square test statistic. Missing data, reported in Tables 1 and 2, were accounted for using Full Information Maximum Likelihood method in Lavaan, which allowed me to estimate factor scores for all individuals, including those with partially missing data and yields unbiased estimates under the assumption of missing at random or missing completely at random (Enders & Bandalos, 2001). Assuming data are either Missing Completely At Random (MCAR) or Missing At Random (MAR), i.e. the missing data can only be dependent on variables also measured within the same dataset, Full Information Maximum Likelihood (FIML) can be used to estimate a model on the full dataset (including subjects with incomplete data; Enders, 2001; Enders & Bandalos, 2001; Wothke, 2000). Given the considerable richness and diversity of data included here, I consider this a defensible assumption, and one made in many other papers. Tests exist for whether data is missing completely at random (e.g., Little's test), and although that assumption is not required for the implementation of FIML, I ran the test (using R's mcar_test function) to verify that my data were missing completely at random (Chi Square = 57.43; p =0.91). Using FIML for missing data (under multivariate normality) maximizes the utility of all existing data, decreases bias and increases statistical power compared to (for instance) omitting incomplete cases ('complete case analysis'; Baraldi & Enders, 2010). In direct comparisons, FIML usually performs as well or better than alternative methods such as multiple imputation (MI) (Larsen, 2011; von Hippel, 2016) and considerably better than complete case analysis. A practical benefit of FIML compared to MI is the stability of estimation across uses, whereas multiple imputation depends on stochastic sampling and will yield a (slightly) different estimate every time.

Model fit was inspected using the chi-square test, the Root Mean Square Error of Approximation (RMSEA) and its confidence interval, the Comparative Fit Index (CFI) and the Standardized Root Mean Square Residual (SRMR). Good fit was defined as RMSEA < 0.05, CFI > 0.97 and SRMR < 0.05, acceptable fit as approximately RMSEA = 0.08 – 0.05, CFI = 0. 95 – 0.97, SRMR = 0.05 – 0.1 (Schermelleh-Engel et al., 2003).

In order assess the slope shape crystallized and fluid intelligence cross the lifespan, I extracted factor scores from the cognitive models for all participants. I tested whether the lifespan differences of crystallized and fluid intelligence were best captured by linear, quadratic, or cubic curves by comparing the BICs of each of the models.

Finally, I examined the degree to which lifestyle factors made *specific* contributions to fluid versus crystallized cognitive differences. To do so, I refit the models while imposing equality constraints on the lifestyle paths. In other words, I compared a model where the effects of lifestyle factors are estimated individually for each of the two cognitive domains, to a more parsimonious model where the path coefficients are assumed to be identical for fluid and crystallized healthy ageing. If the effects of lifestyle factors are equal for both cognitive domains, then one would expect an equality constrained model (where the effects of lifestyle factors on cognitive domains are presumed to be equal) to fit better. However, if certain lifestyle factors have stronger, or weaker, effects on each domain, then one would expect a model that estimates all structural paths freely to fit better.

2.3.7 Exploratory Analyses

I performed a series of exploratory analyses to assess the presence of i) an interaction effect of age and lifestyle using a median split and ii) sex effects.

2.4 Results

2.4.1 Exploratory structural equation model

The sample-size adjusted BIC scores are shown in Figure 2-2 (the first number in each model name refers to the number of cognitive variables, and the second number refers to the number of lifestyle variables). The ESEM analyses revealed that, generally, two- and three factor models of cognitive abilities fit the data substantially better than a one factor model. The three factor solutions had marginally better fit than the two factor

solutions (e.g., ΔBIC = 13.55 for the 2_5 versus 3_5 model). However, I opted for a two factor solution for theoretical reasons, as the two factor solution closely resembled the canonical distinction between fluid and crystallized abilities, in line with Cattell (Cattell, 1943) and a large body of body of work on cognitive ageing (Baltes et al., 1999; Ghisletta et al., 2012). Moreover, I note that, in the two factor cognitive model, although the strongest factor loading on the first 'fluid' factor is the Cattell test, it includes a relatively large, and broad, number of cognitive abilities, several of which are beyond the traditional remit of pure fluid intelligence (Horn & Cattell, 1967).

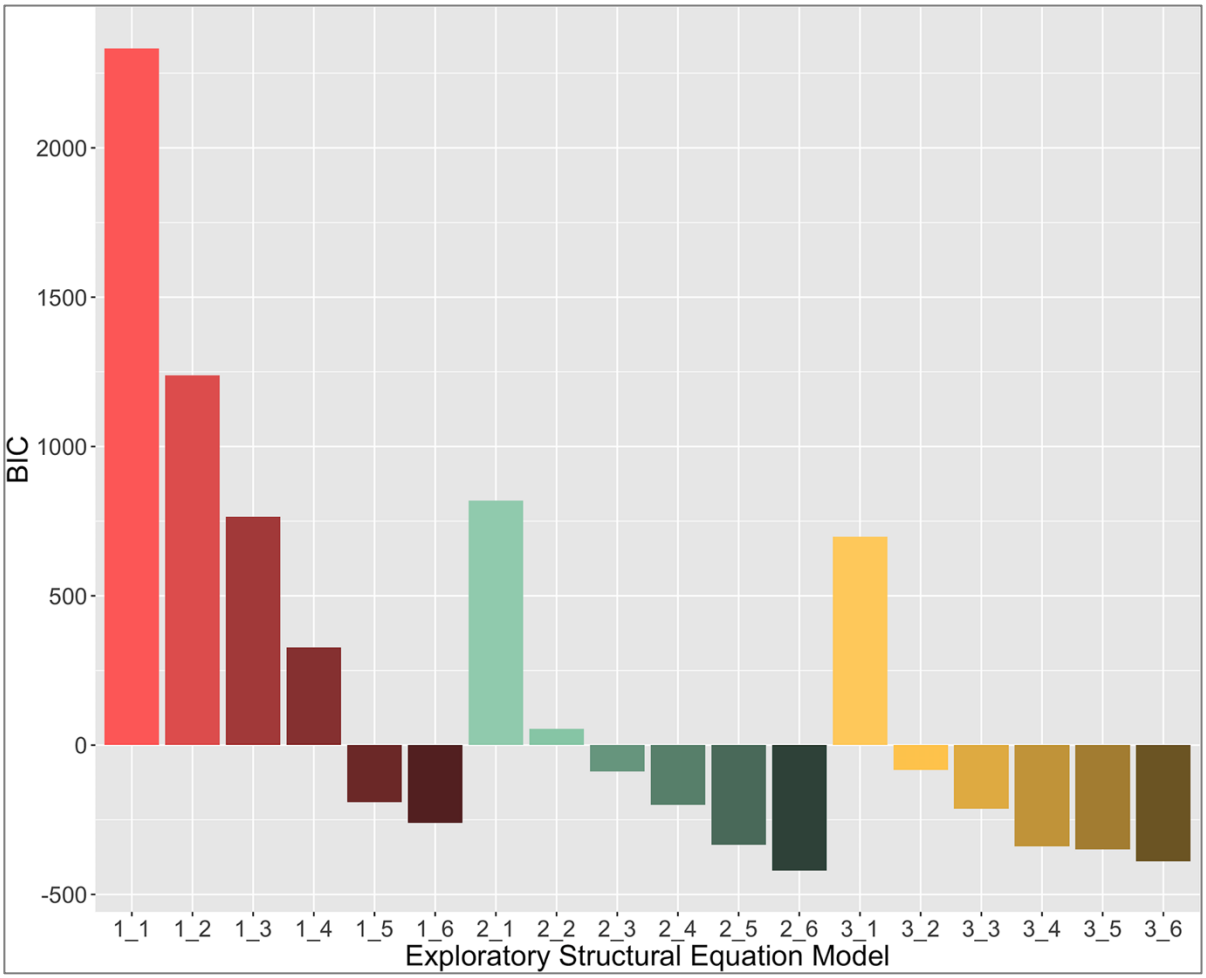


Figure 2-2: Exploratory structural equation model results. Y-axis reflects Bayesian Information Criterion (BIC) measure of model fit; X-axis labels consist of two digits separated by an underscore (e.g., 2_4), where the first refers to the number of cognitive latent variables, and the second to the number of lifestyle latent variables. Model 2_6 has the best overall fit, then Model 3_6; however, Model 2_5 was selected for further examination due difficulties interpreting the sixth lifestyle factor in the 2_6 and 3_6 models.

2.4.2 CFA: Cognitive Model

First, I fit the cognitive data with a two-factor model that mirrors the canonical distinction between crystallized and fluid abilities (Cattell, 1943). One notable exception was that this model required a single data-driven cross-loading for the sentence comprehension task, which may reflect the nature of the task as a combination of being knowledge-based (whether a sentence is grammatical) and benefiting from fluid ability. This cognitive measurement model, shown in Figure 2-3 (A) fit the data adequately: χ^2 = 233.87 (N = 708), df = 63, p <0.001, RMSEA = 0.057 [0.049 0.066], CFI = 0.93, SRMR = 0.048, suggesting that the cognitive data were well captured by a two-factor model.

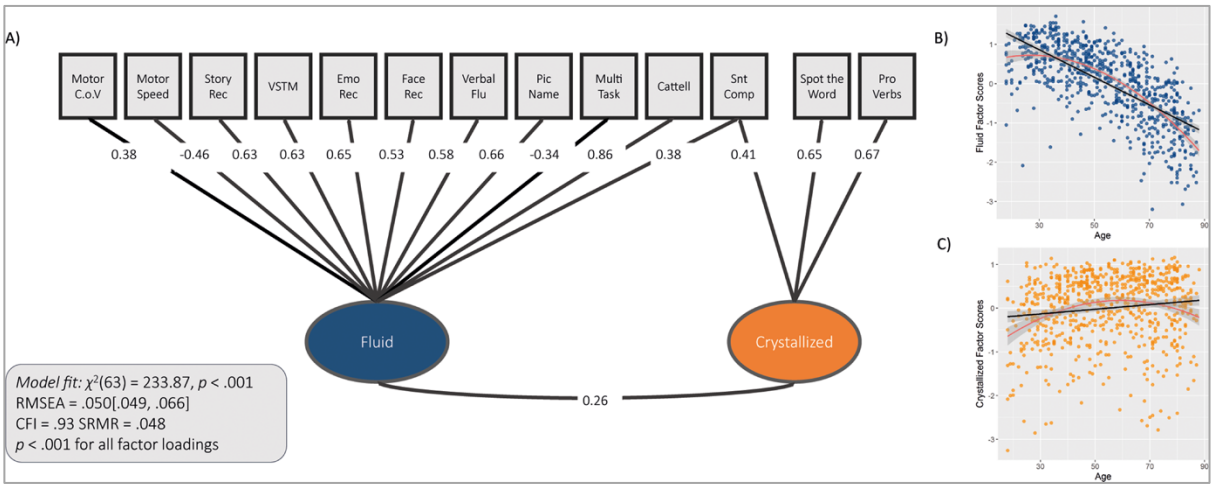


Figure 2-3: Confirmatory factor model. A) Cognitive CFA. For multitasking and motor speed, lower scores indicate better performance (hence the negative factor loadings). B) Fluid factor scores for each participant. Fluid abilities decline with age. C) Crystallized factor scores for each participant; crystallized abilities show slight increase and then decrease. All parameters shown are fully standardized.

Next, I extracted factor scores for all individuals to examine the most appropriate lifespan trajectory for each domain (linear or quadratic). As expected, fluid and crystallized factors showed different lifespan patterns. Scores on the fluid latent variable showed a strong age-related decline, with a modest acceleration of this decline in old age (Figure 2-3 B) consistent with the best-fitting model including a quadratic component (BIC_{Quadr} = 1391.15, BIC_{Lin} = 1458.09, BIC_{Cubic} = 1393.17). Scores on the

crystallized latent variable were less strongly associated with age, with a slight increase until middle age but suggestion of decline in old age Figure 2-3 C), again consistent with a quadratic component (BIC $_{Quadr}$ = 1676.27, BIC $_{Lin}$ = 1696.91, BIC $_{Cubic}$ = 1678.06).

2.4.3 Age-residualized cognitive abilities

Age-residualized measures of fluid and crystallized abilities (shown in Figure 2-4) were significantly positively correlated (Pearson's r = 0.59 [0.53 0.63], df = 706, p = < 0.001). The median (age 55) split analysis showed that the Gf-Gc correlation of residuals did not differ significantly for the two age groups (z = 0.8, p = 0.42).

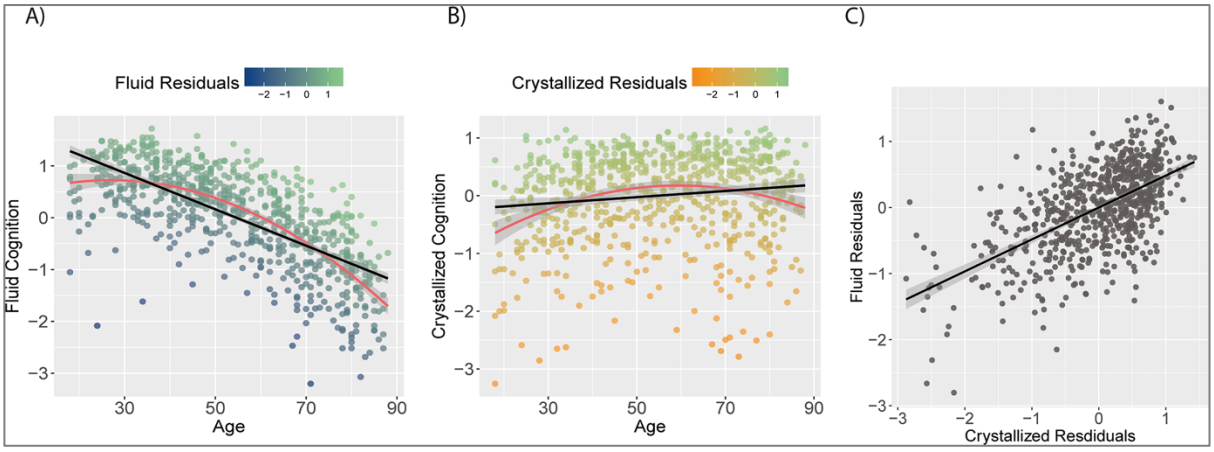


Figure 2-4: Age adjusted residuals. Residuals as measure of healthy cognitive ageing. A) Crystallized residuals, B) fluid residuals, C) correlation between crystallized and fluid residuals; r(706) = .59, p < .001.

2.4.4 CFA: Lifestyle Model

Next, I examined the lifestyle domains in more detail. To do so, I used the ESEM results to specify a simpler (fewer cross-loadings) CFA that captured the observed variables across five latent factors (Figure 2-5). Based on the pattern of loadings, I refer to these five latent variables as follows: i) Education/Socio-Economic Status (SES), ii) Physical Health, iii) Mental Health, iv) Social Engagement, v) Intellectual Engagement. Education/SES consisted of eight variables, namely years of education, income, language, travel, smoking, TV watching and instrument playing. Physical Health

consisted of systolic blood pressure, internet usage, the PAEE score and the LEQ exercise score. Mental health was captured by alcohol usage, depression, self-reported health, and sleep quality. The factor loadings of Intellectual Engagement were reading, brain training games, strategic games, Sudoku/Crossword, and the degree of engagement in artistic pastime. Lastly, Social Engagement was characterized by religious activity, social outing, the social activity score from the Home Interview and the LEQ physical exercise score. Note, the labels of the factors are for convenience and based on the strongest loadings–some include factor loadings on variables which are not canonically associated with the construct. Consider, for instance, *alcohol* for Mental Health or *smoking* for Education/SES – both are plausibly part of the respective factors but may not have been placed there using a more researcher- (rather than data-) driven categorization method. As was the case for the cognitive CFA, this lifestyle model required one data-driven cross-loading for the LEQ exercise variable, which may reflect that fact that many physical activities (e.g., basketball, hiking) include significant social aspects. The model showed adequate fit to the data in most respects: $\chi^2(241) = 747.69$ ($N = 708$), $p < 0.001$, RMSEA = 0.055 [0.050 0.059], CFI = 0.720, SRMR = 0.060, although the CFI is lower than preferable, likely due to the modest factor loadings of some variables. Given the nature of the observed scores (see Table 2), higher scores in Social and Intellectual Engagement and SES/Education reflect more engagement and increased socioeconomic status, respectively. In contrast, higher scores in the Physical Health and Mental Health factors, however, reflect poorer health as their indicators (e.g., blood pressure, mental health symptoms) are considered poor outcomes.

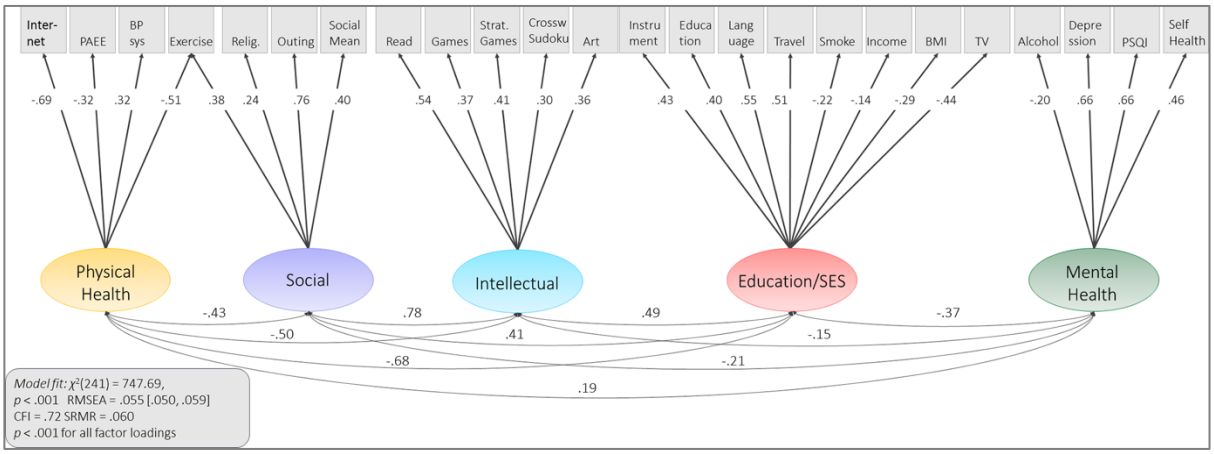


Figure 2-5: Lifestyle CFA. Following the factor loadings obtained via the ESEM, 24 broad lifestyle variables loaded onto five latent lifestyle variables: mental health, social engagement, intellectual engagement, education/SES and physical health. All parameters shown are fully standardized. All but three lifestyle factor loadings (income, internet usage and alcohol) were in the expected direction.

Note that all but three factor loadings were in the expected direction. First, income loaded negatively onto education/SES, where usually higher income is associated with higher SES. One explanation for this could be that Cam-CAN represents a wealthier and more educated sample than the general population, and that in the absence of the "full" range, the effects of income diminish. In addition, although significant, this factor loading of -0.14 was small, and should be interpreted with caution. Second, lower alcohol consumption was associated with poorer mental health, where some might have hypothesized the opposite. However, as was the case for income, the factor loading was small (-0.12), and interpretability is therefore limited. Third, more internet usage was associated with better physical health. I believe that this is largely an SES effect, such that people with higher SES (who, on average, have better physical health) also spend more time browsing the internet.

2.4.5 Separate Regressions

Next, I investigated the extent to which the five lifestyle factors determined my measures of healthy ageing. As the simultaneous estimation of the measurement models (across cognitive and lifestyle domains) and the structural model (regressing cognitive domains on lifestyle variables) could not achieve robust convergence, I used a two-step

procedure. First, I extracted the factor scores for both cognitive factors and computed age-adjusted residuals. Second, I regressed measures of age-residualized fluid and crystallized abilities on the lifestyle factor scores. Doing so, I observed significant associations between each individual lifestyle factor and both fluid and crystallized ageing, as depicted in Figure 2-6. The strongest associations were those between Education/SES and fluid (std β = 0.26) and crystallized cognition (std β = 0.33), followed by Intellectual Engagement (fluid std β = 0.24, crystallized std β = 0.22), Mental Health (fluid std β = -0.17, crystallized std β = -0.19), Physical Health (fluid std β = -0.17, crystallized std β = -0.14) and finally Social Engagement (fluid std β = 0.15, crystallized std β = 0.10).

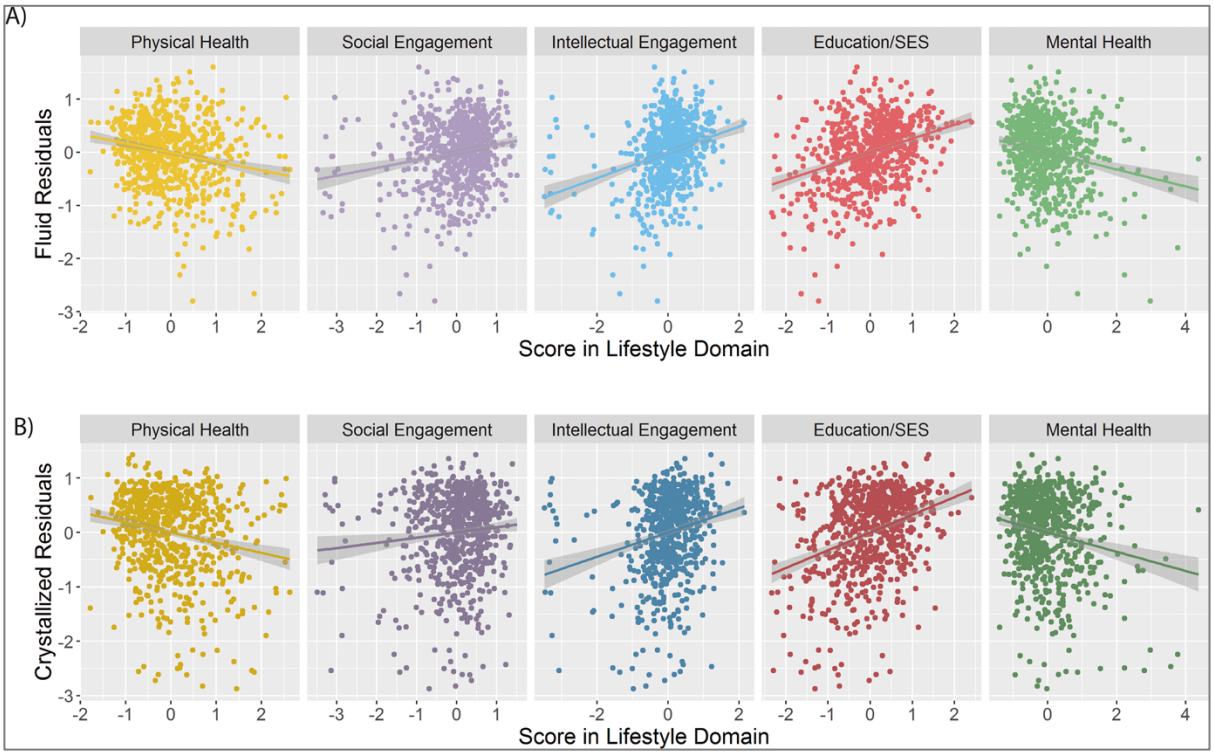


Figure 2-6: Individual path models. Separate regression results for A) fluid abilities and B) crystallized abilities. All five lifestyle factors were significantly associated with cognitive health across the lifespan.

Following recent effect size guidelines (Gignac & Szodorai, 2016), I interpret the associations between the lifestyle factors and cognition to range from relatively large (Education/SES) to typical (Intellectual Engagement, Mental Health, Physical Health), with small associations found for Social Engagement. In summary, these findings

suggest that having higher levels of education/SES as well as physical and mental health, and partaking in intellectually and socially engaging activities, are all individually associated with better fluid and crystallized cognitive outcomes throughout the lifespan, above and beyond age.

All regressions showed modest deviations of the assumption of homoscedastic residuals (all Breusch–Pagan tests $\chi^2>10$, df = 1, $p<0.01$), with a general increase in variability across the lifespan (Figure 2-7). To ensure that these heteroscedastic residuals did not affect my inferences concerning lifestyle-cognition associations, I re-estimated all models using a heteroscedasticity-consistent robust sandwich estimator (using the package 'sandwich'; Zeileis, 2004)). As can be seen in Table 2-3, the parameter estimates and standard errors are virtually identical, suggesting negligible consequences of the heteroscedastic residuals.

Cognitive domain	Lifestyle factor	Standardized beta	Standard error	*p*	R^2	Robust sandwich beta	Robust sandwich SE	p
Fluid abilities	Mental Health	-0.16	0.03	<0.001	0.04	-0.16	0.04	<0.001
	Social	0.15	0.03	<0.001	0.03	0.15	0.03	<0.001
	Intellectual	0.24	0.03	<0.001	0.08	0.24	0.04	<0.001
	Education/SES	0.26	0.03	<0.001	0.11	0.26	0.03	<0.001
	Physical Health	-0.17	0.03	<0.001	0.05	-0.17	0.03	<0.001
Crystallized abilities	Mental Health	-0.18	0.04	<0.001	0.04	-0.17	0.04	<0.001
	Social	0.10	0.03	<0.001	0.009	0.79	0.04	<0.001
	Intellectual	0.22	0.04	<0.001	0.05	0.22	0.04	<0.001
	Education/SES	0.33	0.04	<0.001	0.11	0.33	0.04	<0.001
	Physical Health	-0.19	0.04	<0.001	0.04	-0.19	0.04	<0.001

Table 2-3: Separate regression results for fluid and crystallized abilities.

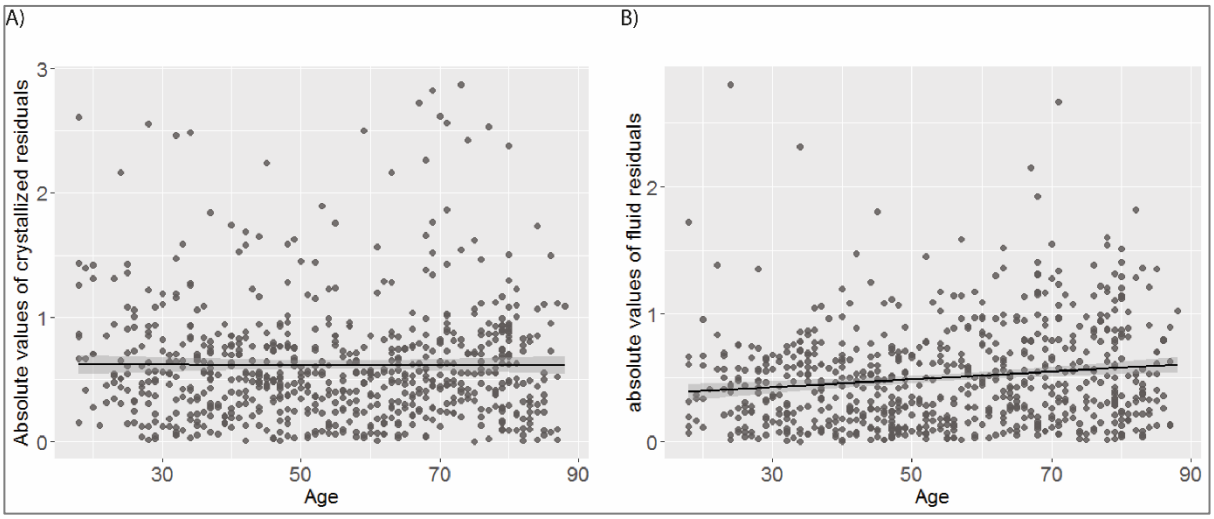
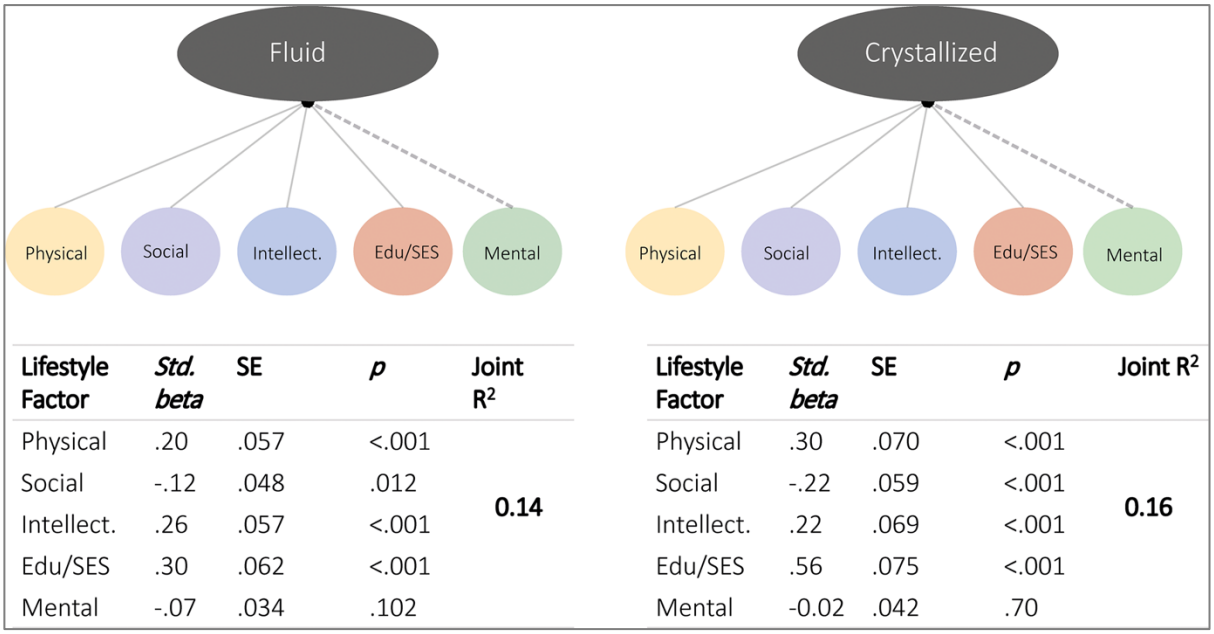


Figure 2-7: Test for homoscedasticity. Figure shows modest deviations of homoscedasticity across the lifespan.

2.4.6 Multiple Regressions

Next, I examined the joint effects of lifestyle factors on healthy cognitive ageing, by simultaneously regressing scores of age-adjusted fluid and crystallized abilities on all five lifestyle factors (Figure 2-8). Doing so allowed me to examine the degree to which each of the five lifestyle factors make unique contributions to cognitive health. For fluid abilities, Education/SES (std $\beta = 0.30$, SE = 0.06, $p < 0.001$), Social Engagement (std $\beta = -0.12$, SE = 0.048, $p = 0.012$), Intellectual Engagement (std $\beta = 0.26$, SE = 0.06, $p < 0.001$) and Physical Health (std $\beta = 0.20$, SE = 0.06, $p = 0.001$), were significant predictors, predicting unique variance in fluid age-residualized abilities, and together explaining 14% of the variance. I found a similar pattern for crystallized abilities, with Education/SES (std $\beta = 0.56$, SE = 0.075, $p < 0.001$), Social Engagement (std $\beta = -0.22$, SE = 0.059, $p < 0.001$), Intellectual Engagement (std $\beta = 0.22$, SE = 0.069, $p < 0.001$) and Physical Health (std $\beta = 0.30$, SE = 0.07, $p < 0.001$) each significant and together explaining 16% of the variance. I did not find evidence that mental health made unique contributions to fluid or crystallized abilities beyond the other lifestyle factors. Notably, in these joint models, the directionality of the effect of Social Engagement changed from

positive to negative, while Physical Health changed from negative to positive. These sign inversions may reflect a true conditional association, or rather a quantitative consequence of the dataset and procedure employed here–I discuss these matters in more detail below.



Lifestyle Factor	*Std. beta*	SE	*p*	Joint R^2
Physical	.20	.057	<.001	**0.14**
Social	-.12	.048	.012	
Intellect.	.26	.057	<.001	
Edu/SES	.30	.062	<.001	
Mental	-.07	.034	.102	

Lifestyle Factor	*Std. beta*	SE	*p*	Joint R^2
Physical	.30	.070	<.001	**0.16**
Social	-.22	.059	<.001	
Intellect.	.22	.069	<.001	
Edu/SES	.56	.075	<.001	
Mental	-0.02	.042	.70	

Figure 2-8: Results of multiple regressions. Four out of five lifestyle factors made unique contributions.

Regarding the specificity of the contribution of lifestyle to crystallized versus fluid abilities, I found that the freely estimated model fit marginally better ($\Delta\chi^2$ (5) = 13.92, p = .016; AIC & BIC (free) = 3580.0; 3639.3; AIC & BIC (constrained) = 3583.9; 3620.4)), suggesting small differences in path estimates. Closer inspection of the parameter estimates showed that this difference was driven almost exclusively by SES, which has a stronger association with crystallized abilities (standardized beta: .56) than with fluid (.30).

2.4.7 Exploratory Analyses

My final set of analyses investigated whether the relationship between lifestyle and cognition differed for different ages or sexes: Testing for an interaction effect therefore assessed whether any of the lifestyle factors increase or decrease in importance for

cognitive health across the lifespan, or between the sexes. To do so, I performed a multi-group model based on a median age split (median = 55 years), which showed that imposing equality constraints across age group did not adversely affect the estimation of the association between lifestyle and cognitive outcomes ($\Delta\chi^2$ (5) = 3.799, p = .58). In other words, a model that assumed the association between lifestyle and cognitive outcomes was the same for those younger and older than 55 did not perform meaningfully worse than a model which assumed them to be different. Finally, I then tested for the presence of sex effects, which again found that the joint model could be equally constrained across sexes without a notable drop in model fit, $\Delta\chi^2$ (10) = 12.96, p = .23. This suggests that the associations between lifestyle and cognitive health are similar across age and for both sexes.

2.5 Discussion

2.5.1 Summary of main findings

In a large lifespan cohort with a broad set of measures, I examined the associations between healthy cognitive ageing and potentially modifiable lifestyle factors. I observed that, in isolation, better physical and mental health, increased social and intellectual engagement and higher levels of education/SES were significantly associated with age-residualized crystallized and fluid cognition (i.e., cognitive abilities higher than those expected for one's age).

Three out of five lifestyle factors showed typical effect sizes, with Education/SES having a strong association, and Social Engagement having a small association (Gignac & Szodorai, 2016). Individual lifestyle domains have previously been correlated with cognitive health in old age and my bivariate results provide further evidence for this relationship. However, as described in the introduction, few studies have investigated combinations of lifestyle factors in a way that allows for statistical inferences regarding their complementary effects (e.g., Clare et al., 2017 who used five separate linear

regressions to investigate the associations between cognition and cognitive and social activity, physical activity, diet, alcohol consumption and smoking). Here, when all lifestyle factors were incorporated into the same model, social and intellectual engagement as well as physical health made independent contributions to fluid and crystallized age-adjusted abilities, above and beyond the effect of education/SES. These relationships were robust across age and sex, and highly similar for fluid and crystallized domains, suggesting general effects, rather than effects specific to cognitive domain. Importantly, social, physical, and intellectual activities are potentially modifiable. Assuming they are causally related to cognitive health (please refer to Chapter 4 for a more detailed discussion on potential causal mechanisms), interventions to increase them may help boost the cognitive abilities which, in turn, may support independent functioning in old age.

In both the linear regressions and the joint models, the strongest associations were those between education/SES and cognitive health. This ties in well with the literature: for example, a recent systematic review comprising over 130,000 individuals (Kralj et al., 2018) showed that a positive education/SES and healthy ageing was reported in 20 of the 25 included studies. One possible explanation is the notion of cognitive reserve (see Chapter 1 for a more detailed explanation of this concept), which suggests that education and occupational attainment determine the brain's reserve capabilities (Stern, 2002). Arguably, however, a person's education or socio-economic status are difficult to alter, particularly later in life. My finding that physical health and intellectual and social engagement are associated with cognitive health above and beyond education/SES therefore offers further support for the promise that potentially modifiable activities also contribute to cognitive reserve.

Previous studies have also reported beneficial associations between higher level of social contact and healthy ageing (Gow et al., 2012; Lövdén et al., 2005; Pruchno & Wilson-Genderson, 2015; Small et al., 2012a; Zaslavsky et al., 2014), consistent with the association found here. Although here the best solution is found by positing separate factors for social and intellectual engagement, note they are highly correlated ($r = .78$),

suggesting a large degree of overlap between the social and intellectual components of lifestyles. Recent work finds a similar overlap: Köhncke and colleagues (2016), for example, used the construct of 'leisure activities', which comprised both social and intellectual activities, to examine its impact on longitudinal change in brain structure and cognitive decline in a large, healthy Swedish cohort (Köhncke et al., 2016) . Notably, they observed that greater current leisure activities were associated with slower rates of decline in mental processing speed, and that this effect could be fully explained (mediated) by slower decline of white matter microstructure. Thus, although my study was cross-sectional, it fits in well with an increased focus on the benefits of an active socio-intellectual lifestyle in old age. Similarly, recent longitudinal evidence suggests possible reciprocal effects between subjective (self-assessment) and objective (test performance) measures of memory in old age (Snitz et al., 2015). In other words, individuals who think their memory is poor may avoid (social) situations where their memory is challenged, thus accelerating memory decline. It may be that social engagement functions as a method of frequent low-grade cognitive challenge, which helps support cognitive performance in old age.

One key contribution of this chapter, echoing recent calls (Kremen et al., 2019), is the simultaneous inclusion of multiple lifestyle factors, in order to better understand their relations and independent contributions. Doing so, I show that four of the five lifestyle factors (all except mental health) contribute uniquely to explaining individual differences in cognitive outcomes. Interestingly, two of the path estimates, namely social and physical, changed sign: While they were, as expected, positively associated with outcomes in isolation, the sign of the association changed in the presence of other, collinear predictors. Both substantive and statistical explanations (which are not mutually exclusive) of these patterns are possible, and I outline both below.

Firstly, I found that social activities became negatively associated with cognitive performance. A possible interpretation is that high levels of social activity which are devoid of intellectual activity may be associated with poorer outcomes. For example, social and intellectual activities may tend to co-occur in people (e.g., frequently meeting

with family to play games), but once the intellectual component is accounted for, the remaining types of social activity may actually be detrimental to cognitive ability (e.g., drinking alcohol regularly with friends). Further research using more refined lifestyle measures is needed to address this possibility.

Secondly, in the simple regressions I observed that better physical health was associated with better cognitive outcomes–but this association changed in sign in the full model. The simple association is in line with several other papers, including intervention studies, which have suggested that physical activity reliably reduces the risk of cognitive impairment (Colcombe & Kramer, 2003; Gill et al., 2015; Kramer et al., 2006; Middleton et al., 2010). However, not all studies observe the same pattern – the UK Whitehall II study found no evidence between physical activity and subsequent 27 year cognitive decline (Sabia et al., 2017), and Gow and colleagues found that mid-life intellectual and social activities, but not physical activity, were associated with late-life cognitive health (Gow et al., 2017). Notably, sign reversals need not be counterintuitive. For example, in the same Cam-CAN sample, Fuhrmann et al. observed strong associations such that low diastolic blood pressure (usually associated with lower overall blood pressure) was associated with worse neural health–but only when the model also included systolic blood pressure (Fuhrmann et al., 2019). This pathway thus captured the conditional effect of a large difference between systolic and diastolic blood pressure, known as 'pulse pressure' often associated with (precursors to) diabetes and other medical conditions (Schram et al., 2002). Similarly, there may be indirect conditional pathways which substantively explain the sign inversion.

Alternatively, there are more purely quantitative explanations of these sign flips. It is well-known that high collinearity between predictors (here Intellectual Engagement and Social Engagement r = .61; Physical Health and Education/SES r =.-68) inflates the standard errors of the parameter estimates, which can produce changes in sign of the mean (Goldberger & Goldberger, 1991; Yoo et al., 2014). However, this increase in standard error would normally render tests no longer significant, which is not the case here (and the standard errors for these paths in the full model were not especially large).

More likely is that my findings reflect a type of 'reversal paradox' (Tu et al., 2008). This phenomenon can occur when parts of a causal chain (i.e., both antecedents and consequences) are incorporated in the same model, inducing – especially in observational data with correlated predictors – reversals of path estimates depending on the nature of the predictors included. In this light, it is worth considering the 'reverse causation' hypothesis of Kremen and colleagues: They state that many of the protective effects of individual differences in lifestyle factors (such as greater cognitive and social engagements, and even education) are themselves the consequence of early life differences in cognitive ability (Kremen et al., 2019).

In the absence of direct access to underlying causal mechanisms generating the data, I cannot conclusively say which of the above explanations are most plausible. As such converging lines of evidence from longitudinal studies, interventions and multivariate approaches will be required to understand the true aetiology of these effects. However, it unambiguously demonstrates the importance of simultaneous assessment of multiple lifestyle-cognition associations if one wishes to better understand the complex lifespan process of risk and resilience factors.

The effect of mental health, while significant in univariate analyses, disappeared in the joint models. This null result should be interpreted with caution as NHST does not allow us to infer whether it reflects a "true null" (see Chapter 1). It suggests that the association between mental health (measured, in this paper, as an emergent latent construct that was measured by depression, quality of sleep, alcohol consumption and self-reported health) and cognitive health is either less strong compared to other lifestyle factors, or fully explained by co-occurrence with other lifestyle factors. This finding differs from those of other cross-sectional studies, which found associations between depression and poorer cognitive function in old age (Bunce et al., 2008; Elderkin-Thompson et al., 2007; Reppermund et al., 2011). However, this discrepancy can, in part, be explained by the high degrees of comorbidity between depression and dementia, given that the above studies (unlike the current one) included participants with mild cognitive impairment (MCI) and/or Alzheimer's disease (AD). Indeed, a longitudinal study that employed

latent growth models showed that, when participants with MCI and AD were removed from the models, the association between cognitive health and depression disappeared (Bunce et al., 2012).

I observed no significant difference of the lifestyle-cognition associations for crystallized compared to fluid age-adjusted abilities; both were captured best by models including education/SES, social engagement, and intellectual engagement. I interpret this to suggest that lifestyle is likely to benefit cognition in a global, rather than specific manner. This might have important ramifications for the interpretation of cognitive intervention studies, which often fail to find positive transfer effects. Assessing cognition on latent and global levels, rather than by performance on individual tasks might be – as has been suggested elsewhere (Schmiedek et al., 2010) – a more desirable statistical approach.

2.5.2 Strengths and limitations

A strength of my analyses is the inclusion of an unusually broad and rich set of lifestyle and cognitive variables in a large lifespan cohort. Uniquely, this allows me to directly compare the relative strength of associations of distinct lifestyle factors within the same healthy population.

The most important limitation of this study is that the data investigated here are cross-sectional. For this reason, although my findings align well with other work, I cannot make direct causal inferences regarding the observed associations, as they may be explained by a variety of causal pathways, included omitted third causes. Moreover, as noted above, causality may flow in both directions: better cognitive health may facilitate the desire, as well as capacity, to maintain an active life in old age (Gow et al., 2012). These issues can be addressed to some extent by longitudinal studies, and most directly by interventional studies; issues I discuss further in the subsequent two chapters. However, it may be all but impossible to engage in a true randomized intervention study of factors as integral to individuals as education, social and intellectual engagement. As

such, large observational studies relying on powerful multivariate methodology may offer an imperfect, but nonetheless valuable, insight into which lifestyle factors are most likely to have beneficial protective effects in ageing, and therefore provide candidate factors which might be more amenable to intervention studies (as well as advising what other factors should be controlled for in such studies). Moreover, I only examined relationships between current activities and current cognitive abilities: it is possible that many years are needed before lifestyle changes affect cognitive abilities. For example, one's current lifestyle activities in old-age may be of little value if similar beneficial activities were not conducted earlier in life, consistent with previous Cam-CAN findings using retrospective questionnaires, where people's activity scores in their current, old age did not make a unique contribution above the same activity scores reported from their previous, mid-life period (Chan et al., 2018). Further work is needed to more precisely reveal the temporal development of the beneficial effects of lifestyle engagement on cognitive abilities.

Methodologically, my approach comes with strengths and limitations. The use of exploratory structural equation modelling (ESEM) allowed me to categorize the observed variables in a mainly data-driven fashion–an approach that has the potential to decrease researchers' subjectivity and selection bias and improve statistical power. However, some loadings of the data-driven lifestyle factors may strike some as counterintuitive. Relatedly, by grouping lifestyle variables into factors, I decrease the specificity of associations of individual variables, and render the hypothetical translation to intervention targets (i.e., to encourage the increase of purportedly beneficial activities) less straightforward. This reflects a general issue, namely that the assessment of lifestyle-cognition associations warrants a trade-off between generalizability and reduction of measurement error (using latent variables) versus specificity and ease of interpretation (using observed variables). The latter approach has led researchers to conclude, for instance, that knitting, doing odd jobs and gardening all reduce the risk of dementia (Fabrigoule et al., 1995). However, a defence of latent lifestyle factors would posit that such activities are better seen as reflecting a class of

activities with similar purported beneficial effects. If there is causal efficacy to, say, knitting, then a coherent causal account would likely posit that activities with similar features (subjective enjoyment, social engagement) would lead to similar beneficial accounts. This line of reasoning is implicitly present in intervention studies that focus on e.g. 'physical activity', 'cardiovascular training' or 'coordination training' (rather than 'walking' or 'using a fitness ball'; e.g., Voelcker-Rehage et al., 2011). Additionally, even with individual variables, the notion of modifiability of lifestyle factors is not entirely straightforward, since the behaviours and personality characteristics that are amenable to intervention or modification, and the circumstances that enable alterations, have yet to be established. Factors like personality, mood, people's perception of their abilities, as well as more external limitations including mobility and financial security, are all likely to affect the extent to which people alter the various aspects of their lives. Theory- or prediction-based approaches, such as mixture models or decision-tree based methods (Brandmaier et al., 2016), might provide useful tools to explore these open questions.

Next, although several indicators of model fit are in the acceptable or good range, the CFI is lower than ideal. As the CFI is an index of comparative fit compared to the null model, a lower CFI often occurs for larger measurement models with moderate to low factor loadings. Although several of my factor loadings are strong (e.g., social outings on the social factor) others are lower (e.g., alcohol consumption on mental health). This is likely a consequence of reporting the best fitting exploratory model, which, in a large lifespan observational sample such as Cam-CAN, is likely to group together variables with only moderately strong relations to each other. In contrast, much more well-established measurement models, refined over multiple cohorts, tend to lead to the selection of only indicators with (very) high loadings. As my goal here is explicitly a descriptive, exploratory factor analysis to reduce a rich sample of indicators to a tractable number of lifestyle factors, including only indicators with high factor loadings would not be appropriate, both for reasons of generalizability (modifying the factor structure purely for reasons of fit) and principle (I wish to convey the full richness of

the data including factor loadings and relationships that perhaps don't fit pre-existing groupings). More importantly, the regressions (both the simultaneous and individual) show moderate to strong effects, suggesting that despite a subset of relatively weak loadings, the factor scores are separable and predictive of external outcomes. As such, I prefer the model as is, with several model fit indices that are good but with a less than optimal CFI, rather than modifying the model to simply achieve a better fit. This reasoning is also in line with my objective to use a data-, as opposed to researcher-driven categorization of variable: While an advantage to modifying the measurement model might be (slightly) better model fit, I believe that the advantages of the data-driven approach (i.e., increased objectivity and greater ease of replicability with other datasets and variables) outweigh these concerns.

Finally, because Cam-CAN represents a sample of healthy adults from a specific region in the UK (City of Cambridge), the generalizability of my findings to other populations remains to be investigated by future research (see General Discussion in Chapter 5 for more detail).

2.6 Conclusion

In conclusion, this chapter's findings suggest that lifestyle variables can be grouped into distinct but correlated factors. Moreover, these factors vary in the strength of their associations with cognitive health, and make specific, complementary contributions in explaining individual age-related differences. Specifically, I found that education/SES, physical health, and social and intellectual engagement, are each simultaneously associated with higher age-adjusted cognitive abilities across the adult lifespan, and that these associations are similar in magnitude and direction for two broad cognitive domains (fluid and crystallized). Mental health, although associated when tested with better cognitive outcomes in isolation, did not make unique contributions above the other three lifestyle factors. Because many of the activities included in my models are,

in principle, modifiable, my findings have encouraging implications for individuals and public health initiatives alike.

Chapter 3: A morphometric double dissociation: cortical thickness is more related to ageing; surface area is more related to cognition

3.1 Chapter Summary

The *thickness* and *surface area* of cortex are genetically distinct aspects of brain structure and may be affected differently by age. However, their potential to differentially predict age and cognitive abilities has been largely overlooked, likely because they are typically aggregated into the commonly used measure of *volume*. In a large sample of healthy adults (N=647, aged 18-88), I investigated the brain-age and brain-cognition relationships of thickness, surface area, and volume, plus five additional morphological shape metrics. Cortical thickness was the metric most strongly associated with age cross-sectionally, as well as exhibiting the steepest longitudinal change over time (subsample N=261, aged 25-84). In contrast, surface area was the best single predictor of age-residualized cognitive abilities (fluid intelligence), and changes in surface area were most strongly associated with cognitive change over time. These findings were replicated in an independent dataset (N=1345, aged 18-93). This chapter's results suggest that cortical thickness and surface area make complementary contributions to the age-brain-cognition triangle and highlight the importance of considering these volumetric components separately.

The work in this chapter is currently under review and has been published as preprint online (Borgeest et al., 2021)**.**

3.2 Introduction

As the human brain ages, it undergoes a pronounced structural transformation. Even in the absence of neuropathology, overall brain volume shrinks – from age six onwards into old age (Bethlehem et al., 2021). This volume decline is associated with various physiological changes, including grey-matter reductions caused largely by the regression of dendrites (see Dickstein et al., 2007 for a review), and white-matter reductions stemming from axon demyelination (Fotenos et al., 2005; Gunning-Dixon et al., 2009; Raz, 2005; Scheltens et al., 1995). There are also morphological changes, with sulci for example becoming shallower (Burgmans et al., 2011; Jin et al., 2018; Madan, 2021; Peters, 2007) and cortex becoming more curved (Deppe et al., 2014).

Traditionally, studies investigating human brain structure with Magnetic Resonance Imaging (MRI) have relied largely on volumetric or thickness measures (see Oschwald et al., 2020 for a review), which only capture a small proportion of the richness of age-related morphometric changes (Ecker et al., 2010; Im et al., 2008). Indeed, the number of papers that include both the term “aging” and “brain volume” (N=2715 in a PubMed search as of 01/06/2021) or “cortical thickness” (N=597) far exceeds those investigating other aspects of morphology, such as “aging” combined with “surface area” (N=125) or “curvature” (N=23). Even though several authors have pointed out that volume is a product of cortical thickness and surface area (Norbom et al., 2021; Storsve et al., 2014; Walhovd et al., 2016; Winkler et al., 2018), which in turn are two genetically independent aspects of brain structure (Hofer et al., 2020; McKay et al., 2014; Panizzon et al., 2009; van der Meer et al., 2020), the implication that thickness and area may have dissociable causes (e.g., in ageing) and consequences (e.g., for cognition) have rarely been discussed, especially in adult samples. Moreover, additional detailed morphometric shape measures (such as curvature or sulcal depth) may provide further insight into brain development across the adult lifespan and its relationship with cognitive performance.

In this chapter, I explore multiple morphometric measures in two large adult-lifespan cohorts. I show, firstly, that the most pronounced structural changes in the ageing brain are the decrease in apparent cortical thickness (see Walhovd et al., 2017 for the interpreation of MR-derived cortical thickness) and increase in cortical curvature, in line with other studies (Deppe et al., 2014; Hogstrom et al., 2013; Lemaitre et al., 2012). Secondly, I find that incorporating multiple shape measures into a single model outperforms any individual metrics' ability to capture age-related and fluid cognitive differences. This chapter's main contribution, however, lies in providing cross-sectional and longitudinal evidence of a double dissociation in two independent, large-sample cohorts. Specifically, cortical thickness was more strongly associated with age than cortical surface area, while surface area was more strongly associated with cognition (as indexed by fluid intelligence). This pattern was most apparent longitudinally, but we also observed it cross-sectionally after adjusting for age. This double dissociation points to possibly distinct underlying biological processes (discussed below), and supports recent calls to investigate thickness and surface area separately (Winkler et al., 2018) as brain volume (a product of cortical thickness and surface area) likely conflates and therefore masks these differentiable effects.

3.3 Methods

3.3.1 Initial Cohort: Cam-CAN

3.3.1.1 Participants

Participants were drawn from the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study, which has been described in more detail elsewhere (Shafto et al., 2014; Taylor et al., 2017). 708 healthy adults (359 women, 349 men) from the larger cohort were scanned, with approximately 100 people in each decade (age range 18-88,

Mean=53.4, Standard Deviation (sd) = 18.62). We used calendar age (years) as a measure of participants' age. Cognitive ability was measured using the Cattell Culture Fair test of fluid intelligence (Cattell, 1971). For an age-independent measure of cognition, I calculated age-residualized fluid intelligence scores by regressing the Cattell raw scores on age. As explained in the previous Chapter, residuals adjust for age-expected declines, allowing, for example, an 80-year-old person with a relatively low absolute score to be considered cognitively healthier than a younger individual with a higher score.

A subset of participants (N=261) was scanned twice, with an average interval between the first and the second scan of 1.33 years (sd = 0.66). Additionally, a (partially separate) subset of participants (N=233) completed the Cattell test twice with an average interval between the two cognitive tests of 6.0 years (sd = 0.67). Two waves of both brain and cognitive data were available for 115 participants.

3.3.1.2 Imaging data acquisition and pre-processing

T1- and T2-weighted 1 mm isotropic magnetic resonance imaging scans were available for 647 participants (Taylor et al., 2017). To ensure the quality of the image segmentations, I adapted a recently developed supervised learning tool (Klapwijk et al., 2019), which led me to exclude six participants due to low-quality segmentations. Thus, quality control process is described further in the supplementary materials which are available on the Open Science Framework (https://osf.io/n6b4j/). In order to investigate (cross-sectional) brain morphology in as much detail as possible, I examined a total of eight brain metrics: in addition to three FreeSurfer-derived measures of cortical volume, thickness and surface area (derived from a standard FreeSurfer recon-all pipeline), I examined total grey matter (TGM) derived from SPM 12 (voxel-based morphometry which includes sub-cortical grey-matter too, while FreeSurfer includes only cortical estimates; Ashburner et al., 2021) and four additional morphological measures: from Mindboggle (see Klein et al. 2017 for more detail) I derived sulcal depth, curvature and "thickinthehead" (a recently developed cortical thickness measure that avoids FreeSurfer's reconstruction-based limitations); and from the calcFD toolbox

(Madan & Kensinger, 2016) I calculated fractal dimensionality as a measure of cortical complexity. First, to better understand the relationship between the eight brain structure metrics I calculated a correlation matrix and ran a Principal Component Analysis. To extract reliable brain structure estimates from the longitudinal subsample, images were automatically processed with FreeSurfer's longitudinal stream (Reuter et al., 2012). This yielded co-registered measures of volume, cortical thickness, and surface area for the two waves. Note that I did not explore the other morphological metrics longitudinally because the Mindboggle and calcFD pipeline are not currently optimised for longitudinal data (see discussion). Brain regions were defined according to the Desikan-Killiany-Tourville (DKT) protocol, which yields 62 brain regions (Klein & Tourville, 2012).

3.3.1.3 Cross-sectional analyses

3.3.1.3.1 Whole brain analyses

All analyses were carried out using R (R Core Team, 2013), and the code used for this Chapter is available on the Open Science Framework (https://osf.io/n6b4j/).

First, I calculated whole brain as well as regional correlations between each metric and age, fluid intelligence and age-residualized fluid intelligence. Regional correlations were FDR corrected at alpha = 0.05. I next examined whether the different metrics of brain structure provided unique and complementary information about age and cognitive ability. To do so, I ran frequentist path models in which i) all eight metrics and ii) cortical thickness and surface area only predicted either age, fluid intelligence or age-adjusted fluid intelligence. Path analysis is an extension of multiple linear regressions, allowing researchers to assess the relationships between the predictor variables rather than having several independent variables predict one dependent variable (Streiner, 2005).

3.3.1.3.2 Regional analyses

In Cam-CAN, after looking at whole brain correlations between the eight metrics and age, fluid intelligence and age-residualized fluid intelligence, I investigated these three correlations for each brain region. Regions were assigned 62 labels following the Desikan-Killiany-Tourville (DKT) protocol in the Mindboggle pipeline (Klein et al., 2018), and then averaged across both hemispheres. All correlations were FDR-corrected at alpha = 0.05.

3.3.1.4 Longitudinal analyses

To assess neural and fluid intelligence change between time point 1 and time point 2, I fit a series of longitudinal structural equation models for each longitudinal FreeSurfer metric (whole brain volume, thickness and surface area) and fluid intelligence. Before assessing cognitive change, we also tested for longitudinal measurement invariance (Widaman et al., 2010). Additionally, as the second Cattell test was completed online by approximately half of the participants, versus pencil and paper by the other half, I investigated whether these two groups differed in their measurement properties by assessing metric invariance (constraining factor loadings) and scalar invariance (constraining intercepts).

To understand whether cognitive change was correlated with morphometric change, and if so, whether this relationship differed for the different cortical metrics, I extracted and estimated the rates of cognitive and brain structure change in a series of second order latent change score models (Ferrer et al., 2008; Ferrer & McArdle, 2010; McArdle & Hamagami, 2001; McArdle & Nesselroade, 2003). Second order latent change score models (2LCSM) first estimate latent factors at each time point, and then estimate latent change over time. Steiger's Z-Tests were performed to assess whether the change-change relationships differed significantly between the different metrics (Steiger, 1980). Given that properties of the data, obtaining latent cognitive scores was not possible in the replication sample (see below), so we also ran the models with observed variables

only within Cam-CAN to ensure maximal comparability between the two sets of analyses. We ran models on participants with at least one cognitive score (N=362) using full information maximum likelihood (FIML, which assumes data are missing-at-random, Enders & Mansolf, 2018, and enables robust standard errors to account for missingness).

3.3.2 Replication Cohort: LCBC

To assess the robustness of our results, we investigated whether our core findings replicated in a second, independent dataset. To this end, we analysed data from the Centre for Lifespan Changes in Brain and Cognition at the University of Oslo (LCBC; https://www.oslobrains.no/), which is part of the European Lifebrain project (Walhovd et al., 2018) together with Cam-CAN and other publicly available datasets. The LCBC data consist of a collection of studies, which have been described elsewhere (Walhovd et al., 2016). Briefly, our analyses included 1236 adults aged 18-93 years (median = 37, sd = 20.64). We used WASI Matrix (raw scores) as our measure of fluid intelligence because it is most similar to the Cattell task assessed in Cam-CAN. FreeSurfer-derived cortical thickness, volume and surface area served as our morphological measures (for details on cross-sectional and longitudinal image acquisition and pre-processing see (Walhovd et al., 2016)). At least two waves of cognitive and/or neural data were available for 389 participants. Where participants had more than two waves, we selected their first and last time point, maximizing the interval between waves as well as the data similarity between samples. This allowed us to include the largest possible number of participants in our longitudinal analyses while maintaining two-wave models comparable to those described in Cam-CAN. The mean interval between the two waves so defined was 5.18 years (min = 0.73, max = 10.0, sd = 2.59 years).

Our analysis pipeline mirrored that described above: cross-sectionally, whole brain correlations were followed by frequentist path models. Longitudinally, LCSMs assessed cognitive and neural change separately; and we ran a series 2LCSMs to investigate the

relationship between cognitive change and neural change. The FIML models included 722 participants. Note that it was not possible to derive latent cognitive factor scores for the longitudinal models as individual WASI scores were not available, so the LCBC longitudinal models used observed cognitive variables (but were otherwise identical to Cam-CAN models). The LCSM data and analyses are described in more detail in the supplementary material.

3.4 Results

3.4.1 Descriptive statistics

Tables 3-1 and 3-2 summarize the descriptive statistics of the Cam-CAN and LCBC cohorts.

Cam-CAN	N	Mean	SD	Median	Min	Max	Skew	Kurtosis
Age	641	54.04	18.56	54.00	18.00	88.00	-0.05	-1.15
GM Volume	641	7114.64	899.87	4939.80	930.07	7034.19	0.36	-0.17
Surface Area	641	3177.87	320.97	3157.00	2442.25	4445.75	0.38	0.09
Thickness	641	2.66	0.12	2.67	2.19	2.98	-0.50	0.87
Cattell	622	31.05	6.74	33.00	11.00	44.00	-0.56	-0.16

Table 3-1: Descriptive statistics for Cam-CAN data. Cortical grey matter (GM) volume, area and thickness estimated from FreeSurfer.

LCBC	N	Mean	SD	Median	Min	Max	Skew	Kurtosis
Age	1236	41.55	20.32	31.95	18.0	93.35	0.71	-1.02
GM Volume	1188	7453.09	853.41	5092.91	890.39	7441.81	0.14	-0.41
Surface Area	1188	2630.76	246.85	2618.33	1859.63	3300.62	0.15	-0.32
Thickness	1188	2.60	0.11	2.61	2.09	2.91	-0.38	-0.03
WASI Matrix	1234	27.67	4.64	20.00	6.00	35.00	-.69	4.06

Table 3-2: Descriptive statistics for LCBC data. Cortical grey matter (GM) volume, area and thickness estimated from FreeSurfer.

A correlation matrix of the eight brain structure measures (depicted in Figure 3-1) showed that the strongest correlations were those between FreeSurfer's grey-matter volume (which includes only cortex) and SPM's TGM (which includes subcortical areas, too) (r=.96), surface area and TGM (r=.93), and grey-matter volume and fractal dimensionality (r=.91). Curvature and Thickinthehead are also strongly correlated (r= -.89). The weakest correlations were between depth and curvature (r=-.04), depth and thickness (r = 0.06) and depth and Thickinthehead (r = .13). Surface area and thickness, discussed in more detail below, were correlated only r=.16.

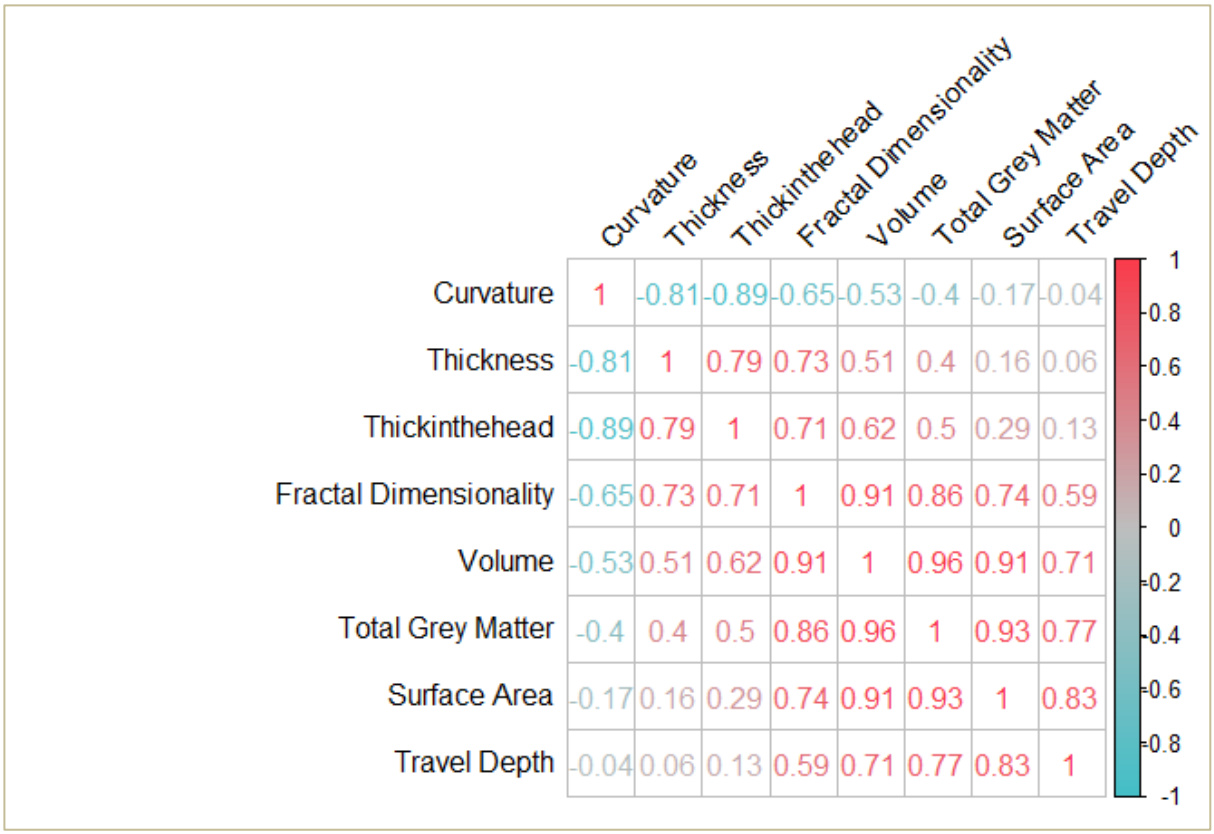
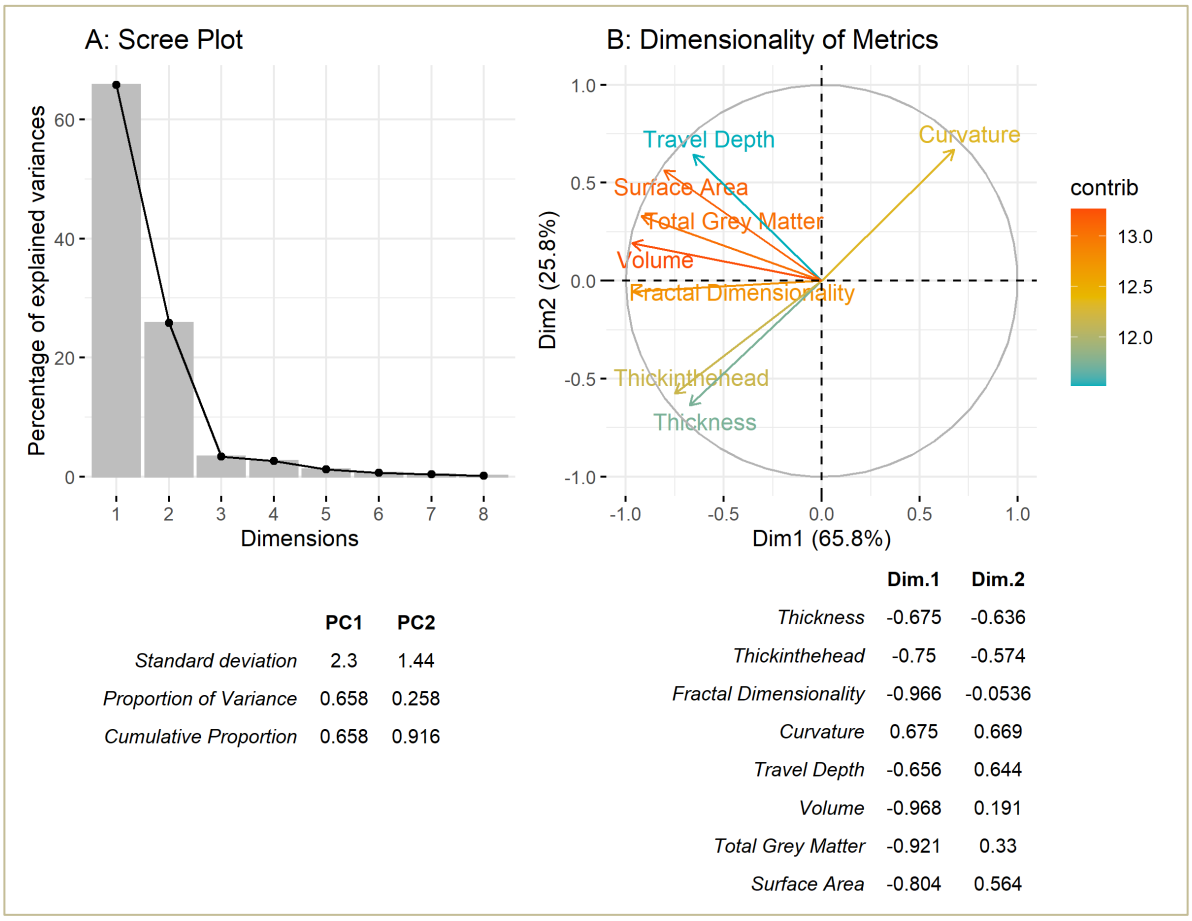


Figure 3-1: Correlation matrix of the eight brain structure metrics. Total Grey Matter (TGM) was derived using SPM 12 (Ashburner et al., 2021). Fractal Dimensionality was calculated using the (Freesurfer-

based) calcFD toolbox (Madan & Kensinger, 2016) . Surface Area, Thickness and Grey Matter Volume stem from FreeSurfer (Fischl, 2012), the remaining three metrics (travel depth, Thickinthehead, and curvature) from (FreeSurfer-based)Mindboggle (Klein et al., 2017).

The Principal Component Analysis (shown in Figure 3-2) suggested that the eight metrics were best captured by two components, which captured 66 percent and 26 percent of the variance, respectively. The first loaded approximately equally on all metrics, whereas the second loaded differentially on thickness metrics versus surface area, as well as curvature and travel depth.



	PC1	PC2
Standard deviation	2.3	1.44
Proportion of Variance	0.658	0.258
Cumulative Proportion	0.658	0.916

	Dim.1	Dim.2
Thickness	-0.675	-0.636
Thickinthehead	-0.75	-0.574
Fractal Dimensionality	-0.966	-0.0536
Curvature	0.675	0.669
Travel Depth	-0.656	0.644
Volume	-0.968	0.191
Total Grey Matter	-0.921	0.33
Surface Area	-0.804	0.564

Figure 3-2: Results of the Principal Component Analysis

3.4.2 Cross-sectional results

3.4.2.1 Whole brain results

I first calculated whole brain as well as regional correlations between each metric and age, cognitive abilities (as indexed by fluid intelligence) and age-residualized cognitive abilities. Residualized cognitive scores allow one to separate concurrent age-related decline in cognitive ability, thus providing an age-independent measure of cognition. Thickinthehead, which is a measure of cortical thickness from the Mindboggle software, showed the strongest age correlation (r = -.83). This was followed by curvature (r = +.77), fractal dimensionality (a measure of cortical complexity; r = -.65) and FreeSurfer's grey-matter volume (r = -..62), as shown in Table 3-3 and plotted in Figure 3-3 and Figure 3-4. Compared to the other metrics, surface area exhibited the weakest age relationship (r = -.36). This order was reversed for age-residualized cognition. Here, surface area was the strongest predictor (r = +.21), while the two thickness metrics and curvature did not show significant brain-cognition correlations after adjusting for age. The two volume measures (FreeSurfer's cortical volume, plus SPM's cortical + subcortical volume) predicted both age and age-residualized fluid-intelligence reasonably well (r_{age} ~ -.58 and r_{cog} ~ +.20), as would be expected since they are proportional to the product of cortical thickness and surface area. Fractal dimensionality was also a good predictor of both age and age-residualized cognition (r_{age}= -.65, r_{cog} = +.19).

	Age		Fluid Intelligence		Age-residualized Fluid Intelligence	
Metric	Pearson's r	*P*	Pearson's r	*P*	Pearson's r	*P*
Cortical Volume (FS)	-.62	<.001	+.56	<.001	+.20	<.001
Cortical Thickness (FS)	-.60	<.001	+.42	<.001	+.04	.33
Surface Area (FS)	-.36	<.001	+.39	<.001	+.21	<.001
Thickinthehead (MB)	-.83	<.001	+.59	<.001	+.04	.34
Curvature (MB)	+.77	<.001	-.56	<.001	-.034	.39
Sulcal Depth (MB)	-.38	<.001	+.51	<.001	+.07	.06
GM Volume (SPM)	-.54	<.001	+.51	<.001	+.20	<.001
Fractal Dimensionality	-.65	<.001	+.56	<.001	+.19	<.001

Table 3-3: correlations between brain structure and age, fluid intelligence and age-residualized fluid intelligence. GM = grey-matter. FS = FreeSurfer. SPM = Statistical Parametric Mapping. MB = Mindboggle.

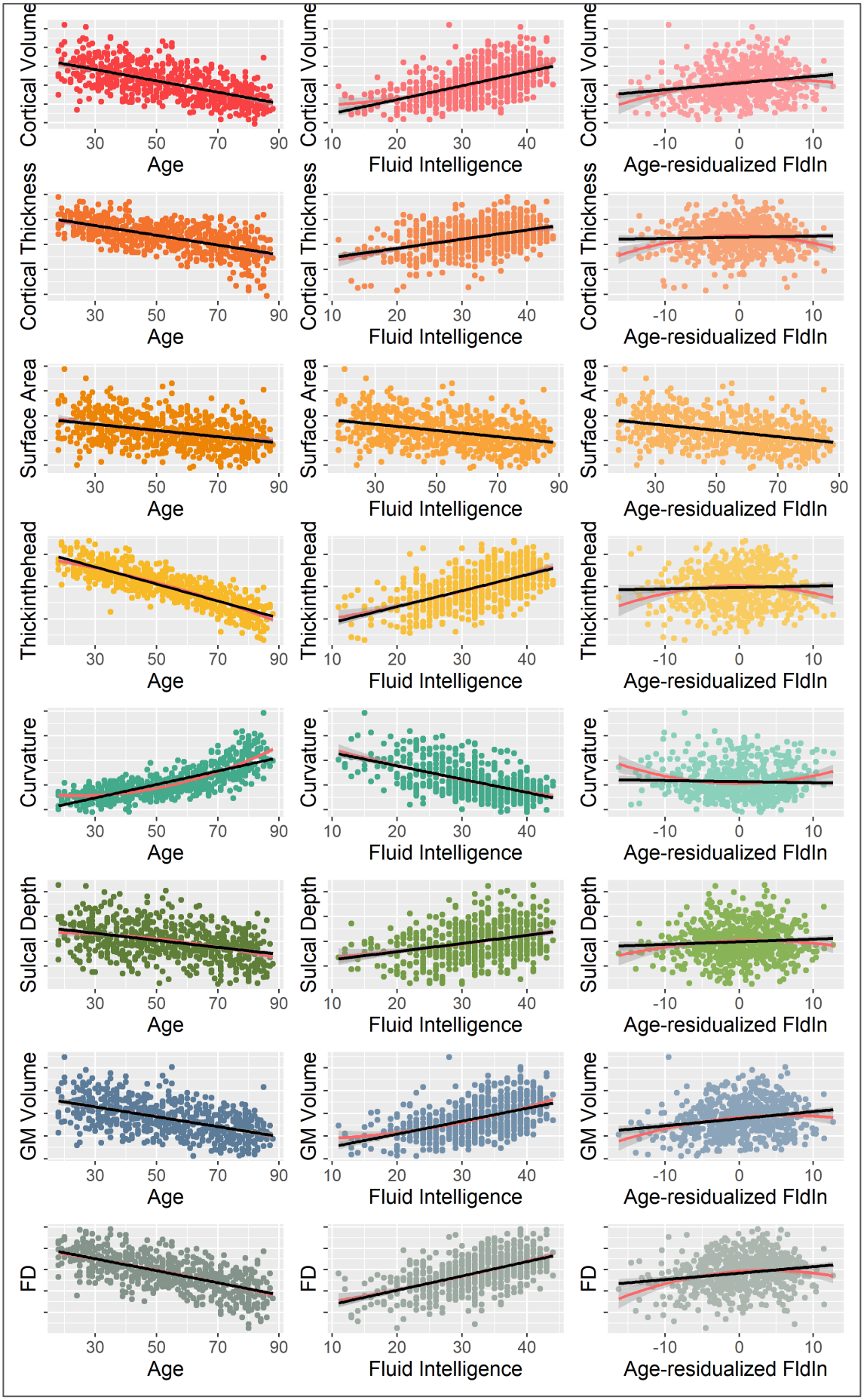

Cortical Volume
Cortical Thickness
Surface Area
Thickinthehead
Curvature
Sulcal Depth
GM Volume
FD
Age
Fluid Intelligence
Age-residualized FldIn
30
50
70
90
10
20
30
40
-10
0
10

Figure 3-3: brain structure-age, -fluid intelligence and -age-residualized fluid intelligence scatterplots of all eight metrics. Black lines show linear fit, red lines show quadratic fit. The metric exhibiting the strongest age relationship is Thickinthehead (a measure of cortical thickness), while surface area is most strongly related to age-residualized cognitive abilities. GM = Grey Matter, FD = Fractal Dimensionality; FldIn = fluid intelligence.

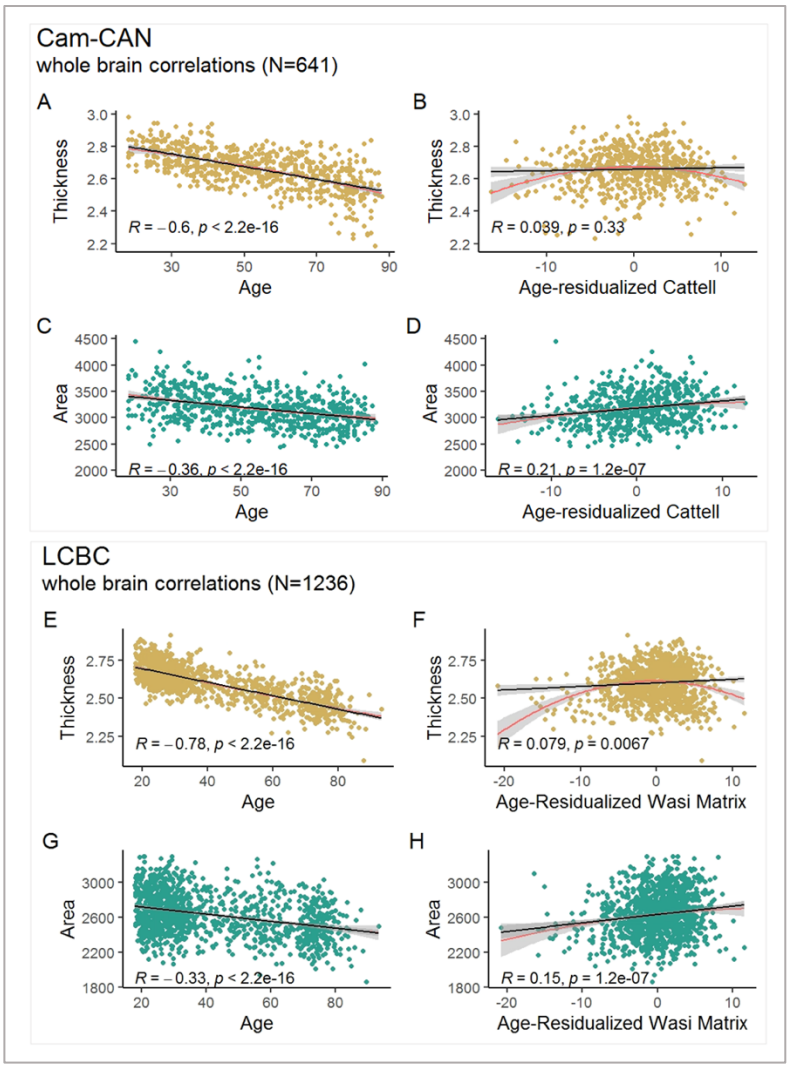


Figure 3-4: Cross-sectional whole brain correlations in Cam-CAN (A-D) and LCBC (E-H). While thickness is associated with age (not age-residualized cognition), surface area captures age-residualized cognition well (and age comparatively poorly).

Next, I estimated a series of path models to assess the relationship between brain structure and age, fluid intelligence and age-residualized fluid intelligence when i)

surface area and cortical thickness, and ii) all eight metrics, are included in the same model. These six models are depicted in Figure 3-5.

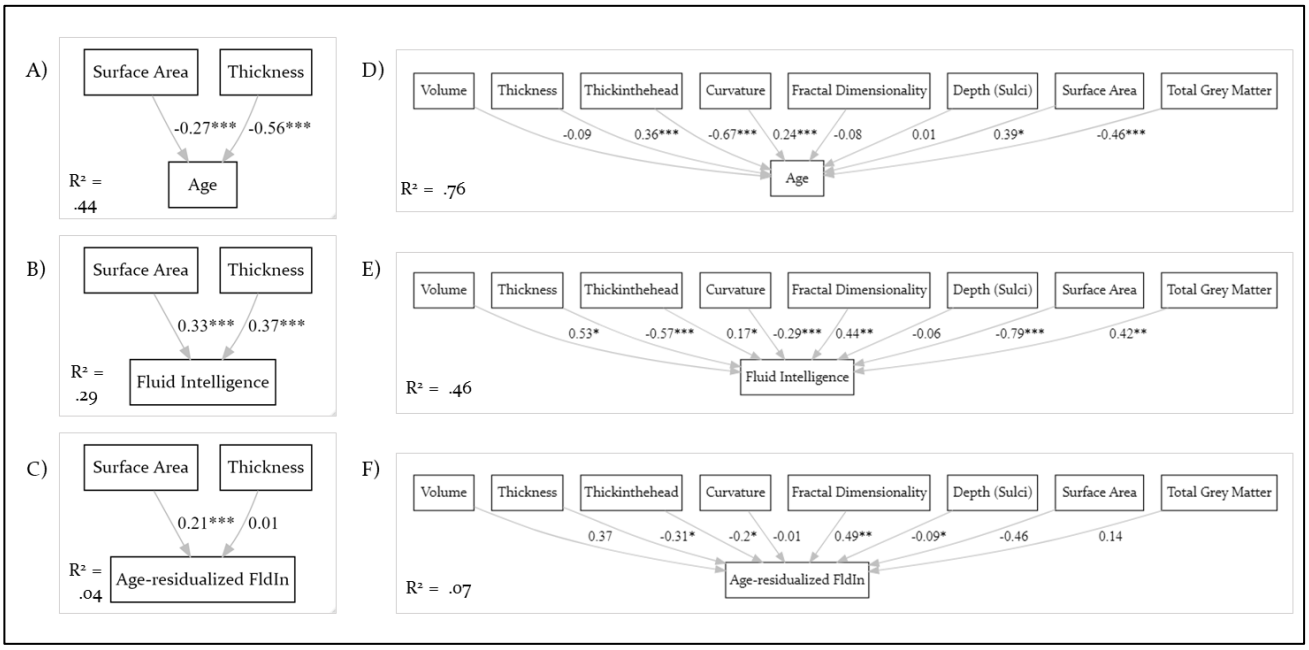


Figure 3-5; Cam-CAN path model results. Models A-C: Both surface area and thickness are significantly associated with age and fluid intelligence, while age-residualized fluid intelligence is captured by surface area only. Full models (D-F).

Age and fluid intelligence were best captured by surface area and cortical thickness, while age-residualized fluid intelligence was associated only with surface area (see Figure 3-5, A-C). For the full models (Figure 3-5, D-F), the total variance explained was 76, 46 and 7 percent for age, fluid intelligence and age-residualized fluid intelligence, respectively – almost double the variance explained by thickness and area alone. Moreover, the fact that multiple morphometric measures provided partially complementary information about the outcome highlights the potential usefulness in assessing various morphological shape measures when investigating the ageing brain and cognitive abilities. For age, the best model included Thickness, Thickinthehead, Curvature, TGM and Surface Area. Fluid intelligence was best captured by Thickinthehead, Curvature, TGM, Surface Area, Thickness, Volume and FD. Finally, the

best model for age-residualized fluid intelligence included FD, Thickness, Depth and Thickinthehead. Interestingly, when FD was not included in the models, the best model for age-residualized fluid intelligence included surface area only, suggesting that surface area and FD capture similar variance.

3.4.2.2 Regional results

Table 3-4 depicts each metric's top-three regions with the strongest age-, raw-cognition and age-residualized cognition effects. I excluded depth from this table because depth is a measure of sulci, not brain regions. Note that data for the entorhinal, superior temporal gyrus and temporal pole were only available for Thickinthehead and Volume. The full table of FDR-corrected correlations between 32 brain regions (averaged across both hemispheres) and age, age-residualized cognitive abilities and raw cognitive abilities can be found in Appendix A.

Overall, there were some interesting similarities and differences between the measures. The precentral gyrus was the region with the strongest age effects in five out of seven metrics: curvature (r=0.74), thickness (r=-0.66), Thickinthehead (r=-0.87), volume (r=-0.71), TGM (r=0.-66). The precentral gyrus, along with the superior temporal and superior frontal gyrus, was also strongly associated with age-residualized cognitive abilities across several imaging metrics. For age-residualized cognition, the lateral orbitofrontal cortex was a significant predictor for volume (r=0.22), surface area (0.19). and TGM (r=0.22).

Metric	Regions with strongest age effect	r	Regions with strongest raw-cognition effect	r	Regions w strongest age-resid. cognition effect	r
Fractal Dimensionality	Inferior parietal	-0.56	Rostral middle frontal	0.43	Isthmus cingulate	0.18
	Rostral middle frontal	-0.55	Precentral	0.43	Transverse temporal	0.16
	Superior frontal	-0.54	Supramarginal	0.43	Median orbitofrontal	0.15
Curvature	Precentral	-0.74	Superior temporal	-0.53	Pericalcarine	0.069
	Supramarginal	-0.74	Supramarginal	-0.51	Corpus callosum	0.068
	Superior frontal	-0.71	Precentral	-0.50	Postcentral	0.049
Thickness	Precentral	-0.66	Precentral	0.50	Insula	0.11
	Superior frontal	-0.65	Superior Temporal	0.46	Lingual	0.092
	Supramarginal	-0.65	Superior frontal	0.44	Superior temporal	0.078
Thickinthehead	Precentral	-0.87	Precentral	0.61	Insula	0.081
	Middle temporal	-0.82	Middle temporal	0.58	Temporal pole	0.079
	Supramarginal	-0.81	Pars opercularis	0.58	Cuneus	0.072
Volume	Precentral	-0.71	Precentral	0.60	Insula	0.22
	Superior frontal	-0.61	Superior frontal	0.53	Lateral orbitofrontal	0.22
	Superior temporal	-0.61	Superior temporal	0.53	Pars orbitalis	0.19
TGM	Precentral	-0.66	Precentral	0.55	Rostral middle frontal	0.22
	Pars opercularis	-0.62	Superior temporal	0.54	Lateral orbitofrontal	0.22
	Superior parietal	-0.61	Rostral middle frontal	0.54	Pars orbitalis	0.21
Surface area	Middle temporal	-0.40	Lateral orbitofrontal	0.40	Posterior cingulate	0.20
	Rostral middle frontal	-0.40	Middle temporal	0.39	Lateral orbitofrontal	0.20
	Lateral orbitofrontal	-0.39	Posterior cingulate	0.38	Superior frontal	0.19

Table 3-4: Regional correlations. Shows the top three regional correlations for age, raw fluid abilities and age-residualized fluid abilities.

The regional analyses further supported the full path models' findings that it can be useful to assess multiple shape measures when investigating the ageing brain: for instance, while volume-age effects were most pronounced in the frontal regions, area-age effects were strongest in the temporal lobes. Importantly, these findings could reflect random noise arising from selecting the "highest statistic" to compare across regions rather than constructing a static across regions. Still, it is plausible that the focus on frontal brain regions in the brain and cognitive ageing literature (Greenwood, 2000; Jung & Haier, 2007) is informed in part by the field's traditional focus on brain volume, and that other aspects of brain structure could point to more underappreciated regional effects.

Finally, the regional results were in line with the morphometric dichotomy found in the whole brain analyses. For cortical thickness, all 32 brain regions (averaged across the hemispheres) were significantly correlated with age (all correlations were FDR corrected at alpha = 0.05), while not a single region predicted age-residualized fluid intelligence. In contrast, for surface area, *all* regions were significantly associated with age-residualized fluid intelligence. While regional surface area also correlated with age, the correlations were substantially weaker than the brain-age correlations for cortical thickness. This dichotomy is shown in Figure 3-6.

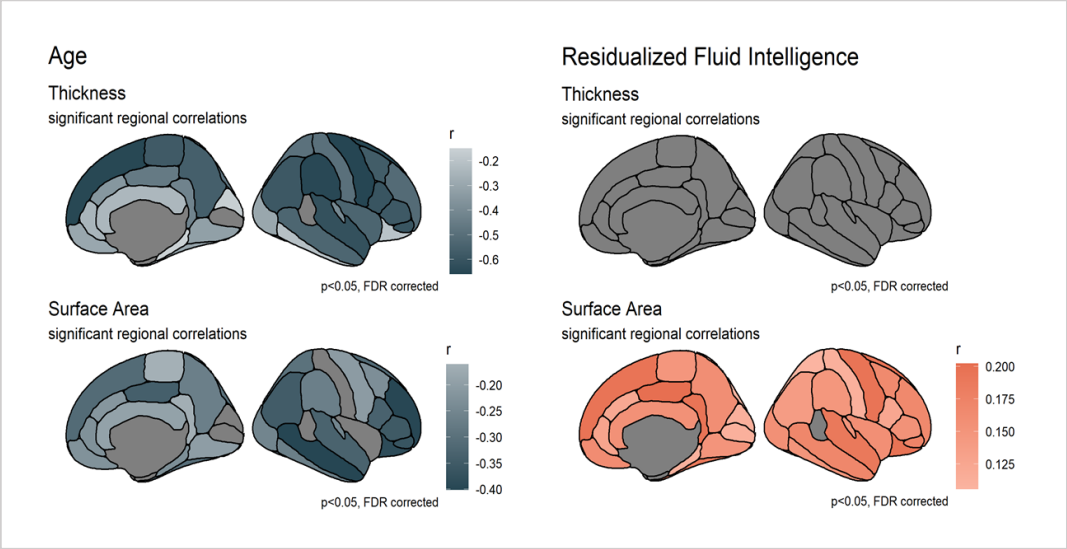


Figure 3-6: Significant regional age- and age-residualized fluid intelligence correlations. Correlations are FDR corrected at alpha = 0.05.

3.4.3 Longitudinal results

Although cross-sectional analyses offer an interesting insight into age-related cognitive and morphometric *differences*, longitudinal data are needed to truly assess how brain and cognition *change* (Oschwald et al., 2020). Doing so, I found that the change-change relationship between surface area and cognition was significantly stronger than the change-change relationship between volume and cognition, as well as that between thickness and cognition. The longitudinal data is plotted in Figure 3-7.

After establishing metric and scalar invariance (described in Appendix A), I used Latent Change Score Models (LCSM) to examine morphometric and cognitive change over time. The cognitive LCSM revealed significant change in cognition over time, as well as significant variability in the rate of change (Table 3-5, variances). The effect size of change of fluid intelligence was -0.09 (Cohen's D, computed by dividing the mean change by the SD at time 1). The three brain-structure LCSMs also showed evidence of change over time (Table 3-5, intercepts) and of significant variability in the rate of

change (Table 3-5, variances). Surface area, volume and thickness all decreased between the first and the second scan. Surface area had the smallest effect size (Cohen's D = -0.02), with cortical thickness and volume exhibiting larger effects (Cohen's D of -0.12 and -0.11, respectively).

		Latent change score model results Cam-CAN					
		Estimate	*SE*	z-value	*p*	Std.all	Effect size
Cattell	Intercepts	-0.633	0.289	-2.192	0.028	-0.145	-0.09
	Variances	19.059	2.808	6.787	<.0001	1.000	
Thickness	Intercepts	-0.012	0.002	-6.234	<.0001	-0.386	-0.12
	Variances	0.001	0.000	7.229	<.0001	1.000	
Surface Area	Intercepts	-5.680	1.632	-3.481	<.0001	-0.215	-0.02
	Variances	695.026	197.495	3.519	<.0001	1.000	
Volume	Intercepts	-50.550	5.887	-8.587	<.0001	-0.530	-0.11
	Variances	9080.25	1057.968	8.587	<.0001	1.000	

Table 3-5: Latent change score model results for change in Cattell, surface area, thickness and volume over time. Effect size is calculated by dividing the mean change by the square root of the variance.

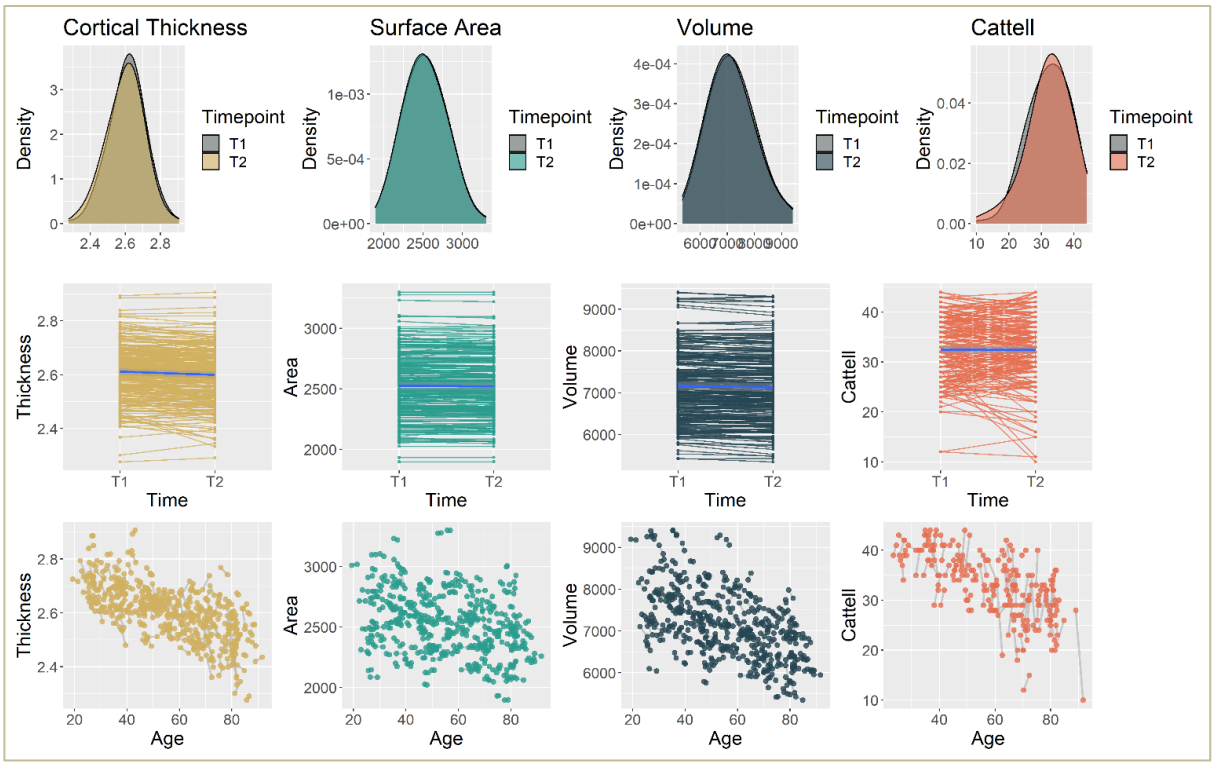


Figure 3-7: In Cam-CAN, cortical thickness, surface area and fluid intelligence declined significantly between time point 1 and time point 2 (average interval between the two time points = 1.33 years).

Next, to investigate the relationship between cognitive change and morphometric change, I fit three second order latent change score models (2LCSM), one for each brain structure metric. I used full information maximum likelihood (FIML, Enders & Mansolf, 2018) with robust standard errors to account for missing data. Results are shown in Table 3-6.

Data	Model	CFI	Change-change correlation (r)	*p*
Cam-CAN	Area – Cognition	0.972	0.23	<0.001
	Thickness – Cognition	0.978	-0.022	0.71
	Volume – Cognition	0.975	0.11	0.068
LCBC	Area – Cognition	0.987	0.35	<0.001
	Thickness – Cognition	0.994	0.21	<0.001
	Volume – Cognition	0.921	0.15	<0.001

Table 3-6: Second order latent change score model results using FIML for missing data. Shows the relationship between change in brain structure (volume, thickness, area) and change in cognition in Cam-CAN and LCBC. In both datasets, change in surface area was most strongly associated with cognitive change.

All three models fit the data well: $CFI_{area} = 0.972$; $CFI_{volume} = 0.975$; $CFI_{thickness} = 0.978$; (further model fit indices can be found in Appendix A). After fitting the models, I extracted and correlated the cognitive rates of change with the brain structural rates of change. Change in surface area showed the largest effect ($r = +.23$, $p < .001$), followed by (non-significantly) volume ($r = -.11$, $p = 0.068$) and cortical thickness ($r = -.022$, $p = 0.71$). The Steiger's-Z tests (Steiger, 1980) in the R package "psych" can directly compare differences in correlation strengths, accounting for the full correlation pattern among variables. Doing so revealed that change in area was significantly more strongly associated with change in cognition than was thickness or volume change (see Table 3-7).

Data	Comparison	r values	N	Z	*p*
Cam-CAN	Thickness / Area	-0.022/0.23	362	3.34	0.001
	Thickness / Volume	-0.022/0.11	362	1.66	0.1
	Volume / Area	0.11/0.23	362	1.77	0.04
LCBC	Thickness / Area	0.21/0.35	722	2.89	0.001
	Thickness / Volume	0.21/0.15	722	1.18	0.24
	Volume / Area	0.15/0.35	722	4.06	0.001

Table 3-7: Steiger's Z Test results. P-value (two-tailed) of <0.05 suggests correlation coefficients are significantly different from each other.

These results suggest that people whose surface area decreased more quickly also showed steeper rates of cognitive decline; an effect not found for thickness or volume.

Note that the models shown above include observed (not latent) variables to ensure maximum comparability between the LCBC and Cam-CAN models (in LCBC, it was not possible to derive latent cognitive scores because only WASI sum scores were available). However, latent variable Cam-CAN models (which we had run initially, before the replication study) show the same pattern, with changes in surface area most strongly associated with changes in cognition (r=0.44, p <0.001). For these models, changes in volume were significantly associated with changes in fluid intelligence (r=0.26, p = <0.001), while this relationship remained insignificant for cortical thickness (r = 0.0047, *p* = 0.94). All longitudinal change score model results are plotted in Appendix A.

3.4.4 Replication results

To examine whether the cross-sectional and longitudinal findings generalise to other cohorts, I next (after finalizing the analyses in Cam-CAN) examined the same associations in an independent sample, the LCBC data. Because of their widespread use

and accessibility, we included the three FreeSurfer-derived metrics (thickness, area, volume) in our replication analyses (Figure 3-8).

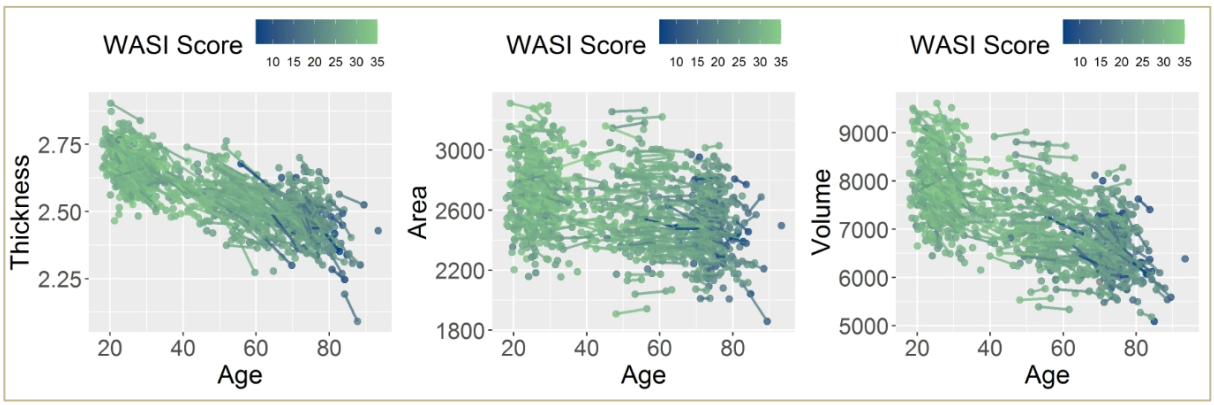


Figure 3-8: The relationship between age, brain structure and cognition in LCBC.

Cross-sectionally, thickness showed the strongest whole brain-age correlation (r= -.78, $p < 0.001$), followed by volume (r = -.64, $p < 0.001$) then surface area (r = -.34, $p < 0.001$). For age-residualized fluid intelligence, thickness had the weakest correlation (r = +.077, $p = 0.009$), followed by surface area (r = +.15, $p = 0.001$) and volume (r = +.16, $p < 0.001$). The surface area and thickness results are plotted Figure 3-4 (E-H) above. As was the case in Cam-CAN, the frequentist path models revealed that the best models to predict age and fluid intelligence were comprised of both surface area and thickness, while age-residualized fluid intelligence was best captured by surface area alone (Figure 3-9).

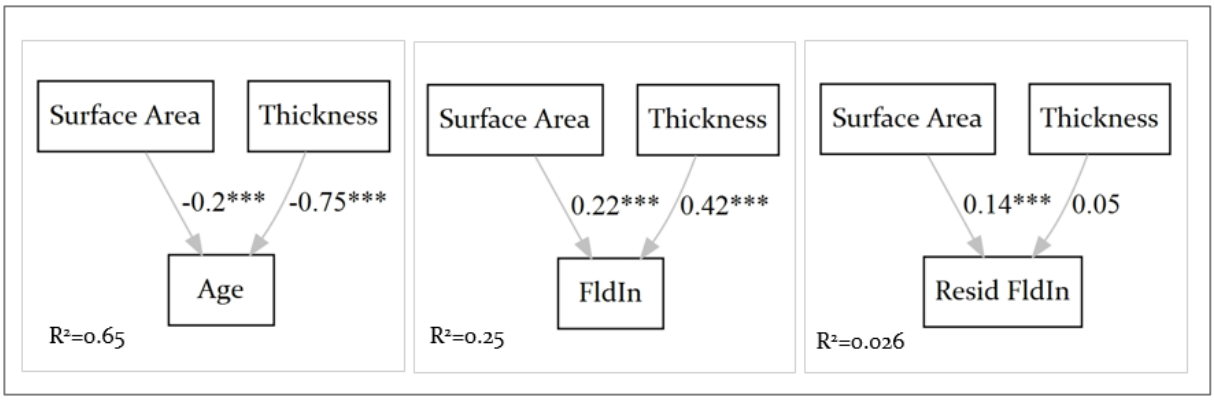


Figure 3-9: LCBC path model results. Both surface area and thickness are significantly associated with age and fluid intelligence, while age-residualized fluid intelligence is captured by surface area only.

Longitudinally, I found evidence of significant change over time for the three-brain metrics (Table 3-8, intercepts), and significant variability over time for the brain metrics and cognition (Table 3-8**Error! Reference source not found.**, variances). A lack of mean cognitive decline can most likely be attributed to test-retest effects, but still allows for investigation of individual differences in change.

		Latent change score model results LCBC					
		Estimate	*SE*	z-value	*p*	Std.all	Effect size
WASI Matrix	Intercepts	-0.247	0.166	-1.488	0.137	-0.078	-0.051
	Variances	10.069	1.246	8.080	<.0001	1.000	
Thickness	Intercepts	-0.039	0.002	-19.815	<.0001	-1.039	-0.340
	Variances	0.001	0.000	12.191	<.0001	1.000	
Surface Area	Intercepts	-14.853	1.935	-7.678	<.0001	-0.412	-0.059
	Variances	1301.028	187.252	20.513	<.0001	1.000	
Volume	Intercepts	-130.745	8.885	-14.716	<.0001	-0.806	-0.15
	Variances	26327.152	2368.341	11.116	<.0001	1.000	

Table 3-8: LCBC data latent change score model results for change in WASI Matrix, surface area, thickness, and volume over time. Effect size is calculated by dividing the mean change by the square root of the variance.

As shown in Table 3-6, the three 2LCMs fit the data well: CFI area = 0.987; CFI volume = 0.921; CFI thickness = 0.994. Change in all structural brain metrics was significantly associated with change in cognition, with surface area showing the largest effect (r = +.35, p <.001), followed by thickness (r=+.21, p <.001), then volume (r=+.15, p =0.001). The Steiger's Z-Test revealed that the change-change relationship between area and cognition was significantly stronger than that between volume and cognition and thickness and cognition (Table 3-7).

The LCBC results therefore successfully replicated Cam-CAN's cross-sectional and longitudinal results, further supporting the finding that changes in surface area predict

changes in cognition, and that this relationship is stronger than that between change in thickness and change in cognition.

3.5 Discussion

3.5.1 A morphometric double dissociation

Across two independent cohorts, I found evidence of a morphometric double dissociation: cortical thickness was more strongly associated with age than cortical surface area, both cross-sectionally and longitudinally, whereas surface area was more strongly associated with cognition (fluid intelligence); certainly longitudinally, and also cross-sectionally, after removing age-related variance. Note that I am not claiming that cortical thickness plays *no* role in cognition – it shows a longitudinal association with cognitive change in one of the two datasets (albeit significantly smaller than that of surface area), and its cross-sectional association with fluid intelligence was significant. The lack of cross-sectional association with age-residualized fluid intelligence could be due to collider bias whereby cortical thickness is causally related to both age and cognition and that any thickness-cognition effect disappears when removing age. This chapter's results do suggest, however, that surface area and thickness, which tend to be investigated together through the aggregate measure of volume, may have dissociable causes (e.g., in ageing) and consequences (e.g., for cognition).

These findings align with previous studies that have pointed to a relationship between surface area and cognition (Cox et al., 2018; Fjell et al., 2015; Gerrits et al., 2016) and support recent calls to focus on the distinctness of cortical thickness and surface area, rather than assessing them jointly through cortical volume (Winkler et al., 2018). Such a shift is not just of theoretical or methodological importance: because surface area and cortical thickness are known to be genetically distinct (Panizzon et al., 2009; Winkler et al., 2010) and to follow different trajectories over the lifespan (Fjell et al., 2015;

Hogstrom et al., 2013), combining them into volume is likely to obscure important biological differences and mechanisms.

While I can, in the present study, only speculate on the biological basis of different morphological metrics (and therefore their age/cognition dichotomy), evidence from animal and histological studies point to a possibly relevant set of mechanisms. With age, the long dendrites of pyramidal neurons have been shown to decrease rapidly across all layers of the cortex (Jacobs et al., 2001; Nakamura et al., 1985; Panizzon et al., 2009) and especially in layer V – the internal pyramidal layer – which contains the majority of large pyramidal neurons and is therefore the thickest of the six cortical layers – at least after the age of 50 (de Brabander et al., 1998). Thus, the steep declines in cortical thickness observed in the present study (and elsewhere, e.g., Lemaitre et al., 2012; Chen et al., 2011) are likely in part due to dendritic shrinkage (Goriounova & Mansvelder, 2019).

Furthermore, the finding that cortical thickness is less strongly associated with cognitive abilities than other measures of brain structure is also supported by animal research, showing that rates of dendritic atrophy in rats did not differ between aged cognitive unpaired and aged cognitive impaired animals (Allard et al., 2012)

What, if not dendritic atrophy, is driving cognitive differences and cognitive change, and why might cognition be related to surface area? According to the radial unit hypothesis (Rakic, 2000), while the development of cortical thickness is driven by the layers in the cortical columns (as described above), the development of surface area is a product of the number of radial columns perpendicular to the pial surface. This theory has been updated via the Supragranular Cortex Expansion Hypothesis (Nowakowski et al., 2016), which postulates that specific cellular mechanisms allow certain types of glial cells to migrate towards the pial surface during development, thereby expanding the cortex, and that this process is, in turn, responsible for many of the cognitive features unique to primates. This is further supported by analyses suggesting that glial cells –

and specifically glial-neural signalling – affect cognition (Chung et al., 2015). A plausible hypothesis therefore is that MR-derived surface area (at least partially) picks up on these glial-dependent neural mechanisms – which likely originate in early development – and thereby cause cognitive differences and changes.

3.5.2 The shape of the ageing brain

A second contribution this chapter makes is to characterize structural age-related differences and changes across multiple morphological metrics. While there have been multiple robust studies comparing different imaging metrics (Hutton et al., 2009; Im et al., 2008; Lövdén et al., 2013b; Pantazis et al., 2010; Shimony et al., 2016; Y. Wang et al., 2019; Wierenga et al., 2014), few have included the breadth of morphometry assessed here. This approach, therefore, allowed me to directly compare the magnitude of cortical age-related differences and changes across a range of metrics.

The biggest age-related change (cross-sectionally and longitudinally) was that of cortical thickness, followed (cross-sectionally) by curvature. This suggests that the most striking structural transformation the human brain undergoes with age – at least of those detectable with MRI – is that the cortex thins while also becoming more 'curved'. The width and depth of cortical sulci might influence the complexity metric, such that more atrophied brains might exhibit an increase in gyral complexity but not a decrease in surface area (Narr, et al., 2004; Lemaitre et al., 2012).

We also show that *combining* shape measures outperforms any individual metrics' ability to capture age-related and cognitive differences: together, the eight morphometric metrics assessed here explained almost double the variance compared to that captured by thickness and surface area alone. Thus, the fact that multiple morphometric measures provided partially complementary information about the outcome highlights the potential usefulness in assessing various morphological shape measures when investigating the ageing brain and cognitive abilities.

3.5.3 Methodological strengths and limitations

In addition to the large sample size and the assessment of multiple shape metrics, the integration of cross-sectional and longitudinal data is of note. As discussed in Chapter 1, recent reviews and commentaries have pointed to the limitations of cross-sectional analyses when investigating brain-cognition relationships in the ageing brain (see Oschwald 2020 for a discussion). While I agree that collecting longitudinal data is almost always preferable, I acknowledge that it is not always attainable. The present approach of integrating cross-sectional and longitudinal data, where the latter largely confirmed the findings of the former, offers some validation of cross-sectional approaches.

Another key strength of this chapter is the successful replication of the cross-sectional and longitudinal findings in an independent cohort. In doing so, I not only validated the apparent existence of the morphological double dissociation, but showed that it is not subject to specific features of the Cam-CAN data. Indeed, replicating our results despite important differences between the two datasets increases the robustness of our findings considerably. For instance, the cognitive tests differed (Cattell in Cam-CAN, WASI Matrix in LCBC), suggesting that surface area captures the broader construct of fluid intelligence (rather than test-specific features). Moreover, while the morphological metrics assessed in our initial Cam-CAN study offered an intriguing description of the ageing brain, obtaining them required five separate processing pipelines: FreeSurfer (Fischl, 2012), FreeSurfer Long (Reuter et al., 2012), Mindboggle (Klein et al., 2017), SPM (Ashburner & Friston, 2000) and the Fractal Dimensionality Toolbox calcFD (Madan & Kensinger, 2016). The fact that our results replicated in canonical metrics (all of which are part of the standard FreeSurfer output) might lower the threshold for future research to, where appropriate, investigate surface area and cortical thickness separately.

The breadth of structural brain metrics reviewed in this Chapter also comes with some important limitations. First, I was not able to investigate the *changes* of several of the

metrics which I had assessed in our cross-sectional analyses. This is because the pipelines used to calculate these additional metrics (e.g., Mindboggle) are not yet optimised for longitudinal data. This is particularly pertinent to curvature, which showed a very strong age effect cross-sectionally, but would have been interesting to explore longitudinally. Likewise, fractal dimensionality, which measures cortical complexity and correlated strongly with age *and* cognition in our cross-sectional analyses, might be a promising candidate for future longitudinal investigations.

3.6 Conclusion

In this Chapter, I describe cross-sectional and longitudinal evidence for a brain-cognition double dissociation: two morphological metrics, surface area and cortical thickness, which tend to be investigated together through grey matter volume, are differentially associated with age and fluid intelligence: while thickness is strongly associated with age, it has weak associations with change in fluid intelligence – a pattern that is reversed for surface area, which captures cognitive change and difference well, and age relatively poorly. I would therefore argue that rather than using grey matter volume as the default measure, researchers should choose structural brain metrics depending on the question under investigation. Doing so will allow us to advance our understanding of the functional significance of these dissociable aspects of brain morphology.

Chapter 4: Morphometry as a mediator in the lifestyle – cognition relationship

4.1 Chapter Summary

After finding relationships between a more active lifestyle and better cognitive abilities (Chapter 2), as well as between brain structure and cognitive ageing (Chapter 3), I wanted to build on these chapters by better understanding the role brain structure plays in the link between lifestyle and cognition. Specifically, I was interested in whether grey matter *mediates* – that is, statistically accounts for part of – the lifestyle-cognition association. In other words, to examine the magnitude of effects under the assumption that lifestyle differences affect brain structure, which in turn affect cognitive performance in old age. To study this question, a first set of analyses explored how grey-matter volume mediates the relationship between Chapter 2's lifestyle factors and fluid cognition. I observed significant mediation effects for all five lifestyle-cognition associations, but the proportion mediated was largest for physical health: grey-matter volume explained 31 percent of the association between the latent Physical Health factor and fluid cognitive abilities.

To further explore the mediating role of brain structure in the physical health – cognition relationship, I next moved away from the latent factor (which was comprised of a set of variables unique to Cam-CAN and is therefore not readily replicable across cohorts) and zoomed in on of the latent factor's key components: *systolic blood pressure*. This proxy for cardiovascular health is widely measured in ageing samples, including both the Cam-CAN and LCBC cohorts. This allowed me to focus on a specific, replicable, and mechanistically plausible question: how does brain structure mediate the

relationship between cardiovascular health (as indexed by systolic blood pressure) and cognitive abilities? In both Cam-CAN (N=641) and LCBC (N=835) grey-matter volume mediated a substantial percentage of the relationship between systolic blood pressure and cognitive abilities. However, as shown in Chapter 3, volume is an aggregate morphometric measure. As such, it may mask the potentially different roles cortical thickness and surface area play in mediating the physical health – cognition relationship. Strikingly, we observed just such a dissociation: Using equality-constrained model comparisons, we found that the grey-matter volume mediation was comprised of two dissociable effects: cortical thickness related to the blood pressure-brain components (a-paths), but surface area related to the brain-cognition components (b-paths) of the mediation models. I discuss the biological mechanisms that might underlie this pattern and note the implications for mediation models relying on composite indicators. These effects were observed in time-lagged cross-sectional data, where we can be confident about the temporal sequence of the measurements and replicated across two large-sample cohorts. However, one limitation on generalisation is that we did not observe the same effects in a smaller subset of the data for which we have true longitudinal measurements: changes in blood pressure did not predict changes in cognitive abilities. I discuss how and why this weakens the strength of the evidence and limits the possibility of causal inference.

4.2 Introduction

The goal of this chapter is to build on the core findings of the previous two chapters by assessing lifestyle, brain structure and cognition simultaneously. The following introductory section frames these analyses by discussing the evidence concerning two important questions: i) how good is the evidence that there is, in fact, a causal link between lifestyle (and especially physical health) and cognition, and ii), what role does brain structure play in this relationship? This introduction will show that the evidence for a causal lifestyle cognition effect in humans is, overall, mixed, but that there is good

evidence supporting the effectiveness of multimodal and exercise interventions on improved or maintained cognitive performance. I will then discuss the biological mechanisms that likely underpin the effect of exercise on cognitive performance as well as whether MR-derived brain structure can be used to partially capture such mechanisms. Finally, I will discuss the concept of mediations: can brain structure mediate the relationship between exercise and cognitive abilities, and if so, what does that tell us about the age-brain-cognition triangle?

4.2.1 How good is the evidence of a causal lifestyle – cognition relationship?

An intriguing promise is often made: "live an active life – one enriched by social, intellectual and physical activities – and you shall be cognitively healthier for longer". An even more hopeful idea is that lifestyle changes in mid- or even late-life – such as picking up a sport or learning an instrument in middle or old age – can still alter the course of cognitive decline (rather than such protective effects needing to accumulate over a lifetime of healthy choices). But just how good is the evidence of a causal link between (changes to) lifestyle activities and cognitive health?

The cross-sectional association between lifestyle and cognitive performance is well established, with studies pointing to positive correlations between social, physical or intellectual engagement and cognitive abilities (Fratiglioni et al., 2004; Hertzog et al., 2008; Lindenberger, 2014; Stern, 2009), as discussed in more detailed in Chapter 2. Longitudinal designs, too, have shown that change in engagement in activities is positively associated with change in cognitive functions in older adults, suggesting that improvements to one's lifestyle attenuate cognitive decline. This effect has been found for physical activities (Angevaren et al., 2010; Ku et al., 2014; Larson et al., 2006; Lindwall et al., 2012; Small et al., 2012b), cognitive activities (Mitchell et al., 2012; Small et al., 2012b), social activities (Brown et al., 2016; Lövdén et al., 2005) as well as for composites of different kinds of lifestyle engagement (Bielak et al., 2007, 2012). Such studies have

led scientists to conclude that about a third of Alzheimer's disease (AD) cases (and an even larger proportion of non-clinical cognitive decline) worldwide were attributable to potentially modifiable risk factors (Norton et al., 2014). This claim is further supported by findings of a large Lancet Commission (Livingston et al., 2017), which recommends being "ambitious about prevention" of dementia and cognitive decline, focusing on interventions to build up resilience and healthier lifestyles.

However, as discussed in the General Introduction (Chapter 1), cross-sectional and longitudinal study designs do not readily allow for a causal interpretation of the data: there could be reverse causation (better cognitive health leading people to engage in more challenging lifestyle activities), or lifestyle and cognition could be causally affected by a third, independent variable such as socioeconomic status. To assess whether lifestyle changes cause improvements in cognition, experimental manipulations in forms of randomized controlled trials (RCTs) are needed: these allow researchers to compare two groups of individuals who differ only in the lifestyle intervention they receive. That way, whatever cognitive difference is found between the two groups following the trial should, in principle, be due to the intervention itself.

To date, there are few large-sample RCTs assessing "multi-domain" activities, where multiple aspects of people's lives such as diet, cognitive training and social activities were targeted simultaneously. The first such study, the Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability (FINGER), which included 1260 elderly adults (age 60-77) who were at risk of cognitive decline, found that the two-year intervention had significant beneficial effects on cognitive change overall, as well as on processing speed, executive function and complex memory tasks (Ngandu et al., 2015). The intervention also benefited non-cognitive outcomes including BMI, diet, physical activity and quality of life. The promising results of the FINGER intervention has led to the launch of the Worldwide FINGERS network in 2017 (https://wwfingers.com/), with data being collected across the globe to further assess the potential usefulness of this intervention. However, two other large-sample RCT studies with similar designs to the FINGER model (the French Multidomain Alzheimer Preventive Trial (MAPT, N = 1860)

and the Dutch Prevention of Dementia by Intensive Vascular Care (PreDIVA, N = 5326)) showed more mixed results: in both studies, the primary outcome – a decrease in dementia incidence – was not shown, even though secondary cognitive measures (such as orientation and short-term memory) showed some improvement following the interventions. Together, these RCT results suggest that multimodal interventions might improve cognitive performance in healthy adults, but are not sufficient to prevent or slow down diseases like Mild Cognitive Impairment (MCI) or AD (Kivipelto et al., 2018).

A 2017 systematic review (Kane et al., 2017) paints a similarly mixed picture of the benefit of interventions. The authors assessed the effectiveness of 13 types of interventions (e.g., cognitive training, physical activity, diet or vitamin supplements) in a total of 263 studies for preventing or delaying the onset of age-related cognitive decline, MCI and AD. They concluded that there was no "high strength" evidence (defined as RCTs with a low risk of bias, with consistent, direct, and precise domains; see discussion on strengths of evidence and their impact on policy in Chapter 5 regarding the effectiveness of any of these interventions). However, the authors note that most of the included studies were small, targeted interventions, and that multimodal, large-sample designs such as the FINGER trial (which they discuss but do not include in their formal systematic review) could be a promising intervention strategy. The one intervention method that likely does have a positive effect on cognition, however, is exercise. Here, the authors conclude that "although physical activity interventions show no consistent benefit in preventing cognitive decline, the proportion of results showing benefit was unlikely to be explained solely by chance, providing a signal of a possible relationship" (Kane. et al., 2017, *p. vii*). Similarly, a meta-analysis of aerobic exercise RCTs concluded that aerobic exercise training is associated with modest improvements in attention and processing speed, executive function and memory (Smith et al., 2010). Studies using observational data also report this positive effect of exercise: in an analysis of four longitudinal studies (N=604, participants aged 80 and older at the time of first examination) using multilevel linear mixed models, Lindwall and colleagues observed change-change relationships between physical

activity and cognitive performance in some but not all cognitive domains. Most consistently, change in physical activity was related to change in reasoning (in 4 of 4 studies), and less so to change in fluency (2 of 3 studies), memory (2 of 4 studies) and semantic knowledge (2 of 4 studies Lindwall et al., 2012).

The follow-up report of the above-mentioned Lancet commission drew a similar conclusion, leading it to categorize seven of the 12 modifiable risk factors for dementia and cognitive decline as "physical health factors" (Livingston et al., 2020): hypertension, obesity, physical inactivity, alcohol consumption, smoking, head injury and diabetes. (The other factors were air pollution, less education, infrequent social contact, depression, and hearing impairment.) The relative robustness of the effect of physical health and activity on cognitive health might, in part, be due to there being more published papers on that association compared to on those other aspects of lifestyle and cognitive abilities. For instance, a PubMed search in February 2022 with the keywords "physical health" and "cognition" in papers' titles or abstracts yielded 7511 results, whereas the words "lifestyle" and "cognition" showed 1173 results. Equally (and perhaps more) plausible, however, is the interpretation that there *is* a relationship between physical health and cognitive performance, and that the former does, in fact, causally affect the latter. If such an effect exists, there should be testable biological mechanisms underpinning this relationship. To examine this question, a good understanding of the neural effects of exercise is warranted.

4.2.1.1 The biological mechanisms of the exercise – cognition relationship

In 1947, Donald Hebb showed that housing rodents in an environment enriched with balls, ladders, running wheels, tunnels and other complicated toys – as opposed to in an empty cage – improved their learning abilities and memory (Hebb, 1947). Beneficial effects of such "enriched environments" on rodent brain and behaviour have since been reported in dozens of studies, suggesting that intellectually, physically and socially complex environments induce neural plasticity and, in turn, enhance cognitive functions (for reviews see Nithianantharajah & Hannan, 2006; van Praag et al., 2000).

This, of course, led researchers to hypothesize that this form of complexity – engaging in novel, cognitively challenging activities – could also protect humans from cognitive decline and dementia (M. C. Carlson et al., 2012; Hertzog et al., 2008). Then, a 2011 study steered the conversation in a different direction: Kobilo and colleagues showed that running wheels were the key determinant of neuroplasticity and cognition-enhancement (rather than the balls, tunnels or toys): Enriched environments *without* running wheels did not improve the rats' memory, but a cage containing *only* running wheels did (Kobilo et al., 2011; Mustroph et al., 2012). This suggests it is the challenging cardiovascular exercise that is the key cognition-enhancing factor.

The brain-derived neurotrophic factor (BDNF) protein plays a crucial role in the mechanisms underlying exercise-induced cognitive improvements. Often called the 'gatekeeper to neural plasticity' (Cowansage et al., 2010), BDNF is known to support neural survival, synaptic functioning, axonal growth and many other aspects of neural development and functioning (Cowansage et al., 2010; Gorski et al., 2003; Matsunaga et al., 2004). Importantly for this chapter, BDNF can be triggered environmentally: not only has BDNF been found to increase following exercise in rodents, but blocking BDNF has been found to diminish the cognitive improvements otherwise induced by exercise, suggesting that BDNF is, indeed, *necessary* to observe exercise-induced cellular effects (Miranda et al., 2019). How does aerobic activity elevate BDNF levels in the brain? The main hypothesis here is that higher metabolic rates during exercise lead to a secretion of signalling molecules that subsequently upregulate BDNF levels (Morland et al., 2017). In adults, this effect is thought to be both immediate and (relatively) long-lasting, with higher BDNF levels found two weeks following a single aerobic exercise session (Hopkins et al., 2011). Regular exercise, conversely, may lead to permanently higher levels of BDNF, priming the brain to be better prepared for learning experience and other cognitive tasks (Miranda et al., 2019): according to a meta-analysis, regular exercise intensifies the effect of a single session of exercise on BDNF levels in humans (Szuhany et al., 2015) – a potentially relevant finding for future interventions. Note, however, that BDNF research (especially in humans) is likely subject to publication bias,

with some authors arguing that the effect of BDNF-induced cognitive improvements is being exaggerated in the literature (Dodds et al., 2013).

Meanwhile, BDNF declines with age, as evident from animal (Matsunaga et al., 2004) and human (Driscoll et al., 2012) studies, the latter of which have shown that BDNF levels (measured in the blood) predict both hippocampal volume and cognitive functioning (Erickson et al., 2010). Importantly, BDNF levels still increase in aged animals following exercise, although these effects are not as robust as those seen in younger animals (Praag et al., 2005). Taken together, it seems likely that this protein, whose levels increase with exercise and decrease with age, and which is known to affect cognition, provides a plausible mechanism for explaining the cognitive benefits of exercise-based interventions.

4.2.1.2 Morphometric correlates

In the absence of direct measures of BDNF, can MR-derived metrics of brain structure capture exercise-induced cognitive improvements? I would argue that they can, partially. As BDNF has been linked to neurogenesis and increased dendritic complexity (Rossi et al., 2006), I argue that cortical thickness (as well as subcortical, e.g., hippocampal, volume) should be most sensitive to this effect (Jan & Jan, 2010). In contrast, MRI measures that are not, or only weakly, associated with dendritic changes (such as FreeSurfer-derived cortical surface area) should be less strongly associated with exercise or cardiovascular measures. I will outline – and then test – this hypothesis in more detail below. First, I will discuss existing empirical evidence regarding the associations between MR-derived brain structure and exercise-cognition relationship in humans.

In one study (N=2000), people who were more physically active in mid-life had larger total brain volume 20 years later compared to less active participants, with regions in the frontal lobe showing the strongest associations (Rovio et al., 2010). Several other papers found hippocampal effects, with aerobic exercise interventions, for instance,

leading to increases in hippocampal volume (Erickson et al., 2011). In general, regional specificity should be expected as post-mortem and animal studies have shown that BDNF does not present uniformly across the brain (Miranda et al., 2019); exercise-induced plasticity should therefore follow the molecular pathways of BDNF expression. White matter, too, has been linked to physical health and exercise, with some studies pointing to increased white matter volume (Driscoll et al., 2012; Strömmer et al., 2020). However, other studies found no such associations (Podewils et al., 2007; Rovio et al., 2010). A 2016 systematic review concluded that there is moderate evidence of links between physical health and white matter structure (Sexton et al., 2016). Overall, these studies support the hypothesis that MR-derived brain structure metrics *can* detect aspects of the exercise-cognition effect, and that they do, in part, reflect BDNF-affected processes.

For observational (i.e., non-interventional) data, yet stronger support for this hypothesis would come by integrating patterns of findings into a plausible mechanistic model, rather than stopping at showing only simple associations between pairs of variables. This is because univariate associations between exercise and brain structure alone could, in principle, reflect changes to brain structure that do not necessarily play a mechanistic or causal role in supporting maintained cognitive performance in old age; in other words, they could be an irrelevant by-product of increased exercise rather than the reason for improved cognitive abilities. Likewise, associations between brain structure and cognition may not be *due* to exercise. Nonetheless, if there is a relationship between aerobic fitness and cognitive performance, and if this relationship is caused by (perhaps BDNF-induced) changes to brain structure, then brain structure should mediate – that is, statistically account for part of – the relationship between aerobic activity and cognition.

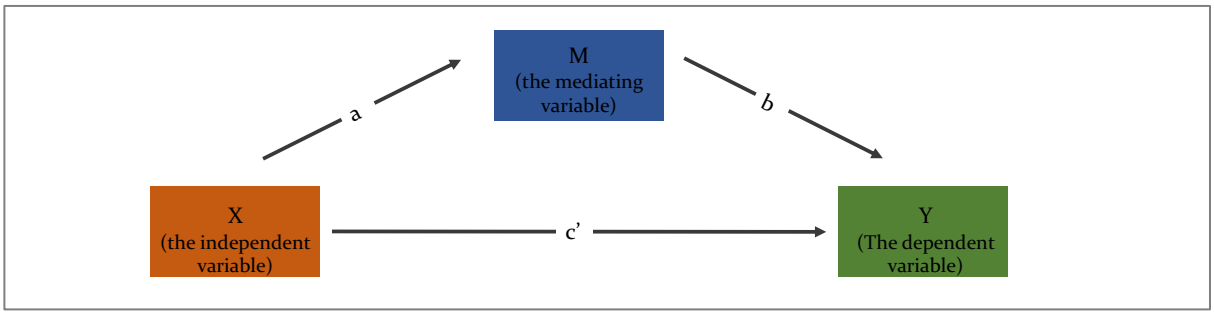


Figure 4-1: Schematic representation of a mediation.

To date, there are only a handful of studies that assess the exercise-grey-matter-cognition relationship with mediation models. A study of 310 older adults with MCI reported that benefits of physical exercise on memory performance were mediated by hippocampal volume (Makizako et al., 2015). Weinstein and colleagues showed that grey-matter volume in several prefrontal brain regions mediated the relationship between fitness and executive function and working memory (Weinstein et al., 2012). Finally, Verstynen and colleagues found that dorsal striatum volume (specifically the caudate nucleus) statistically mediated the relationship between fitness and cognitive flexibility, a function thought to be supported by this region (Verstynen et al., 2012).

Mediation models assume that part of or all variance shared between two variables is also shared by a third variable, which has been used to argue that the third variable is causing the association between the other two variables (Shrout & Bolger, 2002). Such causal interpretations are often problematic, particularly when mediation models are run on cross-sectional data, as is the case in the studies cited above (Raz & Lindenberger, 2011). At their core, mediation models conducted on observational, cross-sectional data contain a series of regression coefficients: 1 bivariate regression coefficient (i.e., a) and 2 partial correlation coefficients (i.e., b and c′; see Figure 4-1). Because the "mediator" variable (M) is only correlational (rather than truly interventional), its causal role is ambiguous: there are many alternative explanations for an observed correlational pattern which cannot be ruled out, such as reverse causation (for example, it is possible that cognition mediates the relationship between lifestyle and brain structure rather than lifestyle mediating the relationship between cognition and brain structure). Cross-

sectional (as opposed to longitudinal) mediation models are particularly vulnerable to reverse causation because in cross-sectional data, the assumption of correct temporal ordering of the variables can potentially be violated (Maxwell et al., 2011; Maxwell & Cole, 2007). In an observational (i.e., non-interventional) setting, if the dependent and independent variables are collected at the same time, then there really is no way of inferring (from the data alone) the directionality of an observed effect. Of course, this basic statistical issue – that correlation is not equal to causation – is not specific to mediations; it affects all regression analyses. However, it is of particular importance in the context of mediations because mediations have, as mentioned above, often been (mis)interpreted as evidence of causation.

Scholars have therefore suggested that longitudinal data – where the predictor variable was collected at an earlier time point than the outcome variable – are methodologically more robust, in terms of temporal causality (Raz & Lindenberger, 2011). Indeed, because temporal precedence enabled by longitudinal analyses allows for the testing of the direction of the paths between the variables, longitudinal data can offer improved insight into the dimensions and dynamics of temporal processes such as cognitive and neural ageing. But even a significant mediation in which the variables of interest have been collected at different timepoints does not mean that causation can safely be inferred. For example, the outcome variable might not change over time, so one would get equivalent results if it had been measured at the same time as the predictor variable. Therefore, the strongest (or "gold standard") causal inference enabled by observational (rather than interventional) data comes when mediation analysis is performed on change scores, i.e., differences in variables across timepoints in longitudinal data. For instance, if a change in cortical thickness mediates the relationship between a change in physical activity and a change in cognition, then the notion that physical activity *cause*s cognitive changes, and that these are *partially explained* by morphometric changes, is more likely (though, as with any statistical modelling, such a result could still be the consequence by an unknown, confounding variable that happens to change in the same way as the mediator variable).

Regardless of the level of statistical control (i.e., whether mediation is performed on simultaneous, lagged or change scores), it is also vital for any causal interpretation of mediation effects to have a solid theoretical and mechanistic explanation of the directionality of an effect. Indeed, for mediation models to address causality, the directionality of causation (i.e., X affecting Y in Figure 4-1) needs to not just be plausible, but its reverse (Y affecting X) should be implausible.

A final issue is that the vast majority of neuroimaging studies assessing the relationship between physical activity and brain structure (whether through mediations or univariate associations) have focused on brain *volume* (the 2013 review by Voelcker-Rehage & Niemann, for instance, does not include a single cortical thickness paper). Volumetric studies have consistently pointed to effects of higher physical fitness or cardiovascular intervention on, for instance, volume in the frontal cortex (Colcombe et al., 2006; Ruscheweyh et al., 2011; Weinstein et al., 2012) and the temporal lobe, which contains the hippocampus (Erickson et al., 2011; Szabo et al., 2011). I think that it is likely that the volumetric effects observed in many of these studies are, in fact, driven by cortical thickness (see Chapter 3). Indeed, more recent studies have shown increased cortical thickness in more physically active older adults in temporal (Raffin et al., 2021; Walhovd et al., 2014; Williams et al., 2017) and frontal brain regions (Lee et al., 2016). In other words, a central question addressed in this Chapter is whether an aggregate measure like grey-matter volume (in FreeSurfer, the product of cortical thickness and surface area) might be conflating separate aspects of the causal exercise-cognition chain, and whether these distinct aspects can be detected using mediation analyses.

4.2.2 Measuring fitness

One advantage of animal and intervention designs is having an objective measure of physical activity: one group exercises, the other does not. Observational studies must rely on more indirect proxies of physical activity, which tend to fall into two categories: self-report data, on one hand, such as Cam-CAN's "EPAQ" questionnaire (see https://dapa-toolkit.mrc.ac.uk/instrument/40) that asks participants to state how

frequently they partake in mild, moderate, or strenuous physical activity. These measures are subjective and are prone to inaccurate (usually exaggerated) levels of exercise (Prince et al., 2020). Physical measures, on the other hand, can assess a person's fitness more objectively. The most accurate measure of physical fitness is maximal oxygen update ($VO2_{max}$) during exercise. $VO2_{max}$ indicates the maximal capacity of the cardiovascular system to provide oxygen to muscle cells during sustained exercise, and is usually measured by an incremental test on a motor-driven treadmill or a bicycle ergometer (Plasqui & Westerterp, 2005); an expensive and time-consuming task which is often not suitable for elderly participants (Huggett et al., 2005). As an alternative to $VO2_{max}$, researchers have relied on a series of non-invasive measures to assess cardiovascular health and fitness levels. Heart rate variability (HRV) has relatively recently been identified as robust proxy of cardiovascular health and has also shown to be related to cognitive performance (Luque-Casado et al., 2013); although HRV has also been criticized for its interpretability (Heathers, 2014). The idea is that a heart rate that is variable and responsive to demands reflects a healthy cardiovascular system, whereas reduced HRV may be associated with poorer cardiovascular health, aerobic fitness levels and other negative health outcomes (Routledge et al., 2010). HRV is best measured with an electrocardiogram, but can also be captured using wearable devices such as smart watches, opening up promising new research avenues (Goessl et al., 2017; Saif et al., 2020). If heart rate metrics are not available, as is the case in Oslo's Lifespan Changes in Brain and Cognition (LCBC) cohort, blood pressure can be used as measure of cardiovascular health. High blood pressure (hypertension) is one of the strongest indicators of poor cardiovascular health (Wu et al., 2015), and can therefore serve as a (albeit indirect, see discussion) measure of fitness. Systolic blood pressure especially, which measures the force the heart exerts on the walls of the arteries, has been linked to cognitive age-related differences and changes (Gottesman et al., 2014; Launer et al., 1995) as well as to reduced brain volume (Beauchet et al., 2013; Gianaros et al., 2006; Muller et al., 2010; Swan et al., 1998) and cortical thickness (Walhovd et al., 2014). Furthermore, a relatively large body of research has pointed to associations between

blood pressure and measures of white matter (e.g., Fuhrmann et al., 2019) but since the focus of this thesis is on grey matter, these studies are beyond the present scope.

4.2.3 Summary

The evidence for lifestyle activities positively affecting cognitive performance is mixed. The most robust effects have been found for physical activity, where increases in aerobic exercise led to (modest to moderate) improvements in cognition. Animal studies suggest that this effect is likely due to neurotrophic processes: exercise causes rises in brain derived neurotrophic factor (BDNF), which in turn stimulates neurogenesis and dendritic arborisation (Nokia et al., 2016). It is likely that MR-derived measures of grey-matter can partially capture this effect. One particularly insightful avenue for analysing this exercise-brain-cognition relationship in observational data is to assess whether grey-matter mediates the longitudinal relationship between changes in exercise and changes in cognition: More specifically, whether the dissociable aspects of cortical thickness and surface area in the previous chapter also represent different mediation pathways.

4.2.4 Present study

This chapter assesses the mediating role of grey matter in the link between cardiovascular health (as indexed by systolic blood pressure) and fluid cognitive abilities. Using (equality) constrained mediation models and model comparisons, it tests the hypothesis that such a mediation is comprised of two dissociable components, which are masked when investigating only the aggregate measure of grey-matter volume: while volume should mediate the overall relationship between blood pressure and cognition, cortical thickness should capture the “blood pressure-brain” aspect of the mediation, while surface area should be related to the “brain-cognition” part of the mediation.

4.3 Methods and Results

4.3.1 Cam-CAN Lifestyle Factors

As described above, the most compelling and mechanistically plausible evidence for a brain-mediated lifestyle-cognition effect comes from studies investigating physical health and exercise, while the neural mechanisms underlying the potential benefits of other aspects of lifestyle (such as intellectual or social engagement) are less well understood. It therefore follows that, of the five latent lifestyle factors created in Chapter 2, brain structure should best mediate (that is, explain the largest amount of variance) of the link between the Physical Health factor and cognition. To test this assumption, I ran five mediation models, one per lifestyle factor, to i) assess whether grey-mater volume mediates the relative lifestyle-cognition association and ii) to calculate (and then compare) the effect size of each mediation by dividing the indirect estimate by the estimate for the total effect. Lifestyle, cognitive and morphometric data were available for 641 participants (age range 18-88, median age 54). Mediation models were estimated in lavaan (Rosseel, 2014). The alpha level was Bonferroni-adjusted (0.05/5) at 0.01.

Lifestyle Factor	**Total effect (lifestyle – fluid intelligence)**		**Mediation effect (grey matter volume)**		**Effect size (indirect/total)**
	Beta (standardized)	*p*	*Beta* (standardized)	*p*	
Physical Health	-0.536	<0.001	-0.164	<0.001	0.31
Social Engagement	0.239	<0.001	0.068	0.002	0.28
Intellectual Engagement	0.361	<0.001	0.10	<0.001	0.28
Education/SES	0.516	<0.001	0.152	<0.001	0.29
Mental Health	-0.302	<0.001	-0.069	0.002	0.23

Table 4-1: mediation results. Brain volume mediates the lifestyle - fluid intelligence relationship of physical health, intellectual engagement, and education/SES

As shown in Table 1, whole brain grey-matter volume significantly mediated the relationship between all five lifestyle factors and fluid intelligence: physical health (β = -0.16, p = <0.001), social engagement (β = 0.068, p = 0.02), intellectual engagement (β = 0.10, p = <0.001), education/SES (β = 0.15, p = <0.001) and mental health (β = -0.069, p = 0.002). Note that the negative sign for physical health is because of how the physical health factor is comprised, with higher scores reflecting worse physical health.

To compare the size of these mediation effects, I divided the indirect estimate by the estimate for the total effect, which provides the percentage of the total effect explained by the mediating mechanisms (in this case brain volume; Iacobucci, 2012). Mental health had the smallest effect size (0.23), and physical health the largest (0.31), suggesting that grey-matter volume explained 23 and 31 percent of the direct effect, respectively. Moreover, not only does grey-matter volume exert greatest mediation for physical health, but the basic effect being mediated – that between lifestyle and cognition – is also greatest in magnitude for physical health (0.536). Note however that the five lifestyle factors are correlated (see Chapter 2), so we cannot determine the unique contribution of each.

However, instead of continuing the analyses with the latent factor of physical health, I was interested in a physical health measure that would be more readily comparable across samples and most closely approximates the biological mechanisms discussed above. As noted, the latent factor created in Chapter 2 was comprised from a relatively unique set of observed Cam-CAN variables, making replication of this factor in a different cohort difficult. Of the four observed variables making up the Physical Health lifestyle factor (see Chapter 2), systolic blood pressure was the one available in both Cam-CAN and LCBC. As discussed in the Introduction, systolic blood pressure measures the force the heart exerts on the walls of the arteries and has been shown to be an effective and reliable approximation of cardiovascular health (Fuhrmann et al., 2019; Gottesman et al., 2014; Launer et al., 1995).

The following sections explore the mediating role of grey matter in the relationship between systolic blood pressure and fluid cognitive abilities in four main steps. Step one assesses grey-matter volume mediations in both synchronously and asynchronously collected (i.e., time-lagged) data in Cam-CAN. Step two (again in Cam-CAN) explores cortical thickness and surface area separately, rather than as the aggregate volumetric measure. Step three replicates the previous two steps in a separate cohort, LCBC. Step four moves beyond cross-sectional inferences to assess change-change mediation models in LCBC, asking whether change in systolic blood pressure predicts change in cognitive abilities, and if so whether this association is mediated by a change in brain structure.

4.3.2 Grey Matter Volume Mediation models

Starting with whole brain grey matter volume in Cam-CAN, this significantly mediated the relationship between systolic blood pressure and fluid cognitive abilities (β = -0.18, p = 0.001), explaining 39 percent of the variance (indirect effect / total effect: -0.12 / -0.305 = 0.39). This mediation model is depicted in Figure 4-2.

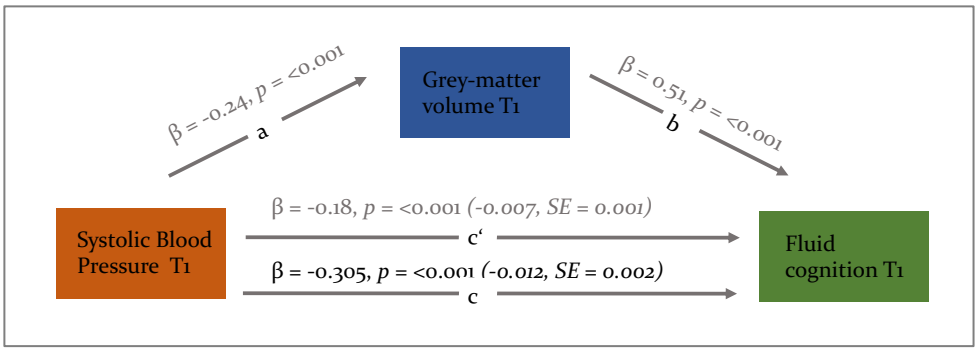


Figure 4-2: grey-matter volume significantly mediates the relationship between systolic blood pressure and fluid cognitive abilities in Cam-CAN (N=541). Shows standardized betas (as well as un-standardized parameters for the c/c' path in italics). Here, all data were collected at the same time (T1).

Because mediation models of cross-sectional data represent age-related differences in target variables, they may not capture time-dependent relations (see Raz & Lindenberger, 2011 and Introduction above). To shed light on the dynamics of cognitive and morphometric ageing, longitudinal data are necessary. As discussed in the

introduction, the gold standard are change-change models (see LCBC section below), but data collected at different time points are still advantageous (that is, they make the possibility of a causal inference more likely) than data collected at the same time. Thus, I next ran the above mediation model in the longitudinal Cam-CAN data (N = 99; cases using FIML: N=552), where cognitive abilities and grey-matter volume data were obtained after systolic blood pressure measures (on average 1.3 years (sd = 0.66 years) later). As shown in Figure 4-3, grey-matter volume at timepoint 2 significantly mediated the relationship between blood pressure at timepoint 1 and fluid cognition and timepoint 2 (indirect effect a*b = ~0.109), explaining 32 percent of the variance (indirect effect / total effect: -0.109/ -0.34 = 0.32). Thus, this time-lagged model provides temporally unambiguous evidence of a mediating role of grey-matter volume in the link between systolic blood pressure and fluid cognition.

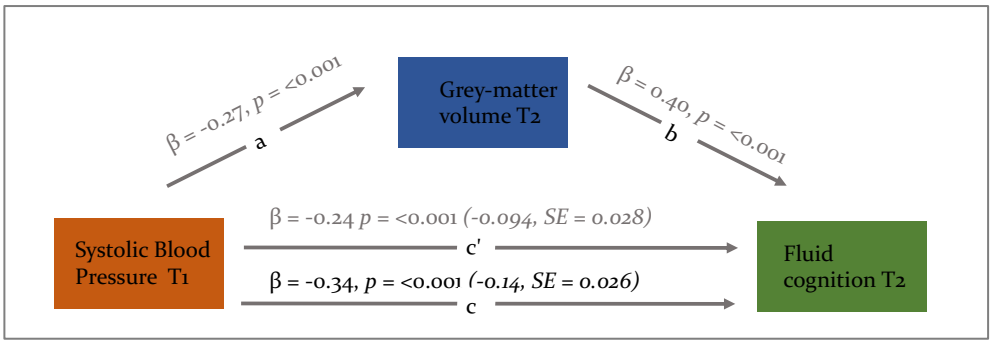


Figure 4-3: grey-matter volume significantly mediates the relationship between systolic blood pressure and fluid cognitive abilities in Cam-CAN (N=99). Shows standardized betas (as well as un-standardized parameters for the c/c' path in italics). Here, brain and cognitive data were collected after blood pressure measures (T2 vs. T1).

4.3.3 Cortical thickness and surface area make dissociable contributions to the mediation model

The above analyses investigated the role of grey-matter volume in mediating the relationship between cardiovascular health and fluid cognitive abilities. As shown in the Introduction, while most neuroimaging papers assessing the relationship between

physical activity and cognition have focused on volume, it is likely that it is cortical thickness – not surface area *and* cortical thickness, i.e., volume – that is most directly affected by (changes in) physical health or fitness. I argue that this is because the biological mechanisms underlying exercise-induced cognitive improvements entail neural and dendritic changes, which cortical thickness (not surface area) might partially capture. I therefore hypothesized the following pattern (depicted in Figure 4-4):

- **Hypothesis 1:** The cardiovascular health - brain aspect of the mediation model should be driven largely by cortical thickness, and not surface area: the association between thickness and systolic blood pressure should be stronger than that between surface area and systolic blood pressure.
- **Hypothesis 2:** Surface area should drive the brain-cognition aspect of the mediation model: the association between surface area and fluid cognitive abilities should be stronger than that between cortical thickness and fluid cognitive abilities.

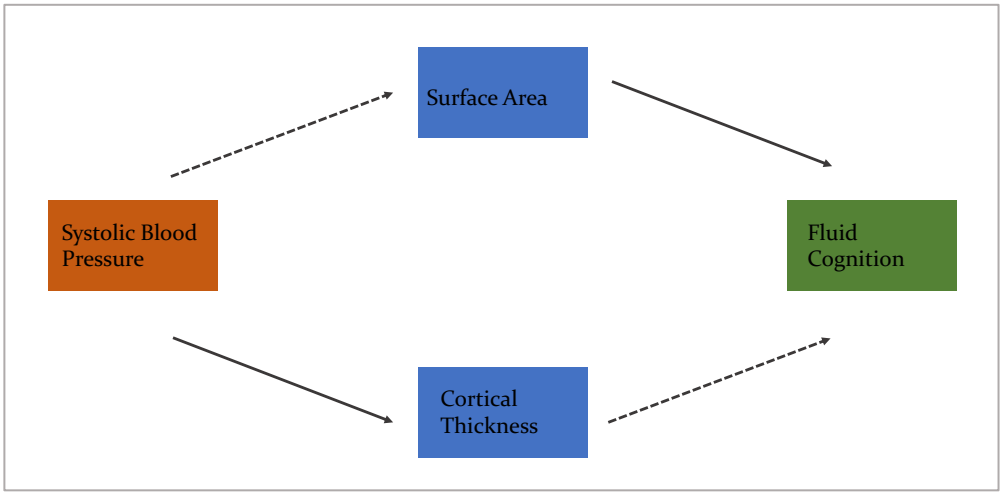


Figure 4-4: Hypothesized pattern of how surface area and cortical thickness could be differentially mediating the blood pressure - cognition relationship. Thick lines reflect hypothesized stronger effects; dashed lines reflect hypothesized weaker effects.

Testing these hypotheses with the methodology adopted here allows this chapter to overcome three important weaknesses of previous studies. Firstly (and perhaps most importantly), mediation models using grey-matter volume do not have a biological or

mechanistic explanatory base because they conflate two separate aspects of morphology into one aggregate measure, which may lead to "artificial" mediations. Conversely, separating cortical thickness and surface area in the mediation models provides an informative and innovative way to shed light onto mechanistically plausible dynamics of the lifestyle-brain-cognition triangle. Secondly, mediation studies often rely on cross-sectional data, making the possibility of causal inference more difficult. Here, cross-sectional mediation findings are followed up with time-lagged and longitudinal analyses. Thirdly, using two large, independent samples increases statistical power and improves the robustness of the results.

To better compare parameter estimates, cortical thickness and surface area were rescaled prior to the mediation analyses which, in turn, were estimated in lavaan (Rosseel, 2014) in in R-Studio 1.1.463 (R version 3.6.3). Note that for all subsequent mediation models in this chapter that include two brain metrics (i.e., cortical thickness and surface area), I did not add a residual covariance between the two simultaneously estimated metrics (that is, I did not allow surface area and cortical thickness to covary). This is because surface area and cortical thickness tend to be uncorrelated in ageing samples (see Chapter 2). Further inspection showed that this simplification as tenable: allowing co-variance did not significantly improve model fit (Cam-CAN χ^2 difference = 0.0024, p = 0.96; LCBC χ^2 difference = 0.0071, p = 0.87).

I first conducted a multiple mediation analysis, which included surface area and cortical thickness as mediators. Because all paths were estimated freely, this model is called the 'free model'. Figure 4-5 shows the standardized betas, while the mediation model results are summarized in Table 4-2. Here, systolic blood pressure at baseline (T1) was significantly associated with cortical thickness at T2 (β. = -0.23, p = 0.01), but not with surface area (β = -0.22, p = 0.21). On the other hand, both cortical thickness (β = 0.25, p = 0.008) and surface area (β = 0.37, p = <0.001) were significantly associated with fluid cognition at T2. Thus, at first glance, this free model offers support for hypothesis 1 (blood-pressure/brain) but not hypothesis 2 (brain/cognition).

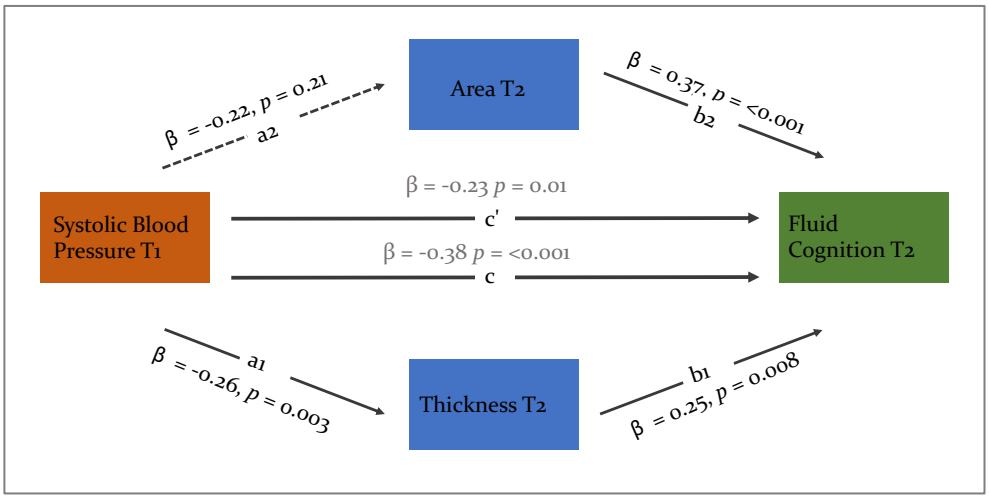


Figure 4-5: Regression paths of the free model with raw cognition in Cam-CAN (Model A). Dashed lines refer to non-significant paths, solid lines reflect significant paths. N=99.

However, as shown in the previous chapter, the correlation between cortical thickness and fluid cognition *was* stronger than that between surface area and fluid cognition – this effect only reversed after age-residualizing the cognitive abilities. Thus, I ran a second free model with age-residualized fluid abilities.

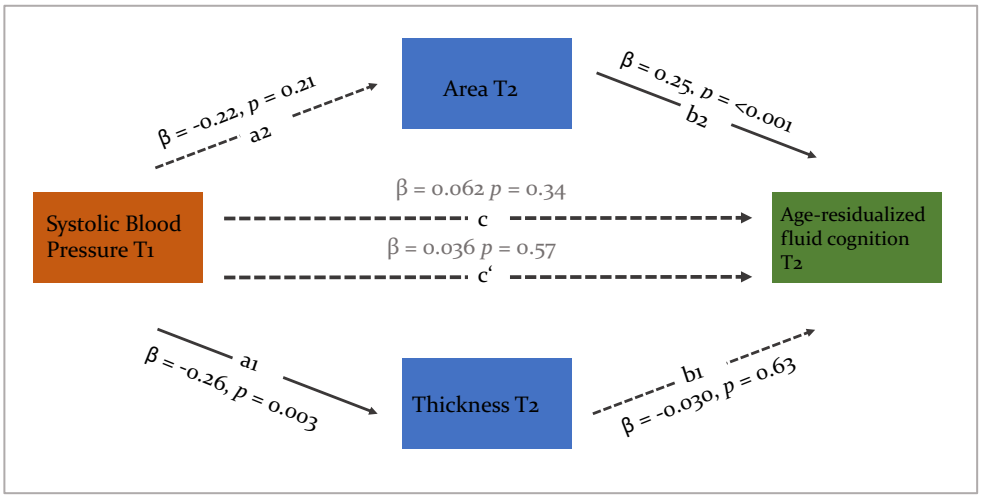


Figure 4-6: Regression paths of the free model with age-residualized cognition in Cam-CAN (Model A). Dashed lines refer to non-significant paths, solid lines reflect significant paths. N=99.

As shown in Figure 4-6, the cardiovascular health – brain pattern mirrored that of the previous free model (significant brain-blood pressure associations for thickness, not for area), while the brain-cognition pattern differed: here, surface area was significantly

associated with age-residualized fluid abilities (β = 0.25, $p < 0.001$), while thickness was not (β = -0.30, $p = 0.63$). Note, however, that in Cam-CAN (unlike LCBC, see below), the original association between blood pressure and age-residualized cognitive abilities was not significant (β = 0.061, $p = 0.34$), i.e., there is no effect to-be-mediated in the first place, which limits the conclusions that can be drawn once cognition is residualized with respect to age in the Cam-CAN data.

4.3.3.1 Formal hypothesis testing: equality constrained and zeroed models

Inspecting the regression paths of the free model offers an interesting conceptual – though not formal – understanding of the mediation patterns (at least when cognition was unadjusted for age). To formally test the above hypotheses, I ran a series of equality-constrained mediation models, both for raw and age-residualized cognition. This approach allowed me to compare the free models to models in which one or more of the paths defining the indirect and direct effects (e.g., paths a1 and a2) are constrained to be equal. If the constrained model fits significantly worse than the free model, this suggests that there is, in fact, a difference in the magnitude of the path coefficients. In other words, if free models are preferred, despite their greater degrees of freedom (dfs), this would suggest that area and thickness differ in their role in mediating the cardiovascular health – cognition relationship.

After constraining the BP-brain paths (paths a1 and a2; Model B) and the brain-cognition paths (paths b1 and b2; Model C), I compared these two constrained models to the free model (Model A) using likelihood ratio tests. Model comparison results are depicted in Table 4.2. Here, Model A fit better than Model B, but not than Model C. This suggests that for the BP-brain paths (a-paths), cortical thickness accounts for significantly more of the mediating variance, but that area and thickness do not differ in their brain-cognition (b-paths). This is in line with first model depicted in Figure 4-5. I repeated the same steps for the age-residualized cognition models, comparing constrained BP-brain paths (Model E) and constrained brain-cognition paths (Model F) to the free model (Model D). Here, the free model fit the data best, suggesting that

cortical thickness and surface area did contribute differentially to the mediating effect of brain structure in the relationship between systolic blood pressure and cognitive abilities, once effects of age are removed from the latter. This, again, is in line with the model depicted in Figure 4-6.

These equality-constrained analyses show that for age-residualized abilities, cortical thickness and surface area differentially mediate the association between blood pressure and cognitive abilities, respectively. For raw (not age-adjusted) cognitive abilities, the a-paths differed significantly, while the b-paths did not. This suggests that the overall GM volume mediation shown earlier (and in several published papers) is likely comprised of two separate and dissociable components, which remain masked when assessing GM volume alone.

Model			**DF**	**AIC**	**BIC**	**Chi-Square**	***p***
Fluid cognition	A	Free	1	1811.0	1858.3	3.79	
	B	BP-brain (a1 & a2) paths constrained	2	1813.4	1856.6	8.18	0.036
	C	Brain-cognition (b1 & b2) paths constrained	2	1809.3	1852.4	4.01	0.64
Age-residualized fluid cognition	D	Free	1	2842.1	2889.6	4.08	
	E	BP-brain (a1 & a2) paths constrained	2	2843.7	2886.9	7.72	0.046
	F	Brain-cognition (b1 & b2) paths constrained	2	2849.6	2892.8	9.53	0.0020

Table 4-2: Likelihood ratio test results comparing free and constrained models in Cam-CAN.

A final question was whether thickness and area just play *different* roles in the mediation of cardiovascular health and cognitive abilities, or if the 'weaker' paths do, in fact, play *no significant role* at all (see Figure 4-7 for schematic representation). To assess this, I investigated whether constraining the weaker paths to zero would impair model fit when comparing the zeroed models to the free models. If the zeroed models are

preferred, the weaker paths of the mediation models are negligible. Conversely, if the free models are preferred, this suggests that the weaker paths (although weaker) still contribute significantly to the mediating effect. Model comparison results are summarized in Table 4-3.

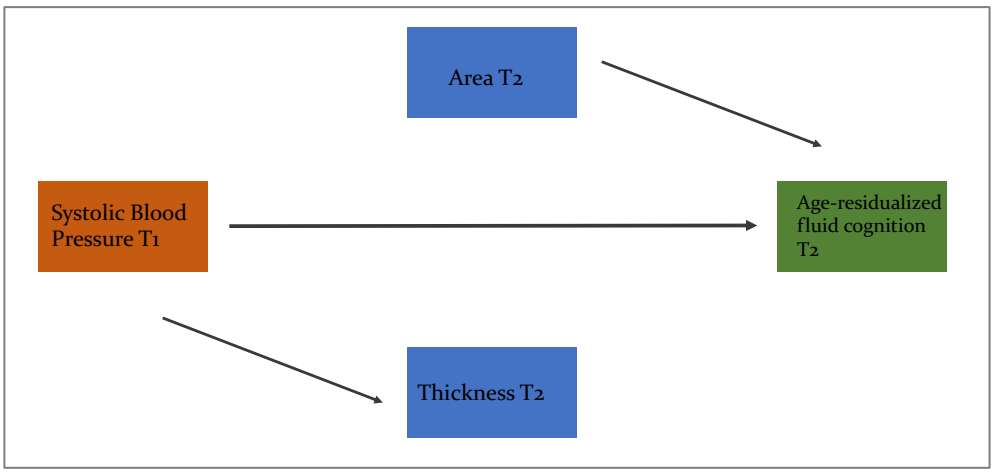


Figure 4-7: a schematic representation of zeroed paths improving model fit, suggesting that the weaker components of the blood pressure - brain and the brain - cognition paths are statistically negligible. This pattern was evident only in LCBC, *not* Cam-CAN (where the weaker paths were not statistically negligible).

For both raw fluid cognition and age-residualized cognition, the free models were preferred, suggesting that the weaker paths did statistically contribute to the mediating effect, and were therefore not negligible.

Model		**DF**	**AIC**	**BIC**	**Chi-Square**	***p***
Fluid cognition	**Free**	1	1811.0	1858.5	3.79	0.0027
	Zeroed	3	1818.9	1857.7	15.64	
Age-residualized fluid cognition	**Free**	1	2842.1	2889.6	4.08	0.047
	Zeroed	3	2844.2	2883.0	6.12	

Table 4-3: Model fit comparison between free and zeroed models in Cam-CAN. The preferred models are bolded.

4.3.4 LCBC

After running the above mediation models in Cam-CAN, I turned to the LCBC data (described in Chapter 2) for two reasons: first, I wanted to investigate whether my findings would replicate in a larger (N=835) longitudinal dataset. Second, because LCBC (unlike Cam-CAN) has change scores of systolic blood pressure, morphometry, and cognition, I was interested in exploring change-change mediation models which, as explained in the Introduction, would offer the best possible evidence of causality in observational data.

4.3.4.1 Replication of Cam-CAN mediation models

I replicated the previous four mediation analyses with the LCBC data, i.e., using BP at T1, and GM and Cognition at T2. Overall, all analyses replicated well, offering further support for the hypothesis that cortical thickness and surface area play differentiable roles in mediating the relationship between blood pressure and cognitive abilities. First, grey-matter volume mediated the relationship between systolic blood pressure at baseline (T1) and fluid cognitive abilities at T2 (on average, 5.18 years (sd = 2.59 years) after the baseline assessment; β. = -0.083, p = 0.001), explaining 35 percent of the variance (indirect effect / total effect: -0.083 / -0.24 = 0.345; (Iacobucci, 2012). Second, the regression paths of the multiple mediation analyses (which included surface area and cortical thickness as mediators) generally mirrored the Cam-CAN ones: The a-paths differed for both raw and age-residualized cognition (that is, systolic blood pressure at baseline was significantly associated with cortical thickness at time point 2, but not with surface area). Conversely, the b-paths differed only in the age-residualized models, not in the raw cognition models. An important difference between the LCBC and the Cam-CAN models was that for LCBC, blood pressure *was* significantly associated with age-residualized cognition, suggesting that there was a to-be-mediated effect. These results are summarized in Table 4-4 and Figure 4-8.

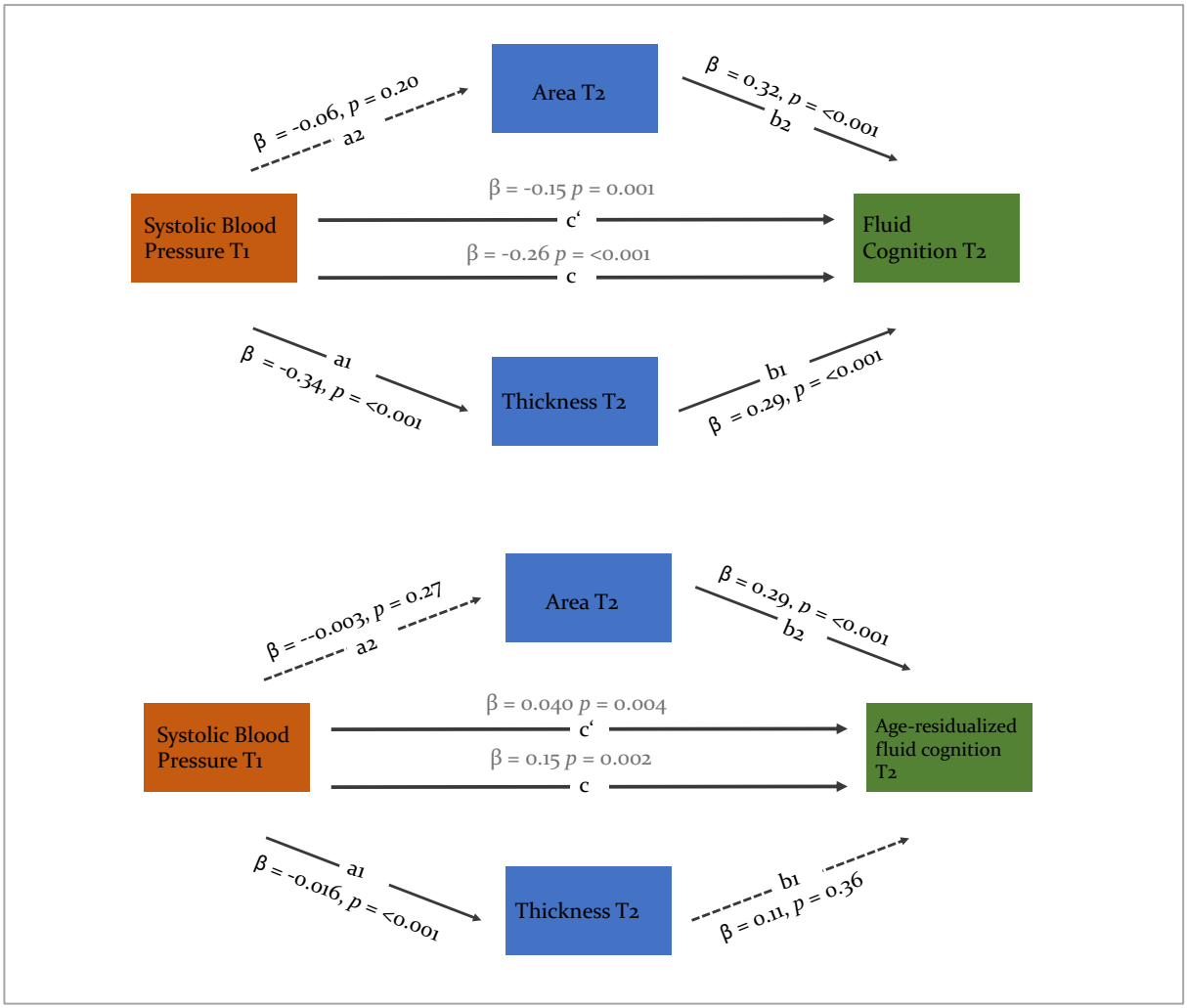


Figure 4-8: Free models in LCBC (N=835). Top: fluid cognitive abilities, bottom: age-residualized cognitive abilities. For age-residualized cognition, cortical thickness and surface area contribute differently to the mediating effect. Dashed lines reflect non-significant effects.

Third, equality constraining the a- and b-paths (to formally test whether cortical thickness and surface area make different contributions to the mediating model as outlined in the hypotheses above) showed that for both raw and age-residualized fluid cognition, the free models were preferred, suggesting that surface area and cortical thickness do, indeed, make differentiable contributions to the relationship between blood pressure and cognitive abilities (see Table 4-4). Note that these findings differed

slightly from the Cam-CAN results, where the free model was only preferred for age-residualized cognition.

Model			DF	AIC	BIC	Chi-Square	*p*
Fluid cognition	A	Free	1	2178.6	2230.6	2.86	
	B	BP-brain (a) paths constrained	2	2188.1	2235.3	14.31	0.00072
	C	Brain-cognition (b) paths constrained	2	2185.1	2232.3	11.29	0.0037
Age-residualized fluid cognition	D	Free	1	3494.9	3546.9	2.62	
	E	BP-brain (a) paths constrained	2	3518.7	3566.0	28.45	0.00037
	F	Brain-cognition (b) paths constrained	2	3498.2	3448.1	3.87	0.034

Table 4-4: Likelihood ratio test results comparing free and constrained models in LCBC. For both fluid intelligence and age-residualized fluid intelligence, free models fit the data better, suggesting that area and thickness play a differential role in mediating the lifestyle-cognition relationship.

Fourth, when comparing models where the weaker a- and b- paths were zeroed, free models fit better for the raw cognition models (χ^2 difference = 55.96, p = <0.0001). For the age-residualized models, the zeroed models were preferred (χ^2 difference = 4.31, p = 0.12). This suggests that for age-residualized cognition (though not the raw cognitive models), the weaker paths can be considered statistically negligible (Table 4-5). This was different in Cam-CAN, where the free models were preferred for both raw and age-residualized cognition.

Model		DF	AIC	BIC	Chi-Square	*P*
Fluid cognition	Free	1	2178.6.1	2230.6	2.87	
	Zeroed	3	2227.7	2270.2	55.96	<0.0001
	Free	1	3494.9	3546.9	2.62	

Age-residualized fluid cognition	Zeroed	3	3495.2	3537.7	6.93	0.12

Table 4-5: Model fit comparison between free and zeroed models in LCBC

4.3.4.2 Change-change mediation models

After finding that the core mediation results replicated well across the two cohorts, I next investigated whether *changes* in systolic blood pressure predicted slower cognitive *decline* and if so, if *changes* in brain structure might mediate this effect. Here, the mediation models included blood pressure, brain structure and cognitive annualized *change scores* (rather than baseline blood pressure and brain and cognitive data from time point 2).

In this 'gold standard' model (which included three change scores: blood pressure, cognitive and morphometric; Figure 4-9), changes in blood pressure did not significantly predict changes in fluid cognition (β = -0.235, p = 0.189), so there was no to-be-mediated effect. However, only 38 participants had all three change scores (N=124 were included in the model using FIML); it is therefore possible that the lack of significant association is due to insufficient statistical power.

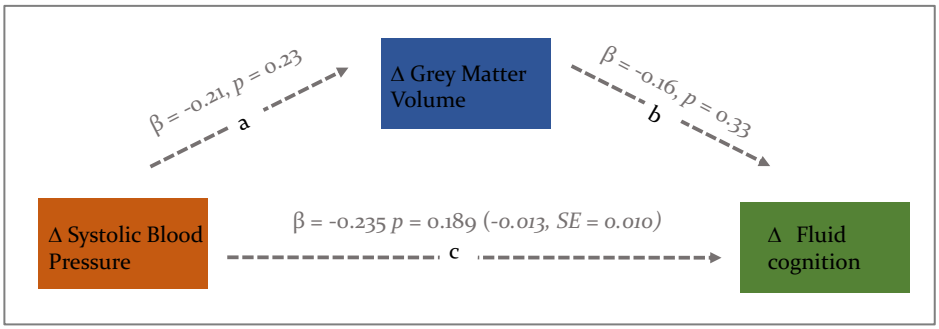


Figure 4-9: The "gold standard" model in LCBC (N=124 with FIML), which assessed changes in all three variables. No association between changes in blood pressure and changes in fluid cognition, suggests that there is no to-be-mediated effect. Dashed lines represent non-significant paths. The numbers in parenthesis reflect raw parameters, SE = standard error.

In the next step, I therefore relaxed requirements to only include *two* change scores (blood pressure and grey-matter volume; this increased the sample size to N=71; N=166 using FIML) and modelled i) whether changes in blood pressure would predict fluid cognition at time point 2 and ii) whether blood pressure at baseline was associated with changes in cognition (and whether, if either of these effects were found, they are mediated by morphometric changes). These models are depicted in Figures 4-10 and 4-11, respectively.

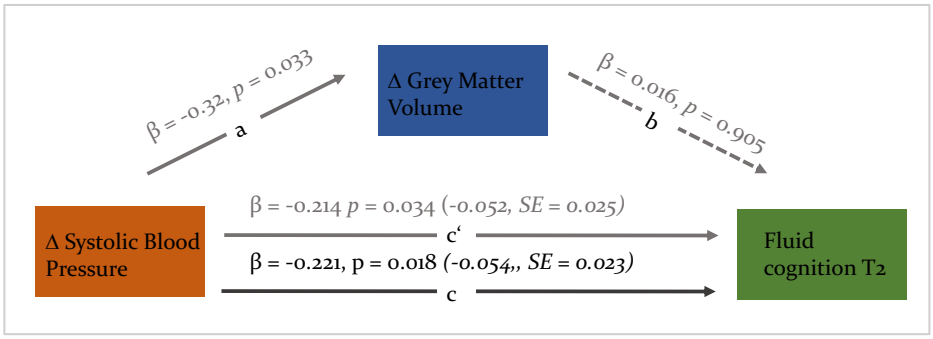


Figure 4-10: In LCBC (N = 166 using FIML), changes in systolic blood pressure predict fluid cognition at time point 2, an effect partially mediated by changes in grey-matter volume. Dashed lines represent non-significant paths. The numbers in parenthesis reflect raw parameters, SE = standard error.

For the first of these two models (Figure 4-10), changes in systolic blood pressure were significantly associated with fluid cognition at time point 2 (β = 0.22, p = 0.018). The mediation effect was also significant (β = 0.21, p = 0.034). Because there was a significant association between changes in systolic blood pressure and changes in grey-matter volume (β = 0.032, p = 0.033), but not between changes in grey-matter volume and fluid cognition and time point 2 (β = 0.016, p = 0.905), it is difficult to interpret this result. Moreover, sample size here was still relatively small (N=71), so this pattern should be investigated further in subsequent research.

Model ii) had a larger sample size (N=267), but baseline systolic blood pressure was not significantly associated with changes in cognition (β = -0.055, p = 0.37), which meant that there was no to-be-mediated effect (Figure 4-11).

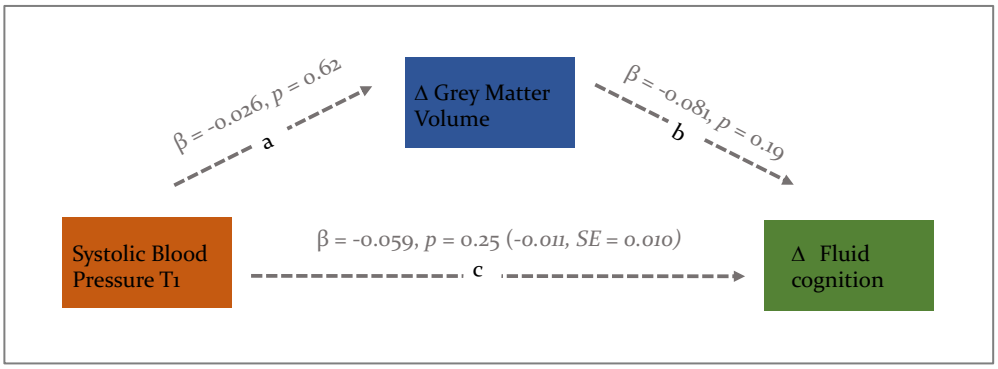


Figure 4-11: In LCBC (N=267), systolic blood pressure at baseline was not significantly associated with changes in fluid cognition. Dashed lines represent non-significant paths.

In summary, there was only relatively weak evidence of any change-change associations, with all but one analysis pointing to null effects. While changes in blood pressure were significantly related to fluid abilities at time point 2 (Figure 4-10), because the sample size was small (N=71) and thus likely underpowered, I decided not to conduct further analyses with this model. I therefore did not test the hypotheses that cortical thickness and surface area contribute differently to this mediating effect in any models involving change scores. This means that, in this thesis, all evidence for a differentiable contribution of thickness and area in mediating the relationship between blood pressure and cognitive abilities comes from time-lagged cross-sectional data (see discussion).

4.4 Discussion

In this chapter, I explored how grey matter mediates the relationship between lifestyle – particularly cardiovascular health as indexed by systolic blood pressure – and cognitive abilities. In two large-scale, independent datasets, I show that i) whole brain grey-matter volume mediates the relationship between blood pressure and cognition overall and that ii) this grey-matter mediation is comprised of two separate, differentiable effects, whereby cortical thickness drives the blood pressure – brain component (a-path) and surface area mediates the brain-cognition component (b-path) of the mediation model. These findings elucidate the relationship between physical health and cognition

in three important ways: First, separating cortical thickness and surface area in the mediation models provides an informative and innovative way to shed light onto mechanistically plausible dynamics of the lifestyle-brain-cognition triangle. Second, mediation studies often rely on cross-sectional data, making the possibility of causal inference more difficult. Here, cross-sectional mediation findings are followed up with time-lagged and longitudinal analyses. Third, using two large, independent samples increases statistical power and improves the robustness of the results.

Below I will discuss these findings in the context of the existing literature and ask what they do and do not tell us about the possibility that MR-derived brain structure captures a potentially causal relationship between cardiovascular health and cognitive abilities.

4.4.1 Summary of findings

I first showed that grey matter volume mediated the relationship between the five Cam-CAN lifestyle factors created in Chapter 1 (physical health, mental health, social engagement, intellectual engagement, and education/SES) and cognitive abilities. The largest mediation effect was found for the association between the Physical Health factor and cognition. Note, however, that the lifestyle factors are correlated (see Chapter 1), so we cannot determine the unique contribution of each. Furthermore, because these analyses were conducted using cross-sectional data (that is, all variables were collected at the same time), not much can be inferred regarding the direction of causality. Finally, because these latent lifestyle factors were comprised of a relatively unique set of observed variables, they were not readily replicable across cohorts.

Systolic blood pressure, on the other hand, was available in both Cam-CAN (where it was one of the observed variables of the Physical health latent factor) and LCBC. It was therefore chosen as proxy for cardio-vascular health. Both cohorts presented the same general and interesting mediation pattern: while grey-matter volume mediated the relationship between systolic blood pressure and fluid cognitive abilities, this single mediating effect consisted of two separate components. Equality-constrained models

revealed that cortical thickness contributed more to the blood pressure – brain (a paths) aspect of the mediation models while surface area contributed more to the brain-cognition (b paths) aspect of the models. Note, however, that this was only true for age-residualized fluid abilities in both cohorts. Moreover, zeroing the weaker paths revealed that for age-residualized cognition, the weaker paths were not just weaker but statistically negligible in LCBC; i.e., there was no evidence for a role of cortical thickness in the brain-cognition aspect of the mediation model, nor a role of surface area in the blood-pressure brain aspect of the model. Note that this was not the case for the raw cognitive ability models in either sample – here the weaker paths were still of statistical significance.

4.4.2 Present findings in the context of other studies

Other mediation studies support the above grey-matter volume findings, pointing to a mediating role of grey-matter volume in the association between physical health and cognition (Makizako et al., 2015; Verstynen et al., 2012; Weinstein et al., 2012). There are, to my knowledge, no studies assessing the mediating role of cortical thickness and/or surface area in the cardio-vascular health / cognition association. This Chapter's finding that cortical thickness drives the blood pressure – brain component of the mediation model is indirectly supported by a handful of studies that have pointed to associations between cardiovascular health and cortical thickness: A small-sample (N=30) study, which does not report a formal mediation analysis, found that a 12-week moderate intensity walking intervention improved fitness in older adults, which in turn, was correlated with cortical thickness change bilateral insula, precentral gyri, precuneus, posterior cingulate, and inferior and superior frontal cortices – regions that are vulnerable to Alzheimer's related atrophy (Reiter et al., 2015). A second (small-sample) study found associations of cardiorespiratory fitness (measured using $VO2_{max}$) with cortical thickness in several cortical areas, with the strongest effect noted in the left supramarginal cortex (Williams et al., 2017). Finally, a recent pooled analysis of N=1218 participants showed significant correlations between Heart Rate Variability and cortical

thickness, particularly in the orbitofrontal cortex (Koenig et al., 2021). It is striking that none of these studies report cognitive data, since one might expect the success of any fitness intervention to be determined by whether cognition improves. One possibility is that the authors did not find any associations between cortical thickness and cognition (which, according to this thesis, is to be expected) and therefore did not report these analyses.

Regarding the surface area-cognition component of the mediation model, the existing evidence is even more scarce than that for the thickness associations. I could not find a single study investigating surface area and physical health in older adults, whether as a mediation analysis or otherwise. A study of obese children (N=101) found that cardiovascular fitness was associated with cortical thickness but not surface area (Esteban-Cornejo et al., 2019). However, because thickness and area are known to play different roles in children than in adults, these findings are only of limited comparative value to the present results. Apart from studies assessing fitness, the link between surface area and cognition in (older) adults is slowly becoming increasingly accepted (see Chapter 2) – a development further supported by the present findings.

4.4.3 Limitations

Although this study has several strengths (including the large cohorts, the successful replication of the main findings, time-lagged mediation models and the inclusion of change-change analyses), there are also several limitations. Firstly, the quality and usefulness of the systolic blood pressure measure as an index of cardiovascular health. Although systolic blood pressure has been used in other studies to reflect cardiovascular health (see Introduction, and e.g., Gottesman et al., 2014) other metrics (such as Heart Rate Variability; Luque-Casado et al., 2013) are arguably better-suited to capture physical health. I chose blood pressure here because of its availability in both datasets. Since HRV might be a more sensitive measure of fitness, it would have been interesting to see how particularly the change-change models would have fared using this metric.

Another limitation is that many inferences are based on time-lagged cross-sectional, rather than on change-change mediations. Had changes in blood pressure predicted changes in cognition, and had this relationship been mediated by changes in grey matter, I could have been more confident in making causal inferences. The null effects reported here could be due to the small sample (N=38) of participants for whom all three change scores were available, or because too little time passed between the measurements to be able to detect meaningful changes in blood pressure, morphometry, or cognition. The null effects could, of course, also reflect a true null finding, where there is no causal relationship between cardiovascular health, at least as indexed by blood pressure and cognitive abilities. The absence of change-change effects makes the interpretation of the time-lagged cross-sectional mediation models less definitive. Finally, the arguably largest limitation of the present analyses is that it used only observational data. As discussed in the General Introduction, the best evidence for causal effects come from intervention studies, such as Randomized Controlled Trials.

4.4.4 Biological mechanisms

Assuming that a causal relationship between cardiovascular health and cognitive abilities is plausible, what biological mechanisms might underlie the mediation pattern found in this chapter? As outlined in the Introduction, one of the physiological responses to exercise in humans is an increase in BDNF (Ruscheweyh et al., 2011), which has been shown to mediate the relationship between exercise and increased hippocampal volume (Colcombe et al., 2006; Erickson et al., 2011) and is associated with greater neural connectivity (Voss et al., 2013). These neurotrophic effects have been hypothesized to contribute to the benefits of exercise on memory performance and induce angiogenesis (development of new blood vessels) in downstream regions such as the motor cortex (Pereira et al., 2007; Swain et al., 2003). It is therefore plausible that the relationship between cortical thickness and blood pressure found in this chapter reflects (at least in part) cardiovascular health-induced synaptogenesis or dendritic arborization, and that MR-derived cortical thickness therefore picks up on these

biological mechanisms. Note that any BDFN-related effects are unlikely to be the only mechanism underlying the exercise-cognition relationship, but that other processes are also at play (see for example Wenger et al., 2017).

The main puzzle in the present findings is arguably not about the mechanisms underlying the a-paths (although speculative here, they are well supported by animal and human research), but about the b-paths: if cardiovascular health affects cortical thickness, but cortical thickness does not impact cognitive abilities, then what does that say about the relationship between cardiovascular health and cognition? One possibility is that that exercise-affected cortical thickness does, in fact, impact cognitive abilities but that this effect is smaller than the general, exercise-independent relationship between surface area and cognition detailed in Chapter 2. In other words, exercise-dependent changes in cortical thickness might result in small cognitive improvements, but the majority of the cognitive variance is still explained by surface area. As explained in Chapter 2, the relationship between surface area and cognition could (in part) be due to certain types of glial cells migrating towards the pial surface during development, thereby expanding the cortex, and that this process is, in turn, responsible for many of the cognitive features unique to primates (Nowakowski et al., 2016; Rakic, 2000). This idea is summarized in Figure 4-12.

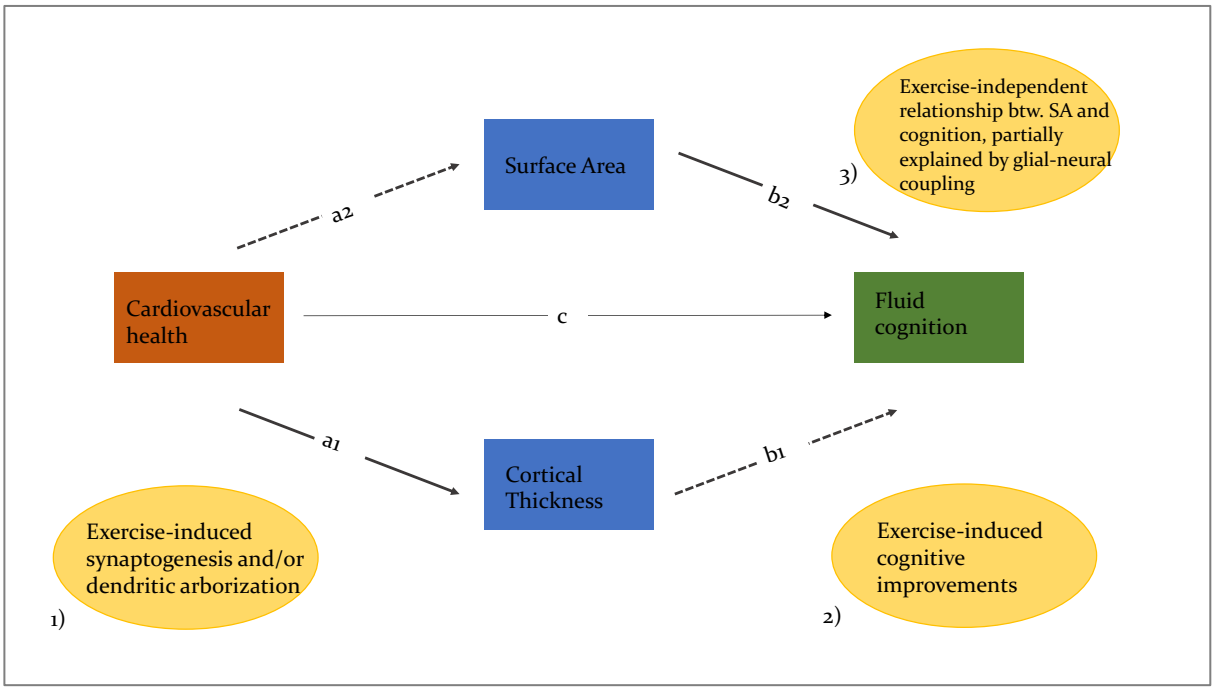


Figure 4-12: Hypothesized mechanisms that could partially explain the relationship between cardiovascular health and fluid cognition, and which can be detected using MR-derived brain structure metrics. Here, exercise increases cortical thickness via synaptogenesis and/or dendritic arborization (1). This, in turn, improves cognition (2), but this effect explains less of the cognitive variance than what is captured by surface area independently of exercise (3). Surface area explaining more of the cognitive variance is the reason why the b1) path (surface area – cognition) is stronger than the b2) (thickness – cognition) path in the mediation model.

Of course, any biological interpretation of the mediation models in these data is purely speculative. However, I would argue that the models presented here, coupled with plausible biological explanations, show how human morphology can be used in an informative and innovative way to shed light onto the dimensions and dynamics of the lifestyle-brain-cognition triangle. As I will discuss in Chapter 5, interdisciplinary and integrative work will be needed to find answers to the questions that the rapidly-ageing world is waiting for: might living an active life – one enriched by social, intellectual, and physical activities – really mean that we can be cognitively healthier for longer?

4.5 Conclusion

Grey matter mediates the relationship between lifestyle engagement and cognitive abilities. The physical health (as indexed by blood pressure) – cognition association, while at first sight mediated by whole-brain grey matter volume, is comprised of a dissociable mediation pattern: cortical thickness contributed more to the blood pressure-brain part of the mediating model, while surface area contributed more to the brain-cognition part of the mediating model. This provides further evidence for the importance to assess surface area and cortical thickness separately and shows how morphology can be used in an informative and innovative way to shed light onto the dimensions and dynamics of the lifestyle-brain-cognition triangle.

Chapter 5: General Discussion

5.1 Chapter summary

Below I summarize this thesis' main findings, as well as its strengths and limitations. Informed by my personal interest and experience in policy, I then discuss how scientific evidence is usually translated into policy as well as the role observational studies and mechanistic science have in this endeavour. Finally, I summarize current policy recommendations regarding modifiable lifestyle activities and cognitive ageing and discuss the extent to which this thesis contributes to future recommendations.

5.2 Summary of findings

In three empirical chapters, this thesis explored the relationships between lifestyle, cognitive abilities, and brain structure. Chapter 2 set out to demonstrate a way to model and assess the complex heterogeneity of people's lives and how it relates to fluid and crystallized cognitive abilities. To tackle the rich, observational dataset that is Cam-CAN, I first used Exploratory Structural Equation Modelling (ESEM, a data-driven technique to categorize a large amount of questionnaire-derived lifestyle variables into five factors: education/SES, physical health, mental health, intellectual engagement, and social engagement. Separate regressions showed that each of these factors was significantly associated with cognitive abilities. Because few studies had assessed different lifestyle aspects' joint relationship with cognition, the multiple regression models set out to capture the complementary role of multiple lifestyle components simultaneously – what effect does, say, social engagement have above and beyond education/SES and the other lifestyle factors? When all lifestyle factors were incorporated into the same model, social and intellectual engagement as well as physical health made independent contributions to fluid and crystallized age-adjusted

abilities, above and beyond the effect of education/SES. These relationships were robust across age and sex, and highly similar for fluid and crystallized domains, suggesting general effects, rather than effects specific to cognitive domain. Because, social, physical, and intellectual activities are potentially modifiable, this chapter offers observational, cross-sectional support in favour a beneficial relationship between lifestyle and cognitive ageing.

Chapter 3 focuses on brain structure. First, after deriving eight neuroimaging metrics from the Cam-CAN data (grey-matter volume, cortical thickness, Thickinthehead (an improved measure of cortical thickness), surface area, curvature, sulcal depth, fractal dimensionality, and total grey matter), I investigated these measures' cross-sectional associations with age and fluid cognitive abilities. This showed that combining shape measures outperforms any individual metrics' ability to capture age-related and cognitive differences, highlighting the potential usefulness in assessing various morphological shape measures when investigating the ageing mind and brain. Next, focusing on cross-sectional and longitudinal measures of surface area and cortical thickness (which together make up grey matter volume) in Cam-CAN and LCBC, I showed how these two separate aspects of morphology are differentially associated with age and cognition: while cortical thickness captured age well, it did not relate strongly to cognitive change (longitudinally) or age-residualized fluid abilities (cross-sectionally). Surface area, on the other hand, had relatively week associations with age, but related strongly to cognitive change and abilities.

Finally, Chapter 4 investigated the mediating role of grey matter in the relationship between potentially modifiable lifestyle activities and cognitive abilities. I showed, firstly, that grey matter mediates the association between Chapter 1's five lifestyle factors and fluid cognition. The strongest mediation effects were found for physical health, a relationship which is also well supported by other observational, interventional, and animal studies. Zooming in on physical health, as indexed by systolic blood pressure, I showed that (in Cam-CAN and LCBC) the grey-matter volume mediation was comprised of two dissociable components. Cortical thickness related more strongly to

the a-paths of the mediation model (the associations between blood pressure on morphology), while surface area related more strongly to the b-paths (the associations between morphology and cognitive abilities).

5.3 Strengths and limitations

This thesis' main strengths stem from its data. First, both the Cam-CAN and the LCBC cohort are large, age-heterogenous samples allowing for well-powered statistical analyses. As discussed in the General Introduction (Chapter 1), cognitive neuroscience has been negatively impacted by low-powered studies, which were part of the cause of the replicability crisis (Open Science Collaboration, 2015). Because high statistical power increases the chance of detecting a true effect, while also reducing the likelihood that a statistically significant result reflects a false effect (Button et al., 2013), the results in this thesis are likely robust. The robustness of the results is further improved because the thesis' core results in Chapter 3 (the morphometric double dissociation, whereby thickness relates to ageing and surface area relates to cognition) and Chapter 4 (the mediating role of surface area and cortical thickness in the relationship between systolic blood pressure and cognitive abilities) successfully replicated in an independent sample. Results replicated despite important differences in the datasets, such as a shorter time interval between the two waves in Cam-CAN compared to LCBC, and different measures of fluid cognitive abilities. Next, the morphometric phenotypes explored here are of note: I assessed eight brain structure measures calculated from multiple imaging pipelines. Investigating more than the canonical neuroimaging metrics of volume and cortical thickness allowed me to describe a previously largely overlooked phenomenon, the double dissociation discussed in Chapter 3, which I argue was overlooked precisely because of the field's focus on a relatively small set of morphometric tools.

The data studied here also come with limitations. Most importantly, the participants I investigated are adults from high-income, well-educated nations with good access to

health care (in fact, England's Cambridge, and Norway's Oslo are particularly well-off cities). These populations have been described as WEIRD: western, educated, industrialized, rich and democratic (Henrich et al., 2010a). The trouble, as anthropologist Joseph Henrich puts it: most people are not WEIRD (Henrich et al., 2010; although see Ghai, 2021). An increasing body of research suggests that many cognitive and behavioural traits thought to be universal apply predominately to the small western minority they were first detected in: amongst other things, people from different cultures differ in how they solve problems, raise their children, treat their elders, in how much risk they take and in what they eat and drink (see Diamond, 2013 for a comprehensive discussion). For instance, most people in western nations grow up speaking only one language, while bilingualism is the norm in many non-Western countries. Bilingualism has been consistently associated with enhanced cognitive functioning, with the largest effect sizes demonstrated for abstract thinking, attention, and problem solving (Adesope et al., 2010). Crucially for this thesis – and the generalizability of its findings – non-WEIRD people also seem to age differently. Although some age-related decline in cognition seem to be universal (e.g., Gurven et al., 2017), scholars have argued that cultural differences can affect both the rate of decline as well as its consequences for the affected individual (Fung, 2013). For example, people from collectivist nations report less loneliness with age, and feel more socially engaged in their communities (Barreto et al., 2021). Given that social engagement might affect cognitive ageing (e.g., Karp et al., 2006), it is possible that this form of social integration might (to some extent) attenuate the effects of cognitive decline. Thus, to better understand how cultural differences affect and interact with cognitive ageing, it is important for shared datasets to not only become larger, but also more diverse. In the meantime, this thesis' results generalise only to WEIRD populations, and (especially Chapter 2) should be interpreted with the appropriate caution on generalisability. Another limitation of the data is that they are purely observational, making causal inference less straightforward. I discussed the importance of experimental science including RCTs in Chapters 1 and 4 and will re-visit this issue in Section 5.4 below.

A second set of strengths and limitations stem from the analysis tools employed in this thesis. First, the data-driven grouping of lifestyle factors and the subsequent multiple regression models in Chapter 2 provide a methodically sound example of how to analyse large, observational datasets while minimizing researcher bias (for example in variable selection) and reflecting complex interactions between multiple simultaneously present variables. Second, the latent variable models, which capture the relationship between peoples' lifestyle and cognitive performance, account for measurement error, assess reliability and validity, and have greater generalizability and statistical power than methods based on observed variables (Jacobucci et al., 2019). Third, the analysis of longitudinal data using Latent Change Score Models, which allowed me to assess how morphometric and cognitive *changes* relate to each other, is of note. Only a small number of studies to date have adopted a similar approach (Hogstrom et al., 2013; Storsve et al., 2014; Y. Wang et al., 2019), and none have investigated the longitudinal relationship between surface area and cognition. Finally, the mediation models in Chapter 4, which explore the differential role of surface area and cortical thickness in mediating the relationship between blood pressure and cognitive abilities, provide (to my knowledge) a new, innovative way of testing MR-detectable mechanistic hypotheses underlying the association between physical and cognitive health. However, while I report strong longitudinal evidence for the morphometric double dissociation in Chapter 3, Chapter 4's mediation models are derived largely from time-lagged cross-sectional mediation models, making casual inference more difficult.

A limitation of the statistical tools employed here is that they rely on researcher-driven input and hypotheses. Could a more data-driven approach, such as machine learning (ML), have benefitted this thesis? It is almost certain that ML techniques (see Kuntzelman et al., 2021 for discussion) could have identified patterns in the data that human researchers have missed. For example, a recent paper investigated grey- and white- matter in 400 participants across multiple waves using a recently developed unsupervised ML technique called mixture-of-experts, which aims to detect heterogeneity in data (Eavani et al., 2016). The algorithm identified five subgroups of

brain ageing, each defined by different patterns of regional atrophy and white matter connectivity (Eavani et al., 2018). The authors conclude that this points to the presence of a variety of biological mechanisms which generate MR-detectable phenotypes. However, these neural subgroups did not correspond with differences in cognitive abilities, which limits their utility in explaining individual differences in cognitive ageing. I think that this reflects a general issue with ML-based techniques: pattern classification should only be the initial goal; ML's true usefulness will, ultimately, be to provide new ways to connect such neural patterns to cognitive or behavioural phenotypes, by being based on theory and plausible mechanisms (see Carlson et al., 2018 for discussion on the future of ML in neuroscience).

A final strength of this thesis is the attempt to link its findings to plausible underlying biological mechanisms. First, in Chapter 1, I discuss how the Locus Coeruleus, and its release of the neuromodulator noradrenaline, might relate to cognitive abilities and thereby serve as a neuro-mechanism of cognitive reserve. An exciting avenue of research would be to extend the morphometric mediation findings of Chapter 4 with MR-derived LC measures to investigate whether the LC mediates the relationship between lifestyle (especially social and intellectual engagement) and cognition. Second, I offer a mechanistic hypothesis for the morphometric double dissociation found in Chapter 3: it is possible that the rapid declines in cortical thickness (observed longitudinally and cross-sectionally) are partially due to age-related dendritic atrophy, while the surface area – cognition association might capture glial-neural coupling at the cortical surface, as postulated by the Supragranular Cortex Expansion Hypothesis (Nowakowski et al., 2016). I hope that surface area will become a more commonly-used measure of brain structure, especially in relation to ageing and cognition, and that future research will explore its biological underpinnings. Finally, I discuss the role of BDNF as one of the mechanisms associated with the relationship between physical exercise and cognitive health. My overall goal here was to provide an example of how different scientific approaches to (i.e., experimental, observational, interventional) and models (i.e., molecular, cellular, rodent, human) of the brain can be used to formulate a biological

mechanism which helps explain an epidemiologically observed phenomenon (the link between aerobic exercise and cognitive abilities). In addition to describing candidate mechanisms throughout the thesis, I show in Chapter 4 how mediation models can be built to reflect concrete, mechanistically plausible frameworks which may help the endeavour to link the brain's underlying biology to MRI-based metrics. Integrating mechanistic science (including molecular, cellular, and animal models) with cognitive and neuroimaging data is, in my option, key to furthering our understanding of cognitive and brain health, decline and disease (see for example (Goriounova & Mansvelder, 2019). Increasingly, funding bodies and universities are supporting this form of integrative and interdisciplinary neuroscience (Waldman, 2013). For instance, in the US, the National Science Foundation initiated a new grant focusing on converging "boundary-crossing" scientific approaches to understand neural and cognitive systems[4]. Likewise, in the UK, King's College London is currently training PhD students in neuroscience *and* immunology, fostering a "new generation of interdisciplinary researchers" to better understand how inflammation causes neurodegenerative and psychiatric diseases [5].

I return to the question of mechanisms at the end of Section 5.4.1. My argument will be that serious interdisciplinary, mechanism-focused work can match the possibility of causal inference policy makers attribute primarily to RCTs, showing that mechanisms are not just important for science, but for science's translation into tangible policies affecting people's lives.

[4] https://beta.nsf.gov/funding/opportunities/integrative-strategies-understanding-neural-and-cognitive-systems-ncs. Recently awarded (for example) to Prof. Lilianne Mujica-Parodi to tackle the relationship between cognitive ageing and insulin resistance, a project combining human and animal models: https://www.nsf.gov/awardsearch/showAward?AWD_ID=1926781&HistoricalAwards=false

[5] https://www.kcl.ac.uk/study-legacy/postgraduate/research-courses/wellcome-trust-phd-training-programme-in-neuro-immune-interactions-in-health-and-disease

5.4 Evidence-based policy

Globally, the number of adults aged 65 or older is estimated to more than double by 2030, with dementia rates exponentially increasing from 1-2% of the population for those age 65, to 58% for those age 94 (Corrada et al., 2010). It is within this demographic shift that the science of healthy aging finds its urgency: identifying risk and protective factors for – and delaying the onset of – cognitive decline and dementia is not just important for affected individuals, but for the societies and nations in which they live.

It might be customary to, at this point of the General Discussion, highlight the promising ways in which this thesis should impact policy, thereby enabling people to age more healthily. However, during the two policy internships I completed during my PhD – one with the Chief Scientific Advisor at the Department of International Development (DFID), the other with the Healthy Aging team at the World Health Organization (WHO) – I was able to gain some insight into the long and complicated process by which research findings are translated into tangible policy recommendations. In the following sections, I would like to discuss this process in some depth: my hope is to enlighten the reader on its challenges, and to provide a balanced view on the role rigorous but fundamental and observational science – such as what I have presented in Chapters 2-4 – might, realistically, play in the science-to-policy pipeline.

5.4.1 Evidence hierarchies: when is science good enough for policy?

5.4.1.1 How policy makers evaluate scientific evidence

One of the WHO's core tasks is to assess the strength of available evidence regarding health policy issues [6]. These recommendations are read and implemented by policy

[6] https://www.who.int/about/who-we-are/our-values

makers, health care providers and individuals, often affecting the lives (and sometimes survival) of millions of people, while potentially costing (often developing) nations large sums of money. Ensuring that the recommendations are based on high-quality science is therefore essential. In 2007, the WHO adopted the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) system to support the formulation of evidence-based recommendations (G. Guyatt et al., 2011; Oxman et al., 2007). This framework, developed by epidemiologists and statisticians, considers several aspects including study design (randomized controlled trials or observational studies), risk of bias, inconsistency of results, indirectness of evidence, imprecision of measures, and risk of reporting bias. The quality of evidence for or against a recommendation is then characterized as either high, moderate, low, or very low, as explained in detail in the *WHO Handbook for Guideline Development* (WHO, 2019) [7].

What struck me during my internship is that high-quality recommendations are generally based on evidence from randomized controlled trials – in fact, mainly on quality-assessed *systematic reviews of RCTs* [8]. The WHO assesses (and often conducts) these systematic reviews, and evaluates their quality based on factors such as the literature search strategy, risk of bias of individual studies, and the statistical methods used to combine the assessed RCTs results. According to GRADE, observational studies generally provide low quality evidence for (or against) a recommendation – this is because cause and effect cannot usually be reliably inferred. There are, however, three circumstances under which GRADE recommends upgrading the quality of evidence from observational studies (G. H. Guyatt et al., 2011): i) when there is a large or a very

[7] https://apps.who.int/iris/handle/10665/145714

[8] The quality assessment of systematic reviews of RCTs is often based on the AMSTAR Checklist (Shea et al., 2007), which is available online, allowing anyone to easily assess the quality of systematic reviews: https://amstar.ca/Amstar_Checklist.php

large magnitude of effect, ii) when consideration of all plausible residual confounders and biases would reduce a demonstrated effect, or suggest a spurious effect when results show no effect, or iii) when there is evidence of a dose–response gradient.

An example of the first condition, a large effect size, is the relationship between infant sleeping position and sudden infant death syndrome (SIDS): one meta-analysis of observational studies found an odds ratio (OR) of 4.1 of SIDS occurring with front vs. back sleeping positions. Moreover "back to sleep" campaigns that were started in the 1980s to encourage back sleeping were associated with a relative decline in the incidence of SIDS by 50-70 percent in multiple countries (Gilbert et al., 2005). Here, according to GRADE, the large effect size and additional population-based epidemiolocal evidence merit upgrading the quality of evidence despite there being no RCTs (which would, of course, be impossible and unethical to conduct). Similarly, but with regards to cognitive aging, while RCTs involving alcohol would be unethical, there is strong observational evidence suggesting that excessive alcohol consumption is a risk factor for cognitive decline and dementia (e.g., Sachdeva et al., 2016). The WHO's 2019 guidelines on "Risk Reduction of Cognitive Decline and Dementia[9]" therefore recommend interventions aimed at reducing or ceasing hazardous and harmful drinking, rating the quality of the evidence as "moderate" (rather than low, as is usually the case with observational evidence).

The second reason for which the GRADE framework recommends up-rating observational evidence is when plausible confounders or biases in an observational study would result in an *underestimate* of a possible treatment effect. For example, a systematic review of observational studies regarding the relationships between for-profit versus not-for-profit hospitals and health outcomes found higher death rates in for-profit hospitals. Since patients in non-for-profit hospitals are likely to be sicker than those in for-profit-hospitals, any residual confounding would bias the results against

[9] https://www.who.int/publications/i/item/risk-reduction-of-cognitive-decline-and-dementia

the non-for-profit hospitals. Thus, because biases should diminish the observed effect, the fact that it was reliably detected suggests that it is sufficiently large as to warrant a "moderate" rather than "low" quality rating.

The final reason for up-rating observational evidence, according to GRADE, is the presence of a dose-response gradient. Of interest to this thesis is, for example, a study assessing the dose of moderate-to-vigorous physical exercise (as measured in minutes per month via self-report) and executive cognitive function in N=2157 adults aged 60 to 85. Rather than finding their hypothesized linear dose-response association, the authors report an inverted u-shape, whereby "very high" levels of exercise were associated with lower cognitive abilities compared to "high" levels of exercise (Loprinzi et al., 2018); an effect also found in a different study by the same authors assessing the relationship between cardiovascular exercise and mortality (here, the highest exercise group did not have greater survival benefits than the second-highest exercise group; Loprinzi, 2015). Despite the inverted U-shape pattern, very high levels of exercise were still associated with better cognitive functions than very low, low, or moderate levels of exercise, suggesting that there is a beneficial dose-response gradient of exercise on cognition. This type of result, according to the GRADE framework, increases the likelihood of there being a causal effect, thus improving the quality of the evidence.

In summary, bodies like the WHO rely largely on (systematic reviews of) RCTs when translating science into policy recommendations, while observational studies are generally interpreted as low-quality evidence, except under certain conditions. Thus, based on current evidence hierarchies, the empirical contributions of this thesis would be ranked as "low" or "very low" quality evidence in favour of a relationship between modifiable lifestyle activities and better cognitive aging. I think that it is important for scientists to realistically gauge the possible policy impacts of their work and would argue that familiarising oneself with frameworks like GRADE can help do so. That said, the following two sections will i) discuss possible downsides of existing evidence hierarchies' focus on RCTs and ii) argue that the successful integration of observational

and mechanistic evidence can match the level of causal inference often attributed to RCTs.

5.4.1.2 Is the focus on RCTs warranted?

One of the most impactful RCTs to date was conducted by economists Michael Kremer and Edward Miguel who wanted to find out how to best improve children's school attendance in Kenya. The researchers found that it was not textbooks, flip charts, smaller teacher-to-student ratios, or the availability of school buses that helped. The only intervention which raised school attendance was treating intestinal worms in school-aged children (Miguel & Kremer, 2004). Three years later, at the 2007 World Economic Forum, Kremer launched *Deworm the World*, an initiative which has since been scaled and is estimated to have reached over 280 million children worldwide. The WHO published the guideline to regularly (i.e., preventatively) treat all school children in affected areas, and millions of dollars of tax-payers' money[10] have been spent on deworming drugs. The golden bullet was found: a simple, cheap, and effective way of improving health, raising school performance, and elevating the educational and employment prospects of millions of children. And all because of a single RCT (and some follow-up studies, e,g., Nga et al., 2011) Then, in 2012, a Cochrane review showed that deworming alone had no effect on growth, cognitive ability or school attendance (Taylor-Robinson et al., 2012), a finding supported by a follow-up review in 2015 (Taylor-Robinson et al., 2015). That same year, researchers from the London School of Hygiene and Tropical Medicine re-analysed the original study's data – spurring what became known in the press as the "Worm Wars" (e.g., Belluz, 2015; Boseley, 2015): treating children or teaching them about worms *did not* improve school attendance (Aiken et al., 2015; Davey et al., 2015). Miguel and Kremen's RCT results, the authors claimed, were caused by biased treatment of missing data. While the details of this dispute have been summarized elsewhere (Clemens & Sandefur, 2015; Majid et al., 2019), I think that the

[10] For example, the UK and US have committed USD 68 and 65 million in 2008 and 2010, respectively (Hotez, 2011).

core lesson from the "Worm Wars" is that the methodological choices researchers make when analysing their data greatly affect a study's results and conclusions (something Orben & Przybylski, 2019 have discussed in detail, for example) – and that this applies to RCTs just as much as it does to observational studies.

The deworming example – where global public health policies were implemented despite warnings from Cochrane reviews and largely because of one RCT – is an exception; as outlined above, usually systematic reviews of RCTs, not individual trials, are used to establish the evidence-base of a policy recommendation. Despite this, the story highlights a greater issue, namely that the value of RCT evidence depends not on a given RTC's reported result, but on the quality of its statistical analysis. For instance, one study found that fewer than half of 193 RCTs assessed statistical power, while all of them used statistical significance tests (Faulkner et al., 2008), suggesting that RCTs might be prone to being insufficiently powered (see also Tsang et al., 2009). Moreover, there is a risk that scholars and policy makers are biased towards RCTs because of their 'gold standard' promise of causal inference, and whether this lowers the methodological rigour with which RCTs are assessed. This concern is reflected, for example, in a 2010 review showing that 35 percent of Cochrane reviews of RCTs had included trials affected by outcome reporting bias, where only some (usually the significant) outcome variables were selected for publication (Kirkham et al., 2010). Sensitivity analyses showed that 19 percent of the included meta-analyses became non-significant after adjusting for other outcome variables, with an additional 26 percent having overestimated the treatment effect by 20 percent or more. Funding bodies and policy agencies are increasingly aware of the issue of reporting bias in RCTs: they have called for pre-registration requirements of RCTs (Huić et al., 2011; Vandenbroucke, 2015) and for all raw data to be made publicly available (Ghersi, 2008) – open science techniques adapted from non-RCT disciplines (e.g., Aguinis et al., 2020; Dienlin et al., 2021).

It is important to contextualize the risk of bias in RCTs and its implications for policy. First and foremost, large-scale reviews show that RCTs generally provide accurate estimates of average treatment effects for groups that differ only with respect to the

intervention of interest (Mansournia et al., 2017). Second, far more often than not, results from RCTs and observational studies complement each other: For example, a meta-analysis of 14 meta-analyses compared more than 1,000 pairs of observational studies and RCTs across 228 medical conditions and found that effect estimates from observational studies were not significantly different from those of RCTs. Third, as outlined earlier, frameworks which evaluate the evidence base of a given recommendation *do* assess studies based, amongst other factors, on their risk of bias (G. Guyatt et al., 2011). While improvements to this "bias filter" may be desirable, especially Cochrane reviews have been found to generally be of very high quality (Moseley et al., 2009; Petticrew et al., 2002).

To summarize, viewing systematic reviews of RCTs as the bedrock of science policy is, in my opinion, largely warranted: there are many examples where such reviews have provided strong evidence for or against a treatment or intervention – a level of causal evidence that can be difficult to obtain using other study designs. However, the risk of bias in (reviews of) RCTs, whereby treatment effects are being overstated, should not be ignored. Moreover, it is impossible or unethical to conduct RCTs in many areas of research; categorizing the quality of evidence most observational studies as "low" can, therefore, seem overly simplistic and punitive. These concerns have led academics and policy makers to question the focus on RCTs as the often *sole* 'gold standard' by which scientific findings should be translated into policy recommendations. In the final part of this section, I outline how mechanistic frameworks could supplement existing RCT-prioritising evidence hierarchies.

5.4.1.3 A mechanistic approach to evidence-based policy

Why *would* deworming cause wide-spread improvements to Kenyan school attendance? Surely, one would only expect to find this effect if the illnesses treated by deworming pills were children's main reason for missing school. This turned out not to be the case: the biggest predictor of pupils' school attendance was the reliability with which *teachers* showed up to school: a World Bank study, for example, found that almost a third of

teaching staff were usually absent (Sabarwal & Abu-Jawdeh, 2018). Teachers' absenteeism seemed to be the problem in schools in Kenya – not the worms.

I think the lesson to be learned here is that in addition to well-powered, high-quality studies (whether interventional or observational), there should be a strong, plausible mechanistic explanation for a given effect. While well-supported psychological, economic, or biological mechanisms are not *sufficient* to translate science into policy, they should form a *necessary* condition which complements the empirical evidence base underlying a given policy recommendation.

I am not the only one in favour of mechanistic explanations. According to the Russo-Williamson distinction, in order to establish a causal claim, evidence that X and Y are probabilistically related and evidence that there is a mechanism between them are both necessary (Russo & Williamson, 2007). Other philosophers of science have since formulated the proposal that evidence of mechanisms – such as animal-models, which evidence-based policy usually places at the bottom of evidence hierarchies – is as important for policy making as RCTs (B. Clarke et al., 2014; Grüne-Yanoff, 2016; Kincaid, 2012). This is because policy recommendations almost always target populations other than those in which they have been tested. Such extrapolative inferences, it is argued, cannot be based exclusively on the statistical evidence produced by methods higher up in the evidence hierarchies. Marchionni and Reijula recently extended this idea to a more formal account; whereby mechanistic evidence forms the causal pathway that connects a policy to an outcome (Marchionni & Reijula, 2019). They argue that knowledge of whether and when a given causal relationship remains stable relies on the variables that mediate and modulate it, and therefore define mechanistic evidence as evidence about those variables.

These philosophical concerns formalize what I hope to have shown throughout this thesis, and which I have summarized in Section 5.3 above: the science we do should not just include large-sample, diverse cohorts, be well-powered and employ high-quality, replicable methodologies – to truly advance our understanding of the aging mind and

brain, our work should be based on probable mechanistic hypotheses that are deeply rooted in transdisciplinary evidence.

5.4.2 Lifestyle, cognition, and policy

5.4.2.1 Current policy recommendations

Section 5.4.1 discussed the ways in which the policy agencies evaluate the evidence-base of given policy recommendations. Against that backdrop, I next summarize how the science of modifiable lifestyle activities and cognitive aging – which this thesis contributes to – has been translated into global public health policies.

Table 5-1 summarizes the WHO's recommendations on physical, social, and cognitive engagement, taken from the *Guidelines on risk reduction of cognitive decline and dementia* (WHO, 2019)[11]. The strongest recommendation is that adults should engage in physical activity, especially high levels of aerobic exercise, to reduce the risk of cognitive decline. This recommendation is based on four systematic reviews of RCTs (Barha et al., 2017; Barreto et al., 2018; Northey et al., 2018; Song et al., 2018), which, according to the report, provide moderate-quality evidence for a beneficial relationship between exercise and cognitive health for adults with normal cognition, and low-quality evidence for adults with MCI. For cognitive interventions, the report assessed the effects of "cognitive stimulation therapy" (CST; defined as participation in a range of activities aimed at improving cognitive and social functioning) and "cognitive training" (which refers to guided practice of specific standardized tasks designed to enhance cognitive functions). For CST in cognitively normal older adults, the evidence was extracted from one systematic review of 18 RCTs (Strout et al., 2016), while the report says no evidence

[11] Note that the report addresses a total of 12 intervention categories, but I focus on the three most relevant to this thesis. The fill list of intervention categories includes: physical activity, tobacco cessation, nutritional interventions, interventions for alcohol use disorders, cognitive interventions, social activity, weight management, as well as management of hypertension, diabetes, dyslipidaemia, depression, and hearing loss.

was available for the effects of CST in adults with MCI. Strout's review (which shows that half of the RCTs were effective in at least one cognitive domain) reports the results in narrative form (rather than by conducting a meta-analysis), leading the WHO to rate the quality of evidence as "low". Similarly, for cognitive training in healthy adults, evidence came from one systematic review which (after reviewing 31 RCTs with a meta-analysis) reported that cognitive training had a moderate positive effect on overall cognitive functioning (Chiu et al., 2017). Finally, for social engagement, the report concludes that the evidence extracted from one (narratively reporting) systematic review of RCTs examining healthy adults was limited and inconclusive (Kelly et al., 2017). Three of the 39 included RCTs were deemed eligible by the WHO, of which only one found that social activity interventions (here group activities in Finnish old people's homes) positively affected cognitive abilities (Pitkala et al., 2011). Therefore, no recommendation was made regarding social interventions. The report states that no recommendation against social engagement because of its potential other benefits to health and wellbeing.

Intervention type	Recommendation	Quality of evidence	Strength of recommendation	Type of studies used as evidence
Physical activity interventions	Physical activity should be recommended to adults with normal cognition to reduce the risk of cognitive decline.	Moderate	Strong	4 systematic reviews of RCTs
	Physical activity may be recommended to adults with mild cognitive impairment to reduce the risk of cognitive decline.	Low	Conditional	
Cognitive interventions	Cognitive training may be offered to older adults with normal cognition and with mild cognitive impairment to reduce the risk of cognitive decline and/or dementia.	Very low to low	Conditional	2 systematic reviews of RCTs

Social activity interventions	There is insufficient evidence for social activity and reduction of risk of cognitive decline/dementia.	NA	NA	1 systemic review of RCTs

Table 5-1: The WHO's physical, cognitive, and social intervention recommendations for better cognitive aging. Adapted from the 2019 *Guidelines on risk reduction of cognitive decline and dementia.*

It is understandable that bodies like the World Health Organization err on the side of caution when making a policy recommendation: interventions are often expensive, and it is important for the WHO's (often developing) member states to trust that spending taxpayer's money on healthy aging policies will indeed reduce cognitive decline and MCI/dementia in their populations. Still, the disregard of observational studies in the recommendations summarized above seems restrictive. In my view, when only few and/or low-quality systemic reviews of RCTs are available (as was the case with cognitive and social interventions), additionally assessing the observational evidence-base may be warranted. This was done, for example, in the Lancet Commission mentioned in Chapter 1 (Livingston et al., 2020). Here, large-sample observational studies and meta-analyses were included alongside RCTs. For cognitive engagement, the report discusses a meta-analysis from 22 longitudinal cohorts (total N ~ 30,000), which reports a summary odds ratio of dementia for high versus low engagement in mentally stimulating activities of 0.5 after controlling for other dementia predictors such as age, sex, general health, cerebrovascular disease, education, occupation and baseline cognition (Valenzuela & Sachdev, 2006). For social engagement, the Lancet Commission considers the wider (arguably mechanistic) context whereby social isolation increases the risk of hypertension, coronary heart disease and depression, all of which are themselves risk factors for dementia. Moreover, it includes a meta-analysis of longitudinal observational studies which concluded that dementia risk was elevated for people with more limited social activity participation (Kuiper et al., 2015). The report warns that the relatively short follow-up period in some studies precludes strong conclusions about the direction of causation. From this (wider than the WHO's considered) evidence-base, Livingston and colleagues conclude:

> "We recommend keeping cognitively, physically, and socially active in midlife and later life although little evidence exists for any single specific activity protecting against dementia […]. Although behaviour change is difficult and some associations might not be purely causal, individuals have a huge potential to reduce their dementia risk." (Livingston et al., 2020, p.413)

.. and warn that:

> "Although a need for more evidence is apparent, recommendations should not wait, as clear indications of ways to reduce the chances of developing dementia without causing harm will also lead to other health and wellbeing benefits." (Livingston et al., 2020, p.429)

In summary, the empirically most strongly supported policy recommendations rely on a large body of observational and interventional studies suggesting that physical activity, and especially strenuous aerobic exercise, improve cognitive abilities and reduce the risk of dementia. The evidence-base for socially and intellectually engaging activities is more ambiguous, leading the WHO to avoid making strong recommendations regarding cognitive and social interventions. The Lancet Commission, which, unlike the WHO, assesses observational evidence alongside RCTs, suggests that there is already strong-enough evidence in favour of social and intellectual engagement, and that governments should adopt policies targeting these activities. This divergence in recommendations between the WHO and Lancet Commission points to the greater issue (discussed in section 5.4.1), of the kinds of studies deemed 'good enough' for policy, and the extent to which RCTs fulfil their 'gold standard' promise of causal inference.

5.4.2.2 Policy implications of the present findings

I mentioned earlier that many scholars, including myself, have the tendency to, in talks, papers or grants, point to the promising policy or clinical implications of their findings – something the wider academic ecosystem (including many funding bodies) actively

encourages. Having outlined the kind of research that *does* get translated into policy above, I think this is not just exaggerative, but potentially dangerous as it inflates the importance of individual findings rather than realistically integrating them into the wider evidence-base.

Based on the WHO's (and other agencies') standards, this thesis is unlikely to make any direct empirical contributions to policy: the observational nature of the data assessed here means that causality cannot be readily inferred. Although more generous in its inclusion criteria, I doubt that a future Lancet Commission on the relationship between modifiable lifestyle activities and cognitive aging would consider the present findings helpful: Chapter 2 is entirely based on cross-sectional data, the longitudinal samples in Chapter 4 are too small, the follow-up periods too short. Thus, the only way this thesis' findings might realistically impact policy is indirectly, through future meta-analyses.

Perhaps this thesis' wider contributions, then, are of a different kind: I hope to have shown, to academics, some of the reality and complexity of evidence-based policy, and, to policy makers, how mechanically plausible, well-conducted observational studies can offer causally meaningful insight into the lifestyle-brain-cognition triangle.

References and Appendices

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A Systematic Review and Meta-Analysis of the Cognitive Correlates of Bilingualism. *Review of Educational Research*, *80*(2), 207–245. https://doi.org/10.3102/0034654310368803

Aguinis, H., Banks, G. C., Rogelberg, S. G., & Cascio, W. F. (2020). Actionable recommendations for narrowing the science-practice gap in open science. *Organizational Behavior and Human Decision Processes*, *158*, 27–35. https://doi.org/10.1016/j.obhdp.2020.02.007

Aiken, A. M., Davey, C., Hargreaves, J. R., & Hayes, R. J. (2015). Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: A pure replication. *International Journal of Epidemiology*, *44*(5), 1572–1580. https://doi.org/10.1093/ije/dyv127

Allard, S., Scardochio, T., Cuello, A. C., & Ribeiro-da-Silva, A. (2012). Correlation of cognitive performance and morphological changes in neocortical pyramidal neurons in aging. *Neurobiology of Aging*, *33*(7), 1466–1480. https://doi.org/10.1016/j.neurobiolaging.2010.10.011

Allen, J. S., Bruss, J., Brown, C. K., & Damasio, H. (2005). Normal neuroanatomical variation due to age: The major lobes and a parcellation of the temporal region. *Neurobiology of Aging*, *26*(9), 1245–1260. https://doi.org/10.1016/j.neurobiolaging.2005.05.023

Allen, N. J., & Barres, B. A. (2009). Glia—More than just brain glue. *Nature*, *457*(7230), 675–677. https://doi.org/10.1038/457675a

Alwin, D. F. (2009). History, cohorts, and patterns of cognitive aging. In *Aging and cognition: Research methodologies and empirical advances* (pp. 9–38). American Psychological Association. https://doi.org/10.1037/11882-001

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *The End of Theory*, 3.

Angevaren, M., Vanhees, L., Nooyens, A. C. J., Wendel-Vos, C. G. W., & Verschuren, W. M. M. (2010). Physical Activity and 5-Year Cognitive Decline in the Doetinchem Cohort Study. *Annals of Epidemiology*, *20*(6), 473–479. https://doi.org/10.1016/j.annepidem.2010.03.007

Ashburner, J., Barnes, G., Chen, C.-C., Daunizeau, J., Flandin, G., Friston, K., Gitelman, D., Glauche, V., Henson, R., Hutton, C., Jafarian, A., Kiebel, S., Kilner, J., Litvak, V., Mattout, J., Moran, R., Penny, W., Phillips, C., Razi, A., & Zeidman, P. (2021). *SPM12 Manual*.

Ashburner, J., & Friston, K. J. (2000). Voxel-Based Morphometry—The Methods. *NeuroImage*, *11*(6), 805–821. https://doi.org/10.1006/nimg.2000.0582

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018

Asparouhov, T., & Muthén, B. (2009). Exploratory Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 397–438. https://doi.org/10.1080/10705510903008204

Auger, M. L., & Floresco, S. B. (2015). Prefrontal Cortical GABA Modulation of Spatial Reference and Working Memory. *International Journal of Neuropsychopharmacology*, *18*(2), pyu013. https://doi.org/10.1093/ijnp/pyu013

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, *34*(28), 3661–3679. https://doi.org/10.1002/sim.6607

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word test: A robust estimate of verbal intelligence based on lexical decision. *British Journal of Clinical Psychology*, *32*(1), 55–65. https://doi.org/10.1111/j.2044-8260.1993.tb01027.x

Baltes, P. B., Cornelius, S. W., Spiro, A., Nesselroade, J. R., & Willis, S. L. (1980). Integration versus differentiation of fluid/crytallized intelligence in old age. *Developmental Psychology*, *16*(6), 625–635. https://doi.org/10.1037/0012-1649.16.6.625

Baltes, P. B., Staudinger, U. M., & Lindenberger, U. (1999). Lifespan psychology: Theory and application to intellectual functioning. *Annual Review of Psychology*, *50*(1), 471–507.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, *48*(1), 5–37. https://doi.org/10.1016/j.jsp.2009.10.001

Barha, C. K., Davis, J. C., Falck, R. S., Nagamatsu, L. S., & Liu-Ambrose, T. (2017). Sex differences in exercise efficacy to improve cognition: A systematic review and meta-analysis of randomized controlled trials in older humans. *Frontiers in Neuroendocrinology*, *46*, 71–85. https://doi.org/10.1016/j.yfrne.2017.04.002

Barnes, J., Bartlett, J. W., van de Pol, L. A., Loy, C. T., Scahill, R. I., Frost, C., Thompson, P., & Fox, N. C. (2009). A meta-analysis of hippocampal atrophy rates in Alzheimer's disease. *Neurobiology of Aging*, *30*(11), 1711–1723. https://doi.org/10.1016/j.neurobiolaging.2008.01.010

Barreto, M., Victor, C., Hammond, C., Eccles, A., Richins, M. T., & Qualter, P. (2021). Loneliness around the world: Age, gender, and cultural differences in loneliness. *Personality and Individual Differences*, *169*, 110066. https://doi.org/10.1016/j.paid.2020.110066

Bartrés-Faz, D., & Arenaza-Urquijo, E. M. (2011). Structural and Functional Imaging Correlates of Cognitive and Brain Reserve Hypotheses in Healthy and Pathological Aging. *Brain Topography*, *24*(3), 340. https://doi.org/10.1007/s10548-011-0195-9

Barulli, D., & Stern, Y. (2013). Efficiency, capacity, compensation, maintenance, plasticity: Emerging concepts in cognitive reserve. *Trends in Cognitive Sciences*, *17*(10), 10.1016/j.tics.2013.08.012. https://doi.org/10.1016/j.tics.2013.08.012

Beauchet, O., Celle, S., Roche, F., Bartha, R., Montero-Odasso, M., Allali, G., & Annweiler, C. (2013). Blood pressure levels and brain volume reduction: A systematic review and meta-analysis. *Journal of Hypertension*, *31*(8), 1502–1516. https://doi.org/10.1097/HJH.0b013e32836184b5

Belluz, J. (2015, July 24). *Worm wars: The fight tearing apart the global health community, explained*. Vox. https://www.vox.com/2015/7/24/9031909/worm-wars-explained

Benton, A. L. (1994). *Contributions to Neuropsychological Assessment: A Clinical Manual*. Oxford University Press.

Bethlehem, R. a. I., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., Adler, S., Alexopoulos, G. S., Anagnostou, E., Areces-Gonzalez, A., Astle, D. E., Auyeung, B., Ayub, M., Ball, G., Baron-Cohen, S., Beare, R., Bedford, S. A., Benegal, V., Beyer, F., … Alexander-Bloch, A. F. (2021). Brain charts for the human lifespan. *BioRxiv*, 2021.06.08.447489. https://doi.org/10.1101/2021.06.08.447489

Bielak, A. A. M., Anstey, K. J., Christensen, H., & Windsor, T. D. (2012). Activity engagement is related to level, but not change in cognitive ability across adulthood. *Psychology and Aging*, *27*(1), 219–228. https://doi.org/10.1037/a0024667

Bielak, A. A. M., Hughes, T. F., Small, B. J., & Dixon, R. A. (2007). It's Never Too Late to Engage in Lifestyle Activities: Significant Concurrent but not Change Relationships Between Lifestyle Activities and Cognitive Speed. *The Journals of Gerontology: Series B*, *62*(6), P331–P339. https://doi.org/10.1093/geronb/62.6.P331

Bishop, N. A., Lu, T., & Yankner, B. A. (2010). Neural mechanisms of ageing and cognitive decline. *Nature*, *464*(7288), 529–535. https://doi.org/10.1038/nature08983

Bollen, K. A., & Diamantopoulos, A. (2017). In defense of causal-formative indicators: A minority report. *Psychological Methods*, *22*(3), 581–596. https://doi.org/10.1037/met0000056

Bookstein, F. L. (2001). "Voxel-based morphometry" should not be used with imperfectly registered images. *NeuroImage*, *14*(6), 1454–1462. https://doi.org/10.1006/nimg.2001.0770

Borgeest, G. S., Henson, R. N., Kietzmann, T. C., Madan, C. R., Fox, T., Malpetti, M., Fuhrmann, D., Knights, E., Carlin, J. D., Cam-CAN, & Kievit, R. A. (2021). *A morphometric double dissociation: Cortical thickness is more related to aging; surface area is more related to cognition* (p. 2021.09.30.462545). bioRxiv. https://doi.org/10.1101/2021.09.30.462545

Borgeest, G. S., Henson, R. N., Shafto, M., Samu, D., Cam-CAN, & Kievit, R. A. (2020). Greater lifestyle engagement is associated with better age-adjusted cognitive abilities. *PLOS ONE*, *15*(5), e0230077. https://doi.org/10.1371/journal.pone.0230077

Boseley, S., & editor, S. B. H. (2015, July 22). New research debunks merits of global deworming programmes. *The Guardian*. https://www.theguardian.com/society/2015/jul/23/research-global-deworming-programmes

Bosma, H., Boxtel, M., Ponds, R., Jelicic, M., Houx, P., Metsemakers, J., & Jolles, J. (2003). *Engaged lifestyle and cognitive function in middle and old-aged, non-demented persons: A reciprocal association?* (Vol. 35). https://doi.org/10.1007/s00391-002-0080-y

Braak, H., Thal, D. R., Ghebremedhin, E., & Del Tredici, K. (2011). Stages of the Pathologic Process in Alzheimer Disease: Age Categories From 1 to 100 Years. *Journal of Neuropathology & Experimental Neurology*, *70*(11), 960–969. https://doi.org/10.1097/NEN.0b013e318232a379

Brandmaier, A. M., Prindle, J. J., McArdle, J. J., & Lindenberger, U. (2016). Theory-guided exploration with structural equation model forests. *Psychological Methods*, *21*(4), 566–582. https://doi.org/10.1037/met0000090

Brodman, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. https://wellcomecollection.org/works/vrnkkxtj

Brown, C. L., Robitaille, A., Zelinski, E. M., Dixon, R. A., Hofer, S. M., & Piccinin, A. M. (2016). Cognitive activity mediates the association between social activity and cognitive performance: A longitudinal study. *Psychology and Aging*, *31*(8), 831–846. https://doi.org/10.1037/pag0000134

Bunce, D., Batterham, P. J., Mackinnon, A. J., & Christensen, H. (2012). Depression, anxiety and cognition in community-dwelling adults aged 70 years and over. *Journal of Psychiatric Research*, *46*(12), 1662–1666. https://doi.org/10.1016/j.jpsychires.2012.08.023

Bunce, D., Tzur, M., Ramchurn, A., Gain, F., & Bond, F. W. (2008). Mental Health and Cognitive Function in Adults Aged 18 to 92 Years. *The Journals of Gerontology: Series B*, *63*(2), P67–P74. https://doi.org/10.1093/geronb/63.2.P67

Burgmans, S., Gronenschild, E. H. B. M., Fandakova, Y., Shing, Y. L., van Boxtel, M. P. J., Vuurman, E. F. P. M., Uylings, H. B. M., Jolles, J., & Raz, N. (2011). Age differences in speed of processing are partially mediated by differences in axonal integrity. *NeuroImage*, *55*(3), 1287–1297. https://doi.org/10.1016/j.neuroimage.2011.01.002

Burke, S. L., Hu, T., Naseh, M., Fava, N. M., O'Driscoll, J., Alvarez, D., Cottler, L. B., & Duara, R. (2019). Factors influencing attrition in 35 Alzheimer's Disease Centers across the USA: A longitudinal examination of the National Alzheimer's Coordinating Center's Uniform Data Set. *Aging Clinical and Experimental Research*, *31*(9), 1283–1297. https://doi.org/10.1007/s40520-018-1087-6

Butler, M., McCreedy, E., Nelson, V. A., Desai, P., Ratner, E., Fink, H. A., Hemmy, L. S., McCarten, J. R., Barclay, T. R., Brasure, M., Davila, H., & Kane, R. L. (2018). Does Cognitive Training Prevent Cognitive Decline? *Annals of Internal Medicine*, *168*(1), 63–68. https://doi.org/10.7326/M17-1531

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*, *28*(2), 193–213. https://doi.org/10.1016/0165-1781(89)90047-4

Bzdok, D., & Yeo, B. T. T. (2017). Inference in the age of big data: Future perspectives on neuroscience. *NeuroImage*, *155*, 549–564. https://doi.org/10.1016/j.neuroimage.2017.04.061

Cabeza, R., & Dennis, N. A. (2013). Frontal Lobes and Aging: Deterioration and Compensation. In *Principles of Frontal Lobe Function*. Oxford University Press. https://oxfordmedicine.com/view/10.1093/med/9780199837755.001.0001/med-9780199837755-chapter-44

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, *12*(1), 1–47. https://doi.org/10.1162/08989290051137585

Carlson, M. C., Parisi, J. M., Xia, J., Xue, Q.-L., Rebok, G. W., Bandeen-Roche, K., & Fried, L. P. (2012). Lifestyle activities and memory: Variety may be the spice of life. The women's health and aging study II. *Journal of the International Neuropsychological Society: JINS*, *18*(2), 286–294. https://doi.org/10.1017/S135561771100169X

Carlson, T., Goddard, E., Kaplan, D. M., Klein, C., & Ritchie, J. B. (2018). Ghosts in machine learning for cognitive neuroscience: Moving from data to theory. *NeuroImage*, *180*, 88–100. https://doi.org/10.1016/j.neuroimage.2017.08.019

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*(3), 153–193. https://doi.org/10.1037/h0059973

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action* (pp. xxii, 583). Houghton Mifflin.

Cattell, R. B., & Cattell, A. K. S. (1960). Measuring Intelligence with the Culture Fair Tests. *Institute for Personality and Ability Testing*.

Chan, D., Shafto, M., Kievit, R., Matthews, F., Spink, M., Valenzuela, M., Cam-CAN, & Henson, R. (2018). Lifestyle activities in mid-life contribute to cognitive reserve in late-life, independent of education, occupation and late-life activities. *BioRxiv*, 267831. https://doi.org/10.1101/267831

Chan, D., Shafto, M., Kievit, R., Matthews, F., Spink, M., Valenzuela, M., & Henson, R. (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

Chang, A. Y., Skirbekk, V. F., Tyrovolas, S., Kassebaum, N. J., & Dieleman, J. L. (2019). Measuring population ageing: An analysis of the Global Burden of Disease Study 2017. *The Lancet Public Health*, *4*(3), e159–e167. https://doi.org/10.1016/S2468-2667(19)30019-2

Chinta, S. J., Woods, G., Rane, A., Demaria, M., Campisi, J., & Andersen, J. K. (2015). Cellular senescence and the aging brain. *Experimental Gerontology*, *68*, 3–7. https://doi.org/10.1016/j.exger.2014.09.018

Chiu, H.-L., Chu, H., Tsai, J.-C., Liu, D., Chen, Y.-R., Yang, H.-L., & Chou, K.-R. (2017). The effect of cognitive-based training for the healthy older people: A meta-analysis of randomized controlled trials. *PloS One*, *12*(5), e0176742. https://doi.org/10.1371/journal.pone.0176742

Choudhury, S., Fishman, J. R., McGowan, M. L., & Juengst, E. T. (2014). Big data, open science and the brain: Lessons learned from genomics. *Frontiers in Human Neuroscience*, *8*, 239. https://doi.org/10.3389/fnhum.2014.00239

Chung, W.-S., Welsh, C. A., Barres, B. A., & Stevens, B. (2015). Do glia drive synaptic and cognitive impairment in disease? *Nature Neuroscience*, *18*(11), 1539–1545. https://doi.org/10.1038/nn.4142

Clare, L., Wu, Y.-T., Teale, J. C., MacLeod, C., Matthews, F., Brayne, C., Woods, B., & CFAS-Wales study team. (2017). Potentially modifiable lifestyle factors, cognitive reserve, and cognitive function in later life: A cross-sectional study. *PLOS Medicine*, *14*(3), e1002259. https://doi.org/10.1371/journal.pmed.1002259

Clarke, A., Taylor, K. I., Devereux, B., Randall, B., & Tyler, L. K. (2013). From Perception to Conception: How Meaningful Objects Are Processed over Time. *Cerebral Cortex*, *23*(1), 187–197. https://doi.org/10.1093/cercor/bhs002

Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the Evidence Hierarchy. *Topoi*, *33*(2), 339–360. https://doi.org/10.1007/s11245-013-9220-9

Clemens, M., & Sandefur, J. (2015). *Mapping the Worm Wars: What the Public Should Take Away from the Scientific Debate about Mass Deworming*. Center For Global Development. https://www.cgdev.org/blog/mapping-worm-wars-what-public-should-take-away-scientific-debate-about-mass-deworming

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

Colcombe, S. J., Erickson, K. I., Scalf, P. E., Kim, J. S., Prakash, R., McAuley, E., Elavsky, S., Marquez, D. X., Hu, L., & Kramer, A. F. (2006). Aerobic Exercise Training Increases Brain Volume in Aging Humans. *The Journals of Gerontology: Series A*, *61*(11), 1166–1170. https://doi.org/10.1093/gerona/61.11.1166

Colcombe, S., & Kramer, A. F. (2003). Fitness Effects on the Cognitive Function of Older Adults: A Meta-Analytic Study. *Psychological Science*, *14*(2), 125–130.

Cole, J. H., & Franke, K. (2017). Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers. *Trends in Neurosciences*, *40*(12), 681–690. https://doi.org/10.1016/j.tins.2017.10.001

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330–351.

Corrada, M. M., Brookmeyer, R., Paganini-Hill, A., Berlau, D., & Kawas, C. H. (2010). Dementia Incidence Continues to Increase with Age in the Oldest Old The 90+ Study. *Annals of Neurology*, *67*(1), 114–121. https://doi.org/10.1002/ana.21915

Counts, S. E., & Mufson, E. J. (2010). Noradrenaline activation of neurotrophic pathways protects against neuronal amyloid toxicity. *Journal of Neurochemistry*, *113*(3), 649–660. https://doi.org/10.1111/j.1471-4159.2010.06622.x

Cowansage, K. K., LeDoux, J. E., & Monfils, M.-H. (2010). Brain-derived neurotrophic factor: A dynamic gatekeeper of neural plasticity. *Current Molecular Pharmacology*, *3*(1), 12–29. https://doi.org/10.2174/1874467211003010012

Cox, S. R., Bastin, M. E., Ritchie, S. J., Dickie, D. A., Liewald, D. C., Muñoz Maniega, S., Redmond, P., Royle, N. A., Pattie, A., Valdés Hernández, M., Corley, J., Aribisala, B. S., McIntosh, A. M., Wardlaw, J. M., & Deary, I. J. (2018). Brain cortical characteristics of lifetime cognitive ageing. *Brain Structure and Function*, *223*(1), 509–518. https://doi.org/10.1007/s00429-017-1505-0

Cox, S. R., Lyall, D. M., Ritchie, S. J., Bastin, M. E., Harris, M. A., Buchanan, C. R., Fawns-Ritchie, C., Barbu, M. C., de Nooij, L., Reus, L. M., Alloza, C., Shen, X., Neilson, E., Alderson, H. L., Hunter, S., Liewald, D. C., Whalley, H. C., McIntosh, A. M., Lawrie, S. M., … Deary, I. J. (2019). Associations between vascular risk factors and brain MRI indices in UK Biobank. *European Heart Journal*, *40*(28), 2290–2300. https://doi.org/10.1093/eurheartj/ehz100

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. *The Journals of Gerontology: Series B*, *58*(5), P249–P255. https://doi.org/10.1093/geronb/58.5.P249

Dahl, M. J., Mather, M., Düzel, S., Bodammer, N. C., Lindenberger, U., Kühn, S., & Werkle-Bergner, M. (2019). Rostral locus coeruleus integrity is associated with better memory performance in older adults. *Nature Human Behaviour*, *3*(11), 1203–1214. https://doi.org/10.1038/s41562-019-0715-2

Davatzikos, C. (2004). Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage*, *23*(1), 17–20. https://doi.org/10.1016/j.neuroimage.2004.05.010

Davey, C., Aiken, A. M., Hayes, R. J., & Hargreaves, J. R. (2015). Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: A statistical replication of a cluster quasi-randomized stepped-wedge trial. *International Journal of Epidemiology*, *44*(5), 1581–1592. https://doi.org/10.1093/ije/dyv128

de Brabander, J. M., Kramers, R. J. K., & Uylings, H. B. M. (1998). Layer-specific dendritic regression of pyramidal cells with ageing in the human prefrontal cortex. *European Journal of Neuroscience*, *10*(4), 1261–1269. https://doi.org/10.1046/j.1460-9568.1998.00137.x

de Lange, A.-M. G., Bråthen, A. C. S., Grydeland, H., Sexton, C., Johansen-Berg, H., Andersson, J. L. R., Rohani, D. A., Nyberg, L., Fjell, A. M., & Walhovd, K. B. (2016). White matter integrity as a marker for cognitive plasticity in aging. *Neurobiology of Aging*, *47*, 74–82. https://doi.org/10.1016/j.neurobiolaging.2016.07.007

de Souto Barreto, P., Demougeot, L., Vellas, B., & Rolland, Y. (2018). Exercise Training for Preventing Dementia, Mild Cognitive Impairment, and Clinically Meaningful Cognitive Decline: A Systematic Review and Meta-analysis. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *73*(11), 1504–1511. https://doi.org/10.1093/gerona/glx234

Deppe, M., Marinell, J., Krämer, J., Duning, T., Ruck, T., Simon, O. J., Zipp, F., Wiendl, H., & Meuth, S. G. (2014). Increased cortical curvature reflects white matter atrophy in individual patients with early multiple sclerosis. *NeuroImage: Clinical*, *6*, 475–487. https://doi.org/10.1016/j.nicl.2014.02.012

Devlieger, I., & Rosseel, Y. (2017). Factor score path analysis: An alternative for SEM? *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *13*(Suppl 1), 31–38. https://doi.org/10.1027/1614-2241/a000130

Diamond, J. (2013). *The World Until Yesterday: What Can We Learn from Traditional Societies?* Penguin.

Dickstein, D. L., Kabaso, D., Rocher, A. B., Luebke, J. I., Wearne, S. L., & Hof, P. R. (2007). Changes in the structural complexity of the aged brain. *Aging Cell*, *6*(3), 275–284. https://doi.org/10.1111/j.1474-9726.2007.00289.x

Dienlin, T., Johannes, N., Bowman, N. D., Masur, P. K., Engesser, S., Kümpel, A. S., Lukito, J., Bier, L. M., Zhang, R., Johnson, B. K., Huskey, R., Schneider, F. M., Breuer, J., Parry, D. A., Vermeulen, I., Fisher, J. T., Banks, J., Weber, R., Ellis, D. A., … de Vreese, C. (2021). An Agenda for Open Science in Communication. *Journal of Communication*, *71*(1), 1–26. https://doi.org/10.1093/joc/jqz052

Dodds, C. M., Henson, R. N., Miller, S. R., & Nathan, P. J. (2013). Overestimation of the effects of the BDNF val66met polymorphism on episodic memory-related hippocampal function: A critique of a recent meta-analysis. *Neuroscience and Biobehavioral Reviews*, *37*(4), 739–741. https://doi.org/10.1016/j.neubiorev.2013.01.019

Driscoll, I., Davatzikos, C., An, Y., Wu, X., Shen, D., Kraut, M., & Resnick, S. M. (2009). Longitudinal pattern of regional brain volume change differentiates normal aging from MCI. *Neurology*, *72*(22), 1906–1913. https://doi.org/10.1212/WNL.0b013e3181a82634

Driscoll, I., Martin, B., An, Y., Maudsley, S., Ferrucci, L., Mattson, M. P., & Resnick, S. M. (2012). Plasma BDNF Is Associated with Age-Related White Matter Atrophy but Not with Cognitive Function in Older, Non-Demented Adults. *PLOS ONE*, *7*(4), e35217. https://doi.org/10.1371/journal.pone.0035217

Duszkiewicz, A. J., McNamara, C. G., Takeuchi, T., & Genzel, L. (2019). Novelty and Dopaminergic Modulation of Memory Persistence: A Tale of Two Systems. *Trends in Neurosciences*, *42*(2), 102–114. https://doi.org/10.1016/j.tins.2018.10.002

Eavani, H., Habes, M., Satterthwaite, T. D., An, Y., Hsieh, M.-K., Honnorat, N., Erus, G., Doshi, J., Ferrucci, L., Beason-Held, L. L., Resnick, S. M., & Davatzikos, C. (2018). Heterogeneity of structural and functional imaging patterns of advanced brain aging revealed via machine learning methods. *Neurobiology of Aging*, *71*, 41–50. https://doi.org/10.1016/j.neurobiolaging.2018.06.013

Eavani, H., Hsieh, M. K., An, Y., Erus, G., Beason-Held, L., Resnick, S., & Davatzikos, C. (2016). Capturing heterogeneous group differences using mixture-of-experts: Application to a study of aging. *NeuroImage*, *125*, 498–514. https://doi.org/10.1016/j.neuroimage.2015.10.045

Ecker, C., Marquand, A., Mourão-Miranda, J., Johnston, P., Daly, E. M., Brammer, M. J., Maltezos, S., Murphy, C. M., Robertson, D., Williams, S. C., & Murphy, D. G. M. (2010). Describing the Brain in Autism in Five Dimensions—Magnetic Resonance Imaging-Assisted Diagnosis of Autism Spectrum Disorder Using a Multiparameter Classification Approach. *Journal of Neuroscience*, *30*(32), 10612–10623. https://doi.org/10.1523/JNEUROSCI.5413-09.2010

Eerola, M., Huurre, T., & Aro, H. (2005). The problem of attrition in a Finnish longitudinal survey on depression. *European Journal of Epidemiology*, *20*(1), 113–120. https://doi.org/10.1007/s10654-004-1657-0

Ekman, P., & Friesen, W. V. (1976). Pictures of facial affect. *Consulting Psychologists Press.*

Elderkin-Thompson, V., Mintz, J., Haroon, E., Lavretsky, H., & Kumar, A. (2007). Executive dysfunction and memory in older patients with major and minor depression. *Archives of Clinical Neuropsychology*, *22*(2), 261–270. https://doi.org/10.1016/j.acn.2007.01.021

Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, *6*(4), 352–370. https://doi.org/10.1037/1082-989X.6.4.352

Enders, C. K., & Bandalos, D. L. (2001). The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(3), 430–457. https://doi.org/10.1207/S15328007SEM0803_5

Enders, C. K., & Mansolf, M. (2018). Assessing the fit of structural equation models with multiply imputed data. *Psychological Methods*, *23*(1), 76–93. https://doi.org/10.1037/met0000102

Erickson, K. I., Prakash, R. S., Voss, M. W., Chaddock, L., Heo, S., McLaren, M., Pence, B. D., Martin, S. A., Vieira, V. J., Woods, J. A., McAuley, E., & Kramer, A. F. (2010). Brain-Derived Neurotrophic Factor Is Associated with Age-Related Decline in Hippocampal Volume. *Journal of Neuroscience*, *30*(15), 5368–5375. https://doi.org/10.1523/JNEUROSCI.6251-09.2010

Erickson, K. I., Voss, M. W., Prakash, R. S., Basak, C., Szabo, A., Chaddock, L., Kim, J. S., Heo, S., Alves, H., White, S. M., Wojcicki, T. R., Mailey, E., Vieira, V. J., Martin, S. A., Pence, B. D., Woods, J. A., McAuley, E., & Kramer, A. F. (2011). Exercise training increases size of hippocampus and improves memory. *Proceedings of the National Academy of Sciences*, *108*(7), 3017–3022. https://doi.org/10.1073/pnas.1015950108

Esteban-Cornejo, I., Mora-Gonzalez, J., Cadenas-Sanchez, C., Contreras-Rodriguez, O., Verdejo-Román, J., Henriksson, P., Migueles, J. H., Rodriguez-Ayllon, M., Molina-García, P., Suo, C., Hillman, C. H., Kramer, A. F., Erickson, K. I., Catena, A., Verdejo-García, A., & Ortega, F. B. (2019). Fitness, cortical thickness and surface area in overweight/obese children: The mediating role of body composition and relationship with intelligence. *NeuroImage*, *186*, 771–781. https://doi.org/10.1016/j.neuroimage.2018.11.047

Fabrigoule, C., Letenneur, L., Dartigues, J. F., Zarrouk, M., Commenges, D., & Barberger-Gateau, P. (1995). Social and leisure activities and risk of dementia: A prospective longitudinal study. *Journal of the American Geriatrics Society*, *43*(5), 485–490.

Faulkner, C., Fidler, F., & Cumming, G. (2008). The value of RCT evidence depends on the quality of statistical analysis. *Behaviour Research and Therapy*, *46*(2), 270–281. https://doi.org/10.1016/j.brat.2007.12.001

Ferrer, E., Balluerka, N., & Widaman, K. F. (2008). Factorial invariance and the specification of second-order latent growth models. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *4*(1), 22–36. https://doi.org/10.1027/1614-2241.4.1.22

Ferrer, E., & McArdle, J. J. (2010). Longitudinal Modeling of Developmental Changes in Psychological Research. *Current Directions in Psychological Science*, *19*(3), 149–154. https://doi.org/10.1177/0963721410370300

Finkel, D., Reynolds, C. A., McArdle, J. J., Gatz, M., & Pedersen, N. L. (2003). Latent growth curve analyses of accelerating decline in cognitive abilities in late adulthood. *Developmental Psychology*, *39*(3), 535–550. https://doi.org/10.1037/0012-1649.39.3.535

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., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Agartz, I., Salat, D. H., Greve, D. N., Fischl, B., Dale, A. M., & Walhovd, K. B. (2009). High Consistency of Regional Cortical

Thinning in Aging across Multiple Samples. *Cerebral Cortex*, *19*(9), 2001–2012. https://doi.org/10.1093/cercor/bhn232

Fjell, A. M., Westlye, L. T., Amlien, I., Tamnes, C. K., Grydeland, H., Engvig, A., Espeseth, T., Reinvang, I., Lundervold, A. J., Lundervold, A., & Walhovd, K. B. (2015). High-expanding cortical regions in human development and evolution are related to higher intellectual abilities. *Cerebral Cortex (New York, N.Y.: 1991)*, *25*(1), 26–34. https://doi.org/10.1093/cercor/bht201

Flynn, J. R. (1996). What environmental factors affect intelligence: The relevance of IQ gains over time. In *The environment* (pp. 17–29). Ablex Publishing. https://doi.org/10.1016/B978-012638281-5/50024-3

Flynn, J. R. (2007). *What Is Intelligence?: Beyond the Flynn Effect*. Cambridge University Press.

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.

Fotenos, A. F., Snyder, A. Z., Girton, L. E., Morris, J. C., & Buckner, R. L. (2005). Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. *Neurology*, *64*(6), 1032–1039. https://doi.org/10.1212/01.WNL.0000154530.72969.11

Frank-Cannon, T. C., Alto, L. T., McAlpine, F. E., & Tansey, M. G. (2009). Does neuroinflammation fan the flame in neurodegenerative diseases? *Molecular Neurodegeneration*, *4*(1), 47. https://doi.org/10.1186/1750-1326-4-47

Fratiglioni, L., Paillard-Borg, S., & Winblad, B. (2004). An active and socially integrated lifestyle in late life might protect against dementia. *The Lancet Neurology*, *3*(6), 343–353. https://doi.org/10.1016/S1474-4422(04)00767-7

Fuhrmann, D., Nesbitt, D., Shafto, M., Rowe, J. B., Price, D., Gadie, A., & Kievit, R. A. (2019). Strong and specific associations between cardiovascular risk factors and white matter micro- and macrostructure in healthy aging. *Neurobiology of Aging*, *74*, 46–55. https://doi.org/10.1016/j.neurobiolaging.2018.10.005

Fung, H. H. (2013). Aging in Culture. *The Gerontologist*, *53*(3), 369–377. https://doi.org/10.1093/geront/gnt024

Ganguli, M. (2017). The times they are a-changin': Cohort effects in aging, cognition, and dementia. *International Psychogeriatrics*, *29*(3), 353–355. https://doi.org/10.1017/S1041610216001927

Gefen, T., Shaw, E., Whitney, K., Martersteck, A., Stratton, J., Rademaker, A., Weintraub, S., Mesulam, M.-M., & Rogalski, E. (2014). Longitudinal Neuropsychological Performance of Cognitive SuperAgers. *Journal of the American Geriatrics Society*, *62*(8), 1598–1600. https://doi.org/10.1111/jgs.12967

Gelman, A., & Carlin, J. (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science*, *9*(6), 641–651. https://doi.org/10.1177/1745691614551642

Gerrits, N. J. H. M., van Loenhoud, A. C., van den Berg, S. F., Berendse, H. W., Foncke, E. M. J., Klein, M., Stoffers, D., van der Werf, Y. D., & van den Heuvel, O. A. (2016). Cortical Thickness, Surface Area and Subcortical Volume Differentially Contribute to Cognitive Heterogeneity in Parkinson's Disease. *PLoS ONE*, *11*(2). https://doi.org/10.1371/journal.pone.0148852

Ghai, S. (2021). It's time to reimagine sample diversity and retire the WEIRD dichotomy. *Nature Human Behaviour*, *5*(8), 971–972. https://doi.org/10.1038/s41562-021-01175-9

Ghersi, D. (2008). Reporting the findings of clinical trials: A discussion paper. *Bulletin of the World Health Organization*, *2008*(6), 492–493. https://doi.org/10.2471/BLT.08.053769

Ghisletta, P., & McArdle, J. J. (2012). Latent Curve Models and Latent Change Score Models Estimated in R. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(4), 651–682. https://doi.org/10.1080/10705511.2012.713275

Ghisletta, P., Rabbitt, P., Lunn, M., & Lindenberger, U. (2012). Two thirds of the age-based changes in fluid and crystallized intelligence, perceptual speed, and memory in adulthood are shared. *Intelligence*, *40*(3), 260–268. https://doi.org/10.1016/j.intell.2012.02.008

Gianaros, P. J., Greer, P. J., Ryan, C. M., & Jennings, J. R. (2006). Higher blood pressure predicts lower regional grey matter volume: Consequences on short-term information processing. *NeuroImage*, *31*(2), 754–765. https://doi.org/10.1016/j.neuroimage.2006.01.003

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Gilbert, R., Salanti, G., Harden, M., & See, S. (2005). Infant sleeping position and the sudden infant death syndrome: Systematic review of observational studies and historical review of recommendations from 1940 to 2002. *International Journal of Epidemiology*, *34*(4), 874–887. https://doi.org/10.1093/ije/dyi088

Gill, S. J., Friedenreich, C. M., Sajobi, T. T., Longman, R. S., Drogos, L. L., Davenport, M. H., Tyndall, A. V., Eskes, G. A., Hogan, D. B., Hill, M. D., Parboosingh, J. S., Wilson, B. J., & Poulin, M. J. (2015). Association between Lifetime Physical Activity and Cognitive Functioning in Middle-Aged and Older Community Dwelling Adults: Results from the Brain in Motion Study. *Journal of the International Neuropsychological Society*, *21*(10), 816–830. https://doi.org/10.1017/S1355617715000880

Giovacchini, G., Giovannini, E., Borsò, E., Lazzeri, P., Riondato, M., Leoncini, R., Duce, V., Mansi, L., & Ciarmiello, A. (2019). The brain cognitive reserve hypothesis: A review with emphasis on the contribution of nuclear medicine neuroimaging techniques. *Journal of Cellular Physiology*, *234*(9), 14865–14872. https://doi.org/10.1002/jcp.28308

Goessl, V. C., Curtiss, J. E., & Hofmann, S. G. (2017). The effect of heart rate variability biofeedback training on stress and anxiety: A meta-analysis. *Psychological Medicine*, *47*(15), 2578–2586. https://doi.org/10.1017/S0033291717001003

Goldberger, A. S., & Goldberger, G., Arthur Stanley. (1991). *A Course in Econometrics*. Harvard University Press.

Goriounova, N. A., & Mansvelder, H. D. (2019). Genes, Cells and Brain Areas of Intelligence. *Frontiers in Human Neuroscience*, *13*. https://www.frontiersin.org/article/10.3389/fnhum.2019.00044

Gorski, J. A., Zeiler, S. R., Tamowski, S., & Jones, K. R. (2003). Brain-Derived Neurotrophic Factor Is Required for the Maintenance of Cortical Dendrites. *Journal of Neuroscience*, *23*(17), 6856–6865. https://doi.org/10.1523/JNEUROSCI.23-17-06856.2003

Gottesman, R. F., Schneider, A. L. C., Albert, M., Alonso, A., Bandeen-Roche, K., Coker, L., Coresh, J., Knopman, D., Power, M. C., Rawlings, A., Sharrett, A. R., Wruck, L. M., & Mosley, T. H. (2014). Midlife hypertension and 20-year cognitive change: The atherosclerosis risk in communities neurocognitive study. *JAMA Neurology*, *71*(10), 1218–1227. https://doi.org/10.1001/jamaneurol.2014.1646

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence Predicts Health and Longevity, but Why? *Current Directions in Psychological Science*, *13*(1), 1–4.

Gow, A. J., Corley, J., Starr, J. M., & Deary, I. J. (2012). Reverse causation in activity-cognitive ability associations: The Lothian Birth Cohort 1936. *Psychology and Aging*, *27*(1), 250–255. https://doi.org/10.1037/a0024144

Gow, A. J., Pattie, A., & Deary, I. J. (2017). Lifecourse Activity Participation From Early, Mid, and Later Adulthood as Determinants of Cognitive Aging: The Lothian Birth Cohort 1921. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *72*(1), 25–37. https://doi.org/10.1093/geronb/gbw124

Greenwood, P. M. (2000). The frontal aging hypothesis evaluated. *Journal of the International Neuropsychological Society*, *6*(6), 705–726. https://doi.org/10.1017/S1355617700666092

Grober, E., Hall, C. B., Lipton, R. B., Zonderman, A. B., Resnick, S. M., & Kawas, C. (2008). Memory impairment, executive dysfunction, and intellectual decline in preclinical Alzheimer's disease. *Journal of the International Neuropsychological Society: JINS*, *14*(2), 266–278. https://doi.org/10.1017/S1355617708080302

Groot, C., Loenhoud, A. C. van, Barkhof, F., Berckel, B. N. M. van, Koene, T., Teunissen, C. C., Scheltens, P., Flier, W. M. van der, & 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

Grüne-Yanoff, T. (2016). WHY BEHAVIOURAL POLICY NEEDS MECHANISTIC EVIDENCE. *Economics & Philosophy*, *32*(3), 463–483. https://doi.org/10.1017/S0266267115000425

Gunning-Dixon, F. M., Brickman, A. M., Cheng, J. C., & Alexopoulos, G. S. (2009). Aging of Cerebral White Matter: A Review of MRI Findings. *International Journal of Geriatric Psychiatry*, *24*(2), 109–117. https://doi.org/10.1002/gps.2087

Gurven, M., Fuerstenberg, E., Trumble, B., Stieglitz, J., Beheim, B., Davis, H., & Kaplan, H. (2017). Cognitive performance across the life course of Bolivian forager-farmers with limited schooling. *Developmental Psychology*, *53*(1), 160–176. https://doi.org/10.1037/dev0000175

Gutchess, A. H., Welsh, R. C., Hedden, T., Bangert, A., Minear, M., Liu, L. L., & Park, D. C. (2005). Aging and the neural correlates of successful picture encoding: Frontal activations compensate for decreased medial-temporal activity. *Journal of Cognitive Neuroscience*, *17*(1), 84–96.

Guyatt, G. H., Oxman, A. D., Sultan, S., Glasziou, P., Akl, E. A., Alonso-Coello, P., Atkins, D., Kunz, R., Brozek, J., Montori, V., Jaeschke, R., Rind, D., Dahm, P., Meerpohl, J., Vist, G., Berliner, E., Norris, S., Falck-Ytter, Y., Murad, M. H., & Schünemann, H. J. (2011). GRADE guidelines: 9. Rating up the quality of evidence. *Journal of Clinical Epidemiology*, *64*(12), 1311–1316. https://doi.org/10.1016/j.jclinepi.2011.06.004

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., deBeer, H., Jaeschke, R., Rind, D., Meerpohl, J., Dahm, P., & Schünemann, H. J. (2011). GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, *64*(4), 383–394. https://doi.org/10.1016/j.jclinepi.2010.04.026

Haenschel, C., Bittner, R. A., Waltz, J., Haertling, F., Wibral, M., Singer, W., Linden, D. E. J., & Rodriguez, E. (2009). Cortical Oscillatory Activity Is Critical for Working Memory as Revealed by Deficits in Early-Onset Schizophrenia. *Journal of Neuroscience*, *29*(30), 9481–9489. https://doi.org/10.1523/JNEUROSCI.1428-09.2009

Harada, C. N., Natelson Love, M. C., & Triebel, K. (2013). Normal Cognitive Aging. *Clinics in Geriatric Medicine*, *29*(4), 737–752. https://doi.org/10.1016/j.cger.2013.07.002

Harlow, J. M. (1848). Passage of an Iron Rod Through the Head. *The Boston Medical and Surgical Journal (1828-1851)*, *39*(20), 0_1.

Harrison, T. M., Weintraub, S., Mesulam, M.-M., & Rogalski, E. (2012). Superior Memory and Higher Cortical Volumes in Unusually Successful Cognitive Aging. *Journal of the International Neuropsychological Society*, *18*(06), 1081–1085. https://doi.org/10.1017/S1355617712000847

Hausman, D. M. (2015). *Valuing Health: Well-Being, Freedom, and Suffering*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190233181.001.0001

Heathers, J. A. J. (2014). Everything Hertz: Methodological issues in short-term frequency-domain HRV. *Frontiers in Physiology*, *5*, 177. https://doi.org/10.3389/fphys.2014.00177

Hebb, D. O. (1947). The effects of early experience on problem-solving at maturity. *Am. Psychol. 2*, 306–307.

Hedden, T., & Gabrieli, J. D. E. (2004). Insights into the ageing mind: A view from cognitive neuroscience. *Nature Reviews Neuroscience*, *5*(2), 87–96. https://doi.org/10.1038/nrn1323

Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Henrich, J., Heine, S. J., & Norenzayan, A. (2010b). Most people are not WEIRD. *Nature*, *466*(7302), 29–29. https://doi.org/10.1038/466029a

Henson, R. N., Campbell, K. L., Davis, S. W., Taylor, J. R., Emery, T., Erzinclioglu, S., & Kievit, R. A. (2016). Multiple determinants of lifespan memory differences. *Scientific Reports*, *6*(1), 32527. https://doi.org/10.1038/srep32527

Hertzog, C., Kramer, A. F., Wilson, R. S., & Lindenberger, U. (2008). Enrichment Effects on Adult Cognitive Development: Can the Functional Capacity of Older Adults Be Preserved and Enhanced? *Psychological Science in the Public Interest*, *9*(1), 1–65. https://doi.org/10.1111/j.1539-6053.2009.01034.x

Hodges, J. (1994). Neurological aspects of dementia and normal aging. *Dementia and Normal Aging. Cambridge University Pres*.

Hofer, E., Roshchupkin, G. V., Adams, H. H. H., Knol, M. J., Lin, H., Li, S., Zare, H., Ahmad, S., Armstrong, N. J., Satizabal, C. L., Bernard, M., Bis, J. C., Gillespie, N. A., Luciano, M., Mishra, A., Scholz, M., Teumer, A., Xia, R., Jian, X., … Seshadri, S. (2020). Genetic correlations and genome-wide associations of cortical structure in general population samples of 22,824 adults. *Nature Communications*, *11*(1), 4796. https://doi.org/10.1038/s41467-020-18367-y

Hogstrom, L. J., Westlye, L. T., Walhovd, K. B., & Fjell, A. M. (2013). The structure of the cerebral cortex across adult life: Age-related patterns of surface area, thickness, and gyrification. *Cerebral Cortex (New York, N.Y.: 1991)*, *23*(11), 2521–2530. https://doi.org/10.1093/cercor/bhs231

Honda, J., & Miyamoto, H. (2020). *Would Population Aging Change the Output Effects of Fiscal Policy?* IMF. https://www.imf.org/en/Publications/WP/Issues/2020/06/12/Would-Population-Aging-Change-the-Output-Effects-of-Fiscal-Policy-49274

Hopkins, M. E., Nitecki, R., & Bucci, D. J. (2011). Physical exercise during adolescence versus adulthood: Differential effects on object recognition memory and brain-derived neurotrophic factor levels. *Neuroscience*, *194*, 84–94. https://doi.org/10.1016/j.neuroscience.2011.07.071

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, *26*, 107–129. https://doi.org/10.1016/0001-6918(67)90011-X

Hotez, P. J. (2011). New antipoverty drugs, vaccines, and diagnostics: A research agenda for the US President's Global Health Initiative (GHI). *PLoS Neglected Tropical Diseases*, *5*(5), e1133. https://doi.org/10.1371/journal.pntd.0001133

Howieson, D. B., Carlson, N. E., Moore, M. M., Wasserman, D., Abendroth, C. D., Payne-Murphy, J., & Kaye, J. A. (2008). Trajectory of mild cognitive impairment onset. *Journal of the International Neuropsychological Society*, *14*(2), 192–198. https://doi.org/10.1017/S1355617708080375

Huggett, D. L., Connelly, D. M., & Overend, T. J. (2005). Maximal Aerobic Capacity Testing of Older Adults: A Critical Review. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *60*(1), 57–66. https://doi.org/10.1093/gerona/60.1.57

Huić, M., Marušić, M., & Marušić, A. (2011). Completeness and Changes in Registered Data and Reporting Bias of Randomized Controlled Trials in ICMJE Journals after Trial Registration Policy. *PLOS ONE*, *6*(9), e25258. https://doi.org/10.1371/journal.pone.0025258

Hurd, M. D., Martorell, P., Delavande, A., Mullen, K. J., & Langa, K. M. (2013). Monetary Costs of Dementia in the United States. *New England Journal of Medicine*, *368*(14), 1326–1334. https://doi.org/10.1056/NEJMsa1204629

Hutton, C., Draganski, B., Ashburner, J., & Weiskopf, N. (2009). A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging. *NeuroImage*, *48*(2), 371–380. https://doi.org/10.1016/j.neuroimage.2009.06.043

Iacobucci, D. (2012). Mediation analysis and categorical variables: The final frontier. *Journal of Consumer Psychology*, *22*(4), 582–594. https://doi.org/10.1016/j.jcps.2012.03.006

Im, K., Lee, J.-M., Won Seo, S., Hyung Kim, S., Kim, S. I., & Na, D. L. (2008). Sulcal morphology changes and their relationship with cortical thickness and gyral white matter volume in mild cognitive impairment and Alzheimer's disease. *NeuroImage*, *43*(1), 103–113. https://doi.org/10.1016/j.neuroimage.2008.07.016

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Jack, C. R., Weigand, S. D., Shiung, M. M., Przybelski, S. A., O'Brien, P. C., Gunter, J. L., Knopman, D. S., Boeve, B. F., Smith, G. E., & Petersen, R. C. (2008). Atrophy rates accelerate in amnestic mild cognitive impairment. *Neurology*, *70*(19 Part 2), 1740–1752. https://doi.org/10.1212/01.wnl.0000281688.77598.35

Jacobs, B., Schall, M., Prather, M., Kapler, E., Driscoll, L., Baca, S., Jacobs, J., Ford, K., Wainwright, M., & Treml, M. (2001). Regional Dendritic and Spine Variation in Human Cerebral Cortex: A Quantitative Golgi Study. *Cerebral Cortex*, *11*(6), 558–571. https://doi.org/10.1093/cercor/11.6.558

Jacobsen, E., Ran, X., Liu, A., Chang, C.-C. H., & Ganguli, M. (2021). Predictors of attrition in a longitudinal population-based study of aging. *International Psychogeriatrics*, *33*(8), 767–778. https://doi.org/10.1017/S1041610220000447

Jacobucci, R., Brandmaier, A. M., & Kievit, R. A. (2019). A Practical Guide to Variable Selection in Structural Equation Modeling by Using Regularized Multiple-Indicators, Multiple-Causes Models. *Advances in Methods and Practices in Psychological Science*, *2*(1), 55–76. https://doi.org/10.1177/2515245919826527

Jan, Y.-N., & Jan, L. Y. (2010). Branching out: Mechanisms of dendritic arborization. *Nature Reviews Neuroscience*, *11*(5), 316–328. https://doi.org/10.1038/nrn2836

Jiang, J., Sachdev, P., Lipnicki, D. M., Zhang, H., Liu, T., Zhu, W., Suo, C., Zhuang, L., Crawford, J., Reppermund, S., Trollor, J., Brodaty, H., & Wen, W. (2014). A longitudinal study of brain atrophy over two years in community-dwelling older individuals. *NeuroImage*, *86*, 203–211. https://doi.org/10.1016/j.neuroimage.2013.08.022

Jin, K., Zhang, T., Shaw, M., Sachdev, P., & Cherbuin, N. (2018). Relationship Between Sulcal Characteristics and Brain Aging. *Frontiers in Aging Neuroscience*, *10*. https://doi.org/10.3389/fnagi.2018.00339

Joshi, S., & Gold, J. I. (2020). Pupil Size as a Window on Neural Substrates of Cognition. *Trends in Cognitive Sciences*, *24*(6), 466–480. https://doi.org/10.1016/j.tics.2020.03.005

Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavioral and Brain Sciences*, *30*(02), 135. https://doi.org/10.1017/S0140525X07001185

Kämpfen, F., Wijemunige, N., & Evangelista, B. (2018). Aging, non-communicable diseases, and old-age disability in low- and middle-income countries: A challenge for global health. *International Journal of Public Health*, *63*(9), 1011–1012. https://doi.org/10.1007/s00038-018-1137-z

Kane, M.D., R. L., Butler, Ph.D., M.B.A., M., & Fink, M.D., M.P.H., H. A. (2017). *Interventions To Prevent Age-Related Cognitive Decline, Mild Cognitive Impairment, and Clinical Alzheimer's-Type Dementia*. Agency for Healthcare Research and Quality (AHRQ). https://doi.org/10.23970/AHRQEPCCER188

Karp, A., Paillard-Borg, S., Wang, H.-X., Silverstein, M., Winblad, B., & Fratiglioni, L. (2006). Mental, Physical and Social Components in Leisure Activities Equally Contribute to Decrease Dementia Risk. *Dementia and Geriatric Cognitive Disorders*, *21*(2), 65–73. https://doi.org/10.1159/000089919

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

Kelley, A. S., McGarry, K., Gorges, R., & Skinner, J. S. (2015). The Burden of Health Care Costs in the Last 5 Years of Life. *Annals of Internal Medicine*, *163*(10), 729–736. https://doi.org/10.7326/M15-0381

Kelly, M. E., Duff, H., Kelly, S., McHugh Power, J. E., Brennan, S., Lawlor, B. A., & Loughrey, D. G. (2017). The impact of social activities, social networks, social support and social relationships on the cognitive functioning of healthy older adults: A systematic review. *Systematic Reviews*, *6*(1), 259. https://doi.org/10.1186/s13643-017-0632-2

Kievit, R. A., Brandmaier, A. M., Ziegler, G., van Harmelen, A.-L., de Mooij, S. M. M., Moutoussis, M., Goodyer, I. M., Bullmore, E., Jones, P. B., Fonagy, P., Lindenberger, U., & Dolan, R. J. (2018). Developmental cognitive neuroscience using latent change score models: A tutorial and

applications. *Developmental Cognitive Neuroscience*, *33*, 99–117. https://doi.org/10.1016/j.dcn.2017.11.007

Kincaid, H. (2012). *The Oxford Handbook of Philosophy of Social Science*. OUP USA.

Kinsella, G. J., Mullaly, E., Rand, E., Ong, B., Burton, C., Price, S., Phillips, M., & Storey, E. (2009). Early intervention for mild cognitive impairment: A randomised controlled trial. *Journal of Neurology, Neurosurgery & Psychiatry*, *80*(7), 730–736. https://doi.org/10.1136/jnnp.2008.148346

Kirkham, J. J., Dwan, K. M., Altman, D. G., Gamble, C., Dodd, S., Smyth, R., & Williamson, P. R. (2010). The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ*, *340*, c365. https://doi.org/10.1136/bmj.c365

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

Klein, A., Ghosh, S. S., Forrest, B. S., Giard, J., Haeme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E. C., & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, *13*(2).

Klein, A., & Tourville, J. (2012). 101 Labeled Brain Images and a Consistent Human Cortical Labeling Protocol. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00171

Kobilo, T., Liu, Q.-R., Gandhi, K., Mughal, M., Shaham, Y., & van Praag, H. (2011). Running is the neurogenic and neurotrophic stimulus in environmental enrichment. *Learning & Memory*, *18*(9), 605–609. https://doi.org/10.1101/lm.2283011

Kochunov, P., Mangin, J.-F., Coyle, T., Lancaster, J., Thompson, P., Rivière, D., Cointepas, Y., Régis, J., Schlosser, A., Royall, D. R., Zilles, K., Mazziotta, J., Toga, A., & Fox, P. T. (2005). Age-related morphology trends of cortical sulci. *Human Brain Mapping*, *26*(3), 210–220. https://doi.org/10.1002/hbm.20198

Koenig, J., Abler, B., Agartz, I., Åkerstedt, T., Andreassen, O. A., Anthony, M., Bär, K.-J., Bertsch, K., Brown, R. C., Brunner, R., Carnevali, L., Critchley, H. D., Cullen, K. R., de Geus, E. J. C., de la Cruz, F., Dziobek, I., Ferger, M. D., Fischer, H., Flor, H., … Quintana, D. S. (2021). Cortical thickness and resting-state cardiac function across the lifespan: A cross-sectional pooled mega-analysis. *Psychophysiology*, *58*(7), e13688. https://doi.org/10.1111/psyp.13688

Köhncke, Y., Laukka, E. J., Brehmer, Y., Kalpouzos, G., Li, T.-Q., Fratiglioni, L., Bäckman, L., & Lövdén, M. (2016). Three-year changes in leisure activities are associated with concurrent changes in white matter microstructure and perceptual speed in individuals aged 80 years and older. *Neurobiology of Aging*, *41*, 173–186. https://doi.org/10.1016/j.neurobiolaging.2016.02.013

Kralj, C., Daskalopoulou, C., Rodriguez-Artalejo, F., & Garcia-Esquinas, E. (2018). *Healthy ageing: A systematic review of risk factors – Athlos Project*. http://athlosproject.eu/healthy-ageing-a-systematic-review-of-risk-factors/

Kramer, A. F., & Erickson, K. I. (2007). Capitalizing on cortical plasticity: Influence of physical activity on cognition and brain function. *Trends in Cognitive Sciences*, *11*(8), 342–348. https://doi.org/10.1016/j.tics.2007.06.009

Kramer, A. F., Erickson, K. I., & Colcombe, S. J. (2006). Exercise, cognition, and the aging brain. *Journal of Applied Physiology*, *101*(4), 1237–1242. https://doi.org/10.1152/japplphysiol.00500.2006

Kremen, W. S., Beck, A., Elman, J. A., Gustavson, D. E., Reynolds, C. A., Tu, X. M., Sanderson-Cimino, M. E., Panizzon, M. S., Vuoksimaa, E., Toomey, R., Fennema-Notestine, C., Hagler, D. J., Fang, B., Dale, A. M., Lyons, M. J., & Franz, C. E. (2019). Influence of young adult cognitive ability and additional education on later-life cognition. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(6), 2021–2026. https://doi.org/10.1073/pnas.1811537116

Ku, P.-W., Fox, K. R., Chang, C.-Y., Sun, W.-J., & Chen, L.-J. (2014). Cross-Sectional and Longitudinal Associations of Categories of Physical Activities with Dimensions of Subjective Well-Being in

Taiwanese Older Adults. *Social Indicators Research*, *117*(3), 705–718. https://doi.org/10.1007/s11205-013-0394-8

Kuhn, T. S. (2011). The Essential Tension: Selected Studies in Scientific Tradition and Change. In *The Essential Tension*. University of Chicago Press. https://doi.org/10.7208/9780226217239

Kuiper, J. S., Zuidersma, M., Oude Voshaar, R. C., Zuidema, S. U., van den Heuvel, E. R., Stolk, R. P., & Smidt, N. (2015). Social relationships and risk of dementia: A systematic review and meta-analysis of longitudinal cohort studies. *Ageing Research Reviews*, *22*, 39–57. https://doi.org/10.1016/j.arr.2015.04.006

Kuntzelman, K. M., Williams, J. M., Lim, P. C., Samal, A., Rao, P. K., & Johnson, M. R. (2021). Deep-Learning-Based Multivariate Pattern Analysis (dMVPA): A Tutorial and a Toolbox. *Frontiers in Human Neuroscience*, *15*. https://www.frontiersin.org/article/10.3389/fnhum.2021.638052

Kunzmann, U., Little, T. D., & Smith, J. (2000). Is age-related stability of subjective well-being a paradox? Cross-sectional and longitudinal evidence from the Berlin Aging Study. *Psychology and Aging*, *15*(3), 511–526. https://doi.org/10.1037/0882-7974.15.3.511

Larson, E. B., Wang, L., Bowen, J. D., McCormick, W. C., Teri, L., Crane, P., & Kukull, W. (2006). Exercise is associated with reduced risk for incident dementia among persons 65 years of age and older. *Annals of Internal Medicine*, *144*(2), 73–81. https://doi.org/10.7326/0003-4819-144-2-200601170-00004

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

Launer, L. J., Masaki, K., Petrovitch, H., Foley, D., & Havlik, R. J. (1995). The association between midlife blood pressure levels and late-life cognitive function. The Honolulu-Asia Aging Study. *JAMA*, *274*(23), 1846–1851.

Lee, J. S., Shin, H. Y., Kim, H. J., Jang, Y. K., Jung, N.-Y., Lee, J., Kim, Y. J., Chun, P., Yang, J.-J., Lee, J.-M., Kang, M., Park, K.-C., Na, D. L., & Seo, S. W. (2016). Combined effects of physical exercise and education on age-related cortical thinning in cognitively normal individuals. *Scientific Reports*, *6*, 24284. https://doi.org/10.1038/srep24284

Lee, R., & Mason, A. (2017). Cost of Aging. *Finance & Development*, *54*(1), 7–9.

Lehert, P., Villaseca, P., Hogervorst, E., Maki, P. M., & Henderson, V. W. (2015). Individually modifiable risk factors to ameliorate cognitive aging: A systematic review and meta-analysis. *Climacteric*, *18*(5), 678–689. https://doi.org/10.3109/13697137.2015.1078106

Lemaitre, H., Goldman, A. L., Sambataro, F., Verchinski, B. A., Meyer-Lindenberg, A., Weinberger, D. R., & Mattay, V. S. (2012). Normal age-related brain morphometric changes: Nonuniformity across cortical thickness, surface area and gray matter volume? *Neurobiology of Aging*, *33*(3), 617.e1-617.e9. https://doi.org/10.1016/j.neurobiolaging.2010.07.013

Lewis, D. A., Curley, A. A., Glausier, J. R., & Volk, D. W. (2012). Cortical parvalbumin interneurons and cognitive dysfunction in schizophrenia. *Trends in Neurosciences*, *35*(1), 57–67. https://doi.org/10.1016/j.tins.2011.10.004

Lezak, Muriel, & Deutsch. (1995). *Neuropsychological assessment*. Oxford University Press.

Lin, F. V., Wang, X., Wu, R., Rebok, G. W., Chapman, B. P., & Initiative, the A. D. N. (2017). Identification of Successful Cognitive Aging in the Alzheimer's Disease Neuroimaging Initiative Study. *Journal of Alzheimer's Disease*, *59*(1), 101–111. https://doi.org/10.3233/JAD-161278

Lindenberger, U. (2014). Human cognitive aging: Corriger la fortune? *Science*, *346*(6209), 572–578.

Lindwall, M., Cimino, C. R., Gibbons, L. E., Mitchell, M. B., Benitez, A., Brown, C. L., Kennison, R. F., Shirk, S. D., Atri, A., Robitaille, A., MacDonald, S. W. S., Zelinski, E. M., Willis, S. L., Schaie, K. W., Johansson, B., Praetorius, M., Dixon, R. A., Mungas, D. M., Hofer, S. M., & Piccinin, A. M. (2012). Dynamic Associations of Change in Physical Activity and Change in Cognitive Function:

Coordinated Analyses of Four Longitudinal Studies. *Journal of Aging Research*, *2012*, e493598. https://doi.org/10.1155/2012/493598

Little, T. D. (2013). *Longitudinal Structural Equation Modeling*. Guilford Press.

Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, E. W. G. (2014). On the Joys of Missing Data. *Journal of Pediatric Psychology*, *39*(2), 151–162. https://doi.org/10.1093/jpepsy/jst048

Liu, K. Y., Kievit, R. A., Tsvetanov, K. A., Betts, M. J., Düzel, E., Rowe, J. B., Howard, R., & Hämmerer, D. (2020). Noradrenergic-dependent functions are associated with age-related locus coeruleus signal intensity differences. *Nature Communications*, *11*(1), 1712. https://doi.org/10.1038/s41467-020-15410-w

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

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., Ballard, C., Banerjee, S., Burns, A., Cohen-Mansfield, J., Cooper, C., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Larson, E. B., Ritchie, K., Rockwood, K., Sampson, E. L., … Mukadam, N. (2017). Dementia prevention, intervention, and care. *The Lancet*, *390*(10113), 2673–2734. https://doi.org/10.1016/S0140-6736(17)31363-6

Lo, A. H. Y., Humphreys, M., Byrne, G. J., & Pachana, N. A. (2012). Test–retest reliability and practice effects of the Wechsler Memory Scale-III. *Journal of Neuropsychology*, *6*(2), 212–231. https://doi.org/10.1111/j.1748-6653.2011.02023.x

López Ulloa, B. F., Møller, V., & Sousa-Poza, A. (2013). How Does Subjective Well-Being Evolve with Age? A Literature Review. *Journal of Population Ageing*, *6*(3), 227–246. https://doi.org/10.1007/s12062-013-9085-0

Loprinzi, P. D. (2015). Dose-response association of moderate-to-vigorous physical activity with cardiovascular biomarkers and all-cause mortality: Considerations by individual sports, exercise and recreational physical activities. *Preventive Medicine*, *81*, 73–77. https://doi.org/10.1016/j.ypmed.2015.08.014

Loprinzi, P. D., Edwards, M. K., Crush, E., Ikuta, T., & Del Arco, A. (2018). Dose–Response Association Between Physical Activity and Cognitive Function in a National Sample of Older Adults. *American Journal of Health Promotion*, *32*(3), 554–560. https://doi.org/10.1177/0890117116689732

Lövdén, M., Bodammer, N. C., Kühn, S., Kaufmann, J., Schütze, H., Tempelmann, C., Heinze, H.-J., Düzel, E., Schmiedek, F., & Lindenberger, U. (2010). Experience-dependent plasticity of white-matter microstructure extends into old age. *Neuropsychologia*, *48*(13), 3878–3883. https://doi.org/10.1016/j.neuropsychologia.2010.08.026

Lövdén, M., Ghisletta, P., & Lindenberger, U. (2005). Social Participation Attenuates Decline in Perceptual Speed in Old and Very Old Age. *Psychology and Aging*, *20*(3), 423–434. https://doi.org/10.1037/0882-7974.20.3.423

Lövdén, M., Wenger, E., Mårtensson, J., Lindenberger, U., & Bäckman, L. (2013a). Structural brain plasticity in adult learning and development. *Neuroscience & Biobehavioral Reviews*, *37*(9, Part B), 2296–2310. https://doi.org/10.1016/j.neubiorev.2013.02.014

Lövdén, M., Wenger, E., Mårtensson, J., Lindenberger, U., & Bäckman, L. (2013b). Structural brain plasticity in adult learning and development. *Neuroscience & Biobehavioral Reviews*, *37*(9, Part B), 2296–2310. https://doi.org/10.1016/j.neubiorev.2013.02.014

Luque-Casado, A., Zabala, M., Morales, E., Mateo-March, M., & Sanabria, D. (2013). Cognitive Performance and Heart Rate Variability: The Influence of Fitness Level. *PLOS ONE*, *8*(2), e56935. https://doi.org/10.1371/journal.pone.0056935

MacCallum, R. C., & Austin, J. T. (2000). Applications of Structural Equation Modeling in Psychological Research. *Annual Review of Psychology*, *51*(1), 201–226. https://doi.org/10.1146/annurev.psych.51.1.201

Madan, C. R. (2021). Age-related decrements in cortical gyrification: Evidence from an accelerated longitudinal dataset. *The European Journal of Neuroscience*, *53*(5), 1661–1671. https://doi.org/10.1111/ejn.15039

Madan, C. R., & Kensinger, E. A. (2016). Cortical complexity as a measure of age-related brain atrophy. *NeuroImage*, *134*, 617–629.

Majid, M. F., Kang, S. J., & Hotez, P. J. (2019). Resolving "worm wars": An extended comparison review of findings from key economics and epidemiological studies. *PLOS Neglected Tropical Diseases*, *13*(3), e0006940. https://doi.org/10.1371/journal.pntd.0006940

Makizako, H., Liu-Ambrose, T., Shimada, H., Doi, T., Park, H., Tsutsumimoto, K., Uemura, K., & Suzuki, T. (2015). Moderate-Intensity Physical Activity, Hippocampal Volume, and Memory in Older Adults With Mild Cognitive Impairment. *The Journals of Gerontology: Series A*, *70*(4), 480–486. https://doi.org/10.1093/gerona/glu136

Manaye, K. F., McIntire, D. D., Mann, D. M. A., & German, D. C. (1995). Locus coeruleus cell loss in the aging human brain: A non-random process. *Journal of Comparative Neurology*, *358*(1), 79–87. https://doi.org/10.1002/cne.903580105

Mansournia, M. A., Higgins, J. P. T., Sterne, J. A. C., & Hernán, M. A. (2017). Biases in Randomized Trials: A Conversation Between Trialists and Epidemiologists. *Epidemiology (Cambridge, Mass.)*, *28*(1), 54–59. https://doi.org/10.1097/EDE.0000000000000564

Marchionni, C., & Reijula, S. (2019). What is mechanistic evidence, and why do we need it for evidence-based policy? *Studies in History and Philosophy of Science Part A*, *73*, 54–63. https://doi.org/10.1016/j.shpsa.2018.08.003

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory Structural Equation Modeling: An Integration of the Best Features of Exploratory and Confirmatory Factor Analysis. *Annual Review of Clinical Psychology*, *10*(1), 85–110. https://doi.org/10.1146/annurev-clinpsy-032813-153700

Martínez, J. H., López, M. E., Ariza, P., Chavez, M., Pineda-Pardo, J. A., López-Sanz, D., Gil, P., Maestú, F., & Buldú, J. M. (2018). Functional brain networks reveal the existence of cognitive reserve and the interplay between network topology and dynamics. *Scientific Reports*, *8*(1), 10525. https://doi.org/10.1038/s41598-018-28747-6

Mather, M., & Harley, C. W. (2016). The Locus Coeruleus: Essential for Maintaining Cognitive Function and the Aging Brain. *Trends in Cognitive Sciences*, *20*(3), 214–226. https://doi.org/10.1016/j.tics.2016.01.001

Matsunaga, W., Shirokawa, T., & Isobe, K. (2004). BDNF is necessary for maintenance of noradrenergic innervations in the aged rat brain. *Neurobiology of Aging*, *25*(3), 341–348. https://doi.org/10.1016/S0197-4580(03)00093-9

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, *12*(1), 23–44. https://doi.org/10.1037/1082-989X.12.1.23

Maxwell, S. E., Cole, D. A., & Mitchell, M. A. (2011). Bias in Cross-Sectional Analyses of Longitudinal Mediation: Partial and Complete Mediation Under an Autoregressive Model. *Multivariate Behavioral Research*, *46*(5), 816–841. https://doi.org/10.1080/00273171.2011.606716

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist*, *70*(6), 487–498. https://doi.org/10.1037/a0039400

McArdle, J. J. (2012). Latent curve modeling of longitudinal growth data. In *Handbook of structural equation modeling* (pp. 547–570). The Guilford Press.

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, *38*(1), 115–142. https://doi.org/10.1037/0012-1649.38.1.115

McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In *New methods for the analysis of change* (pp. 139–175). American Psychological Association. https://doi.org/10.1037/10409-005

McArdle, J. J., & Nesselroade, J. R. (2003). Growth curve analysis in contemporary psychological research. In *Handbook of psychology: Research methods in psychology, Vol. 2.* (pp. 447–480). John Wiley & Sons Inc.

McClelland, G. H., Lynch, J. G., Irwin, J. R., Spiller, S. A., & Fitzsimons, G. J. (2015). Median splits, Type II errors, and false–positive consumer psychology: Don't fight the power. *Journal of Consumer Psychology*, *25*(4), 679–689. https://doi.org/10.1016/j.jcps.2015.05.006

McCormick, E. M. (2021). Multi-Level Multi-Growth Models: New opportunities for addressing developmental theory using advanced longitudinal designs with planned missingness. *Developmental Cognitive Neuroscience*, *51*, 101001. https://doi.org/10.1016/j.dcn.2021.101001

McDonald, C. R., McEvoy, L. K., Gharapetian, L., Fennema-Notestine, C., Hagler, D. J., Holland, D., Koyama, A., Brewer, J. B., & Dale, A. M. (2009). Regional rates of neocortical atrophy from normal aging to early Alzheimer disease. *Neurology*, *73*(6), 457–465. https://doi.org/10.1212/WNL.0b013e3181b16431

McKay, D. R., Knowles, E. E. M., Winkler, A. A. M., Sprooten, E., Kochunov, P., Olvera, R. L., Curran, J. E., Kent, J. W., Carless, M. A., Göring, H. H. H., Dyer, T. D., Duggirala, R., Almasy, L., Fox, P. T., Blangero, J., & Glahn, D. C. (2014). Influence of age, sex and genetic factors on the human brain. *Brain Imaging and Behavior*, *8*(2), 143–152. https://doi.org/10.1007/s11682-013-9277-5

Middleton, L. E., Barnes, D. E., Lui, L.-Y., & Yaffe, K. (2010). Physical Activity Over the Life Course and its Association with Cognitive Performance and Impairment in Old Age. *Journal of the American Geriatrics Society*, *58*(7), 1322–1326. https://doi.org/10.1111/j.1532-5415.2010.02903.x

Miguel, E., & Kremer, M. (2004). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*, *72*(1), 159–217. https://doi.org/10.1111/j.1468-0262.2004.00481.x

Millikan, R. G. (1989). In Defense of Proper Functions. *Philosophy of Science*, *56*(2), 288–302.

Miranda, M., Morici, J. F., Zanoni, M. B., & Bekinschtein, P. (2019). Brain-Derived Neurotrophic Factor: A Key Molecule for Memory in the Healthy and the Pathological Brain. *Frontiers in Cellular Neuroscience*, *0*. https://doi.org/10.3389/fncel.2019.00363

Mitchell, A. J., & Shiri-Feshki, M. (2009). Rate of progression of mild cognitive impairment to dementia—Meta-analysis of 41 robust inception cohort studies. *Acta Psychiatrica Scandinavica*, *119*(4), 252–265. https://doi.org/10.1111/j.1600-0447.2008.01326.x

Mitchell, M. B., Cimino, C. R., Benitez, A., Brown, C. L., Gibbons, L. E., Kennison, R. F., Shirk, S. D., Atri, A., Robitaille, A., MacDonald, S. W. S., Lindwall, M., Zelinski, E. M., Willis, S. L., Schaie, K. W., Johansson, B., Dixon, R. A., Mungas, D. M., Hofer, S. M., & Piccinin, A. M. (2012). Cognitively Stimulating Activities: Effects on Cognition across Four Studies with up to 21 Years of Longitudinal Data. *Journal of Aging Research*, *2012*. https://doi.org/10.1155/2012/461592

Morcom, A. M., & Henson, R. N. A. (2018). Increased Prefrontal Activity with Aging Reflects Nonspecific Neural Responses Rather than Compensation. *Journal of Neuroscience*, *38*(33), 7303–7313. https://doi.org/10.1523/JNEUROSCI.1701-17.2018

Morland, C., Andersson, K. A., Haugen, Ø. P., Hadzic, A., Kleppa, L., Gille, A., Rinholm, J. E., Palibrk, V., Diget, E. H., Kennedy, L. H., Stølen, T., Hennestad, E., Moldestad, O., Cai, Y., Puchades, M., Offermanns, S., Vervaeke, K., Bjørås, M., Wisløff, U., … Bergersen, L. H. (2017). Exercise induces cerebral VEGF and angiogenesis via the lactate receptor HCAR1. *Nature Communications*, *8*(1), 15557. https://doi.org/10.1038/ncomms15557

Mortimer, J. A. (1997). Brain reserve and the clinical expression of Alzheimer's disease. *Geriatrics*, *52 Suppl 2*, S50-53.

Moseley, A. M., Elkins, M. R., Herbert, R. D., Maher, C. G., & Sherrington, C. (2009). Cochrane reviews used more rigorous methods than non-Cochrane reviews: Survey of systematic reviews in physiotherapy. *Journal of Clinical Epidemiology*, *62*(10), 1021–1030. https://doi.org/10.1016/j.jclinepi.2008.09.018

Mukadam, N., Anderson, R., Knapp, M., Wittenberg, R., Karagiannidou, M., Costafreda, S. G., Tutton, M., Alessi, C., & Livingston, G. (2020). Effective interventions for potentially modifiable risk factors for late-onset dementia: A costs and cost-effectiveness modelling study. *The Lancet Healthy Longevity*, *1*(1), e13–e20. https://doi.org/10.1016/S2666-7568(20)30004-0

Muller, M., van der Graaf, Y., Visseren, F. L., Vlek, A. L., Mali, W. Pt., Geerlings, M. I., & Group, for the S. S. (2010). Blood pressure, cerebral blood flow, and brain volumes. The SMART-MR study. *Journal of Hypertension*, *28*(7), 1498–1505. https://doi.org/10.1097/HJH.0b013e32833951ef

Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, *35*(8), 4140–4154. https://doi.org/10.1002/hbm.22466

Mustroph, M. L., Chen, S., Desai, S. C., Cay, E. B., DeYoung, E. K., & Rhodes, J. S. (2012). Aerobic exercise is the critical variable in an enriched environment that increases hippocampal neurogenesis and water maze learning in male C57BL/6J mice. *Neuroscience*, *219*, 62–71. https://doi.org/10.1016/j.neuroscience.2012.06.007

Nakamura, S., Akiguchi, I., Kameyama, M., & Mizuno, N. (1985). Age-related changes of pyramidal cell basal dendrites in layers III and V of human motor cortex: A quantitative Golgi study. *Acta Neuropathologica*, *65*(3), 281–284. https://doi.org/10.1007/BF00687009

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

Neuropathology Group. Medical Research Council Cognitive Function and Aging Study. (2001). Pathological correlates of late-onset dementia in a multicentre, community-based population in England and Wales. Neuropathology Group of the Medical Research Council Cognitive Function and Ageing Study (MRC CFAS). *Lancet (London, England)*, *357*(9251), 169–175. https://doi.org/10.1016/s0140-6736(00)03589-3

Newson, R. S., & Kemps, E. B. (2005). General Lifestyle Activities as a Predictor of Current Cognition and Cognitive Change in Older Adults: A Cross-Sectional and Longitudinal Examination. *The Journals of Gerontology: Series B*, *60*(3), P113–P120. https://doi.org/10.1093/geronb/60.3.P113

Nga, T. T., Winichagoon, P., Dijkhuizen, M. A., Khan, N. C., Wasantwisut, E., & Wieringa, F. T. (2011). Decreased Parasite Load and Improved Cognitive Outcomes Caused by Deworming and Consumption of Multi-Micronutrient Fortified Biscuits in Rural Vietnamese Schoolchildren. *The American Journal of Tropical Medicine and Hygiene*, *85*(2), 333–340. https://doi.org/10.4269/ajtmh.2011.10-0651

Ngandu, T., Lehtisalo, J., Solomon, A., Levälahti, E., Ahtiluoto, S., Antikainen, R., Bäckman, L., Hänninen, T., Jula, A., Laatikainen, T., Lindström, J., Mangialasche, F., Paajanen, T., Pajala, S., Peltonen, M., Rauramaa, R., Stigsdotter-Neely, A., Strandberg, T., Tuomilehto, J., … Kivipelto, M. (2015). A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): A randomised controlled trial. *The Lancet*, *385*(9984), 2255–2263. https://doi.org/10.1016/S0140-6736(15)60461-5

Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., Kriegeskorte, N., Milham, M. P., Poldrack, R. A., Poline, J.-B., Proal, E., Thirion, B., Van Essen, D. C., White, T., & Yeo, B. T. T. (2017). Best practices in data analysis and sharing in neuroimaging using MRI. *Nature Neuroscience*, *20*(3), 299–303. https://doi.org/10.1038/nn.4500

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

Nithianantharajah, J., & Hannan, A. J. (2006). Enriched environments, experience-dependent plasticity and disorders of the nervous system. *Nature Reviews. Neuroscience*, *7*(9), 697–709. https://doi.org/10.1038/nrn1970

Nokia, M. S., Lensu, S., Ahtiainen, J. P., Johansson, P. P., Koch, L. G., Britton, S. L., & Kainulainen, H. (2016). Physical exercise increases adult hippocampal neurogenesis in male rats provided it is aerobic and sustained. *The Journal of Physiology*, *594*(7), 1855–1873. https://doi.org/10.1113/JP271552

Norbom, L. B., Ferschmann, L., Parker, N., Agartz, I., Andreassen, O. A., Paus, T., Westlye, L. T., & Tamnes, C. K. (2021). New insights into the dynamic development of the cerebral cortex in childhood and adolescence: Integrating macro- and microstructural MRI findings. *Progress in Neurobiology*, 102109. https://doi.org/10.1016/j.pneurobio.2021.102109

Northey, J. M., Cherbuin, N., Pumpa, K. L., Smee, D. J., & Rattray, B. (2018). Exercise interventions for cognitive function in adults older than 50: A systematic review with meta-analysis. *British Journal of Sports Medicine*, *52*(3), 154–160. https://doi.org/10.1136/bjsports-2016-096587

Norton, S., Matthews, F. E., Barnes, D. E., Yaffe, K., & Brayne, C. (2014). Potential for primary prevention of Alzheimer's disease: An analysis of population-based data. *The Lancet Neurology*, *13*(8), 788–794. https://doi.org/10.1016/S1474-4422(14)70136-X

Nowakowski, T. J., Pollen, A. A., Sandoval-Espinosa, C., & Kriegstein, A. R. (2016). Transformation of the Radial Glia Scaffold Demarcates Two Stages of Human Cerebral Cortex Development. *Neuron*, *91*(6), 1219–1227. https://doi.org/10.1016/j.neuron.2016.09.005

Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U., & Bäckman, L. (2012). Memory aging and brain maintenance. *Trends in Cognitive Sciences*, *16*(5), 292–305.

Nyberg, L., Magnussen, F., Lundquist, A., Baaré, W., Bartrés-Faz, D., Bertram, L., Boraxbekk, C. J., Brandmaier, A. M., Drevon, C. A., Ebmeier, K., Ghisletta, P., Henson, R. N., Junqué, C., Kievit, R., Kleemeyer, M., Knights, E., Kühn, S., Lindenberger, U., Penninx, B. W. J. H., … Fjell, A. M. (2021). Educational attainment does not influence brain aging. *Proceedings of the National Academy of Sciences*, *118*(18). https://doi.org/10.1073/pnas.2101644118

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, *3*(2), 173–182. https://doi.org/10.1038/s41562-018-0506-1

Oschwald, J., Guye, S., Liem, F., Rast, P., Willis, S., Röcke, C., Jäncke, L., Martin, M., & Mérillat, S. (2020). Brain structure and cognitive ability in healthy aging: A review on longitudinal correlated change. *Reviews in the Neurosciences*, *31*(1), 1–57. https://doi.org/10.1515/revneuro-2018-0096

Oxman, A. D., Lavis, J. N., & Fretheim, A. (2007). Use of evidence in WHO recommendations. *The Lancet*, *369*(9576), 1883–1889. https://doi.org/10.1016/S0140-6736(07)60675-8

Pandya, S. Y., Clem, M. A., Silva, L. M., & Woon, F. L. (2016). Does mild cognitive impairment always lead to dementia? A review. *Journal of the Neurological Sciences*, *369*, 57–62. https://doi.org/10.1016/j.jns.2016.07.055

Panizzon, M. S., Fennema-Notestine, C., Eyler, L. T., Jernigan, T. L., Prom-Wormley, E., Neale, M., Jacobson, K., Lyons, M. J., Grant, M. D., Franz, C. E., Xian, H., Tsuang, M., Fischl, B., Seidman, L., Dale, A., & Kremen, W. S. (2009). Distinct genetic influences on cortical surface area and cortical thickness. *Cerebral Cortex (New York, N.Y.: 1991)*, *19*(11), 2728–2735. https://doi.org/10.1093/cercor/bhp026

Pantazis, D., Joshi, A., Jiang, J., Shattuck, D. W., Bernstein, L. E., Damasio, H., & Leahy, R. M. (2010). Comparison of landmark-based and automatic methods for cortical surface registration. *NeuroImage*, *49*(3), 2479–2493. https://doi.org/10.1016/j.neuroimage.2009.09.027

Park, D. C., & Reuter-Lorenz, P. (2009). The Adaptive Brain: Aging and Neurocognitive Scaffolding. *Annual Review of Psychology*, *60*(1), 173–196. https://doi.org/10.1146/annurev.psych.59.103006.093656

Pereira, A. C., Huddleston, D. E., Brickman, A. M., Sosunov, A. A., Hen, R., McKhann, G. M., Sloan, R., Gage, F. H., Brown, T. R., & Small, S. A. (2007). An in vivo correlate of exercise-induced neurogenesis in the adult dentate gyrus. *Proceedings of the National Academy of Sciences*, *104*(13), 5638–5643. https://doi.org/10.1073/pnas.0611721104

Peters, A. (2007). The Effects of Normal Aging on Nerve Fibers and Neuroglia in the Central Nervous System. In D. R. Riddle (Ed.), *Brain Aging: Models, Methods, and Mechanisms*. CRC Press/Taylor & Francis. http://www.ncbi.nlm.nih.gov/books/NBK3873/

Petticrew, M., Wilson, P., Wright, K., & Song, F. (2002). Quality of Cochrane reviews. *BMJ : British Medical Journal*, *324*(7336), 545.

Pitkala, K. H., Routasalo, P., Kautiainen, H., Sintonen, H., & Tilvis, R. S. (2011). Effects of socially stimulating group intervention on lonely, older people's cognition: A randomized, controlled trial. *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry*, *19*(7), 654–663. https://doi.org/10.1097/JGP.0b013e3181f7d8b0

Plasqui, G., & Westerterp, K. R. (2005). Accelerometry and Heart Rate as a Measure of Physical Fitness: Proof of Concept: *Medicine & Science in Sports & Exercise*, *37*(5), 872–876. https://doi.org/10.1249/01.MSS.0000161805.61893.C0

Podewils, L. J., Guallar, E., Beauchamp, N., Lyketsos, C. G., Kuller, L. H., & Scheltens, P. (2007). Physical activity and white matter lesion progression: Assessment using MRI. *Neurology*, *68*(15), 1223–1226. https://doi.org/10.1212/01.wnl.0000259063.50219.3e

Poldrack, R. A. (2012). The future of fMRI in cognitive neuroscience. *NeuroImage*, *62*(2), 1216–1220. https://doi.org/10.1016/j.neuroimage.2011.08.007

Poldrack, R. A., & Gorgolewski, K. J. (2014). Making big data open: Data sharing in neuroimaging. *Nature Neuroscience*, *17*(11), 1510–1517. https://doi.org/10.1038/nn.3818

Praag, H. van, Shubert, T., Zhao, C., & Gage, F. H. (2005). Exercise Enhances Learning and Hippocampal Neurogenesis in Aged Mice. *Journal of Neuroscience*, *25*(38), 8680–8685. https://doi.org/10.1523/JNEUROSCI.1731-05.2005

Prince, S. A., Cardilli, L., Reed, J. L., Saunders, T. J., Kite, C., Douillette, K., Fournier, K., & Buckley, J. P. (2020). A comparison of self-reported and device measured sedentary behaviour in adults: A systematic review and meta-analysis. *International Journal of Behavioral Nutrition and Physical Activity*, *17*(1), 31. https://doi.org/10.1186/s12966-020-00938-3

Pruchno, R. A., & Wilson-Genderson, M. (2015). A Longitudinal Examination of the Effects of Early Influences and Midlife Characteristics on Successful Aging. *The Journals of Gerontology: Series B*, *70*(6), 850–859. https://doi.org/10.1093/geronb/gbu046

Radder, H. (2009). The philosophy of scientific experimentation: A review. *Automated Experimentation*, *1*, 2. https://doi.org/10.1186/1759-4499-1-2

Raffin, J., Rolland, Y., Fischer, C., Mangin, J.-F., Gabelle, A., Vellas, B., & de Souto Barreto, P. (2021). Cross-sectional associations between cortical thickness and physical activity in older adults with spontaneous memory complaints: The MAPT Study. *Journal of Sport and Health Science*. https://doi.org/10.1016/j.jshs.2021.01.011

Rakic, P. (2000). Radial unit hypothesis of neocortical expansion. *Novartis Foundation Symposium*, *228*, 30–42; discussion 42-52. https://doi.org/10.1002/0470846631.ch3

Raykov, T., & Penev, S. (1999). On Structural Equation Model Equivalence. *Multivariate Behavioral Research*, *34*(2), 199–244. https://doi.org/10.1207/S15327906Mb340204

Raz, N. (2005). The Aging Brain Observed in Vivo: Differential Changes and Their Modifiers. In *Cognitive neuroscience of aging: Linking cognitive and cerebral aging* (pp. 19–57). Oxford University Press.

Raz, N., & Lindenberger, U. (2011). Only Time will Tell: Cross-sectional Studies Offer no Solution to the Age-Brain-Cognition Triangle—Comment on. *Psychological Bulletin*, *137*(5), 790–795. https://doi.org/10.1037/a0024503

Raz, N., Rodrigue, K. M., Kennedy, K. M., Head, D., Gunning-Dixon, F., & Acker, J. D. (2003). Differential Aging of the Human Striatum: Longitudinal Evidence. *American Journal of Neuroradiology*, *24*(9), 1849–1856.

Reijnders, J., van Heugten, C., & van Boxtel, M. (2013). Cognitive interventions in healthy older adults and people with mild cognitive impairment: A systematic review. *Ageing Research Reviews*, *12*(1), 263–275. https://doi.org/10.1016/j.arr.2012.07.003

Reiter, K., Nielson, K. A., Smith, T. J., Weiss, L. R., Alfini, A. J., & Smith, J. C. (2015). Improved Cardiorespiratory Fitness Is Associated with Increased Cortical Thickness in Mild Cognitive Impairment. *Journal of the International Neuropsychological Society*, *21*(10), 757–767. https://doi.org/10.1017/S135561771500079X

Reppermund, S., Brodaty, H., Crawford, J. D., Kochan, N. A., Slavin, M. J., Trollor, J. N., Draper, B., & Sachdev, P. S. (2011). The relationship of current depressive symptoms and past depression with cognitive impairment and instrumental activities of daily living in an elderly population: The Sydney Memory and Ageing Study. *Journal of Psychiatric Research*, *45*(12), 1600–1607. https://doi.org/10.1016/j.jpsychires.2011.08.001

Resnick, S. M., Goldszal, A. F., Davatzikos, C., Golski, S., Kraut, M. A., Metter, E. J., Bryan, R. N., & Zonderman, A. B. (2000). One-year Age Changes in MRI Brain Volumes in Older Adults. *Cerebral Cortex*, *10*(5), 464–472. https://doi.org/10.1093/cercor/10.5.464

Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage*, *61*(4), 1402–1418. https://doi.org/10.1016/j.neuroimage.2012.02.084

Reuter-Lorenz, P. A., & Park, D. C. (2014). How Does it STAC Up? Revisiting the Scaffolding Theory of Aging and Cognition. *Neuropsychology Review*, *24*(3), 355–370. https://doi.org/10.1007/s11065-014-9270-9

Revelle, W. (2018). *Procedures for psychological, psychometric, and personality research, version 1.8.3*. Northwestern University, Evanston, IL. https://CRAN.R-project.org/package=psych

Ritchie, S. J., Bastin, M. E., Tucker-Drob, E. M., Maniega, S. M., Engelhardt, L. E., Cox, S. R., Royle, N. A., Gow, A. J., Corley, J., Pattie, A., Taylor, A. M., Hernández, M. del C. V., Starr, J. M., Wardlaw, J. M., & Deary, I. J. (2015). Coupled Changes in Brain White Matter Microstructure and Fluid Intelligence in Later Life. *Journal of Neuroscience*, *35*(22), 8672–8682. https://doi.org/10.1523/JNEUROSCI.0862-15.2015

Robertson, I. H. (2013). A noradrenergic theory of cognitive reserve: Implications for Alzheimer's disease. *Neurobiology of Aging*, *34*(1), 298–308. https://doi.org/10.1016/j.neurobiolaging.2012.05.019

Robertson, I. H. (2014). A right hemisphere role in cognitive reserve. *Neurobiology of Aging*, *35*(6), 1375–1385. https://doi.org/10.1016/j.neurobiolaging.2013.11.028

Rodd, J. M., Longe, O. A., Randall, B., & Tyler, L. K. (2010). The functional organisation of the fronto-temporal language system: Evidence from syntactic and semantic ambiguity. *Neuropsychologia*, *48*(5), 1324–1335. https://doi.org/10.1016/j.neuropsychologia.2009.12.035

Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *The American Psychologist*, *65*(1), 1–12. https://doi.org/10.1037/a0018326

Rogalski, E. J., Gefen, T., Shi, J., Samimi, M., Bigio, E., Weintraub, S., Geula, C., & Mesulam, M.-M. (2013). Youthful Memory Capacity in Old Brains: Anatomic and Genetic Clues from the Northwestern SuperAging Project. *Journal of Cognitive Neuroscience*, *25*(1), 29–36. https://doi.org/10.1162/jocn_a_00300

Rosseel, Y. (2014). The lavaan tutorial. *Department of Data Analysis: Ghent University*. http://dornsifecms.usc.edu/assets/sites/210/docs/GC3/lavaan_tutorial.pdf

Rossi, C., Angelucci, A., Costantin, L., Braschi, C., Mazzantini, M., Babbini, F., Fabbri, M. E., Tessarollo, L., Maffei, L., Berardi, N., & Caleo, M. (2006). Brain-derived neurotrophic factor (BDNF) is required for the enhancement of hippocampal neurogenesis following environmental enrichment. *European Journal of Neuroscience*, *24*(7), 1850–1856. https://doi.org/10.1111/j.1460-9568.2006.05059.x

Rostral locus coeruleus integrity is associated with better memory performance in older adults | Nature Human Behaviour. (n.d.). Retrieved January 19, 2022, from https://www.nature.com/articles/s41562-019-0715-2

Routledge, F. S., Campbell, T. S., McFetridge-Durdle, J. A., & Bacon, S. L. (2010). Improvements in heart rate variability with exercise therapy. *The Canadian Journal of Cardiology*, *26*(6), 303–312.

Rovio, S., Spulber, G., Nieminen, L. J., Niskanen, E., Winblad, B., Tuomilehto, J., Nissinen, A., Soininen, H., & Kivipelto, M. (2010). The effect of midlife physical activity on structural brain changes in the elderly. *Neurobiology of Aging*, *31*(11), 1927–1936. https://doi.org/10.1016/j.neurobiolaging.2008.10.007

Ruscheweyh, R., Willemer, C., Krüger, K., Duning, T., Warnecke, T., Sommer, J., Völker, K., Ho, H. V., Mooren, F., Knecht, S., & Flöel, A. (2011). Physical activity and memory functions: An interventional study. *Neurobiology of Aging*, *32*(7), 1304–1319. https://doi.org/10.1016/j.neurobiolaging.2009.08.001

Russo, F., & Williamson, J. (2007). Interpreting Causality in the Health Sciences. *International Studies in the Philosophy of Science*, *21*(2), 157–170. https://doi.org/10.1080/02698590701498084

Sabarwal, S., & Abu-Jawdeh, M. (2018). *What Teachers Believe: Mental Models about Accountability, Absenteeism, and Student Learning*. World Bank, Washington, DC. https://doi.org/10.1596/1813-9450-8454

Sabia, S., Dugravot, A., Dartigues, J.-F., Abell, J., Elbaz, A., Kivimäki, M., & Singh-Manoux, A. (2017). Physical activity, cognitive decline, and risk of dementia: 28 year follow-up of Whitehall II cohort study. *BMJ*, *357*, j2709. https://doi.org/10.1136/bmj.j2709

Sachdev, P. S., Blacker, D., Blazer, D. G., Ganguli, M., Jeste, D. V., Paulsen, J. S., & Petersen, R. C. (2014). Classifying neurocognitive disorders: The DSM-5 approach. *Nature Reviews Neurology*, *10*(11), 634–642. https://doi.org/10.1038/nrneurol.2014.181

Sachdeva, A., Chandra, M., Choudhary, M., Dayal, P., & Anand, K. S. (2016). Alcohol-Related Dementia and Neurocognitive Impairment: A Review Study. *International Journal of High Risk Behaviors & Addiction*, *5*(3), e27976. https://doi.org/10.5812/ijhrba.27976

Saif, N., Yan, P., Niotis, K., Scheyer, O., Rahman, A., Berkowitz, M., Krikorian, R., Hristov, H., Sadek, G., Bellara, S., & Isaacson, R. S. (2020). Feasibility of using a wearable biosensor device in patients at risk for Alzheimer's disease dementia. *The Journal of Prevention of Alzheimer's Disease*, *7*(2), 104–111. https://doi.org/10.14283/jpad.2019.39

Salat, D. H. (2011). The Declining Infrastructure of the Aging Brain. *Brain Connectivity*, *1*(4), 279–293. https://doi.org/10.1089/brain.2011.0056

Salat, D. H., Buckner, R. L., Snyder, A. Z., Greve, D. N., Desikan, R. S. R., Busa, E., Morris, J. C., Dale, A. M., & Fischl, B. (2004). Thinning of the Cerebral Cortex in Aging. *Cerebral Cortex*, *14*(7), 721–730. https://doi.org/10.1093/cercor/bhh032

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology*, *54*(1–3), 35–54. https://doi.org/10.1016/S0301-0511(00)00052-1

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

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews. Neuroscience*, *10*(3), 211–223. https://doi.org/10.1038/nrn2573

Sara, S. J., & Bouret, S. (2012). Orienting and Reorienting: The Locus Coeruleus Mediates Cognition through Arousal. *Neuron*, *76*(1), 130–141. https://doi.org/10.1016/j.neuron.2012.09.011

Scahill, R. I., Frost, C., Jenkins, R., Whitwell, J. L., Rossor, M. N., & Fox, N. C. (2003). A Longitudinal Study of Brain Volume Changes in Normal Aging Using Serial Registered Magnetic Resonance Imaging. *Archives of Neurology*, *60*(7), 989–994. https://doi.org/10.1001/archneur.60.7.989

Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist*, *49*(4), 304–313. https://doi.org/10.1037/0003-066X.49.4.304

Schaie, K. W., Willis, S. L., & Pennak, S. (2005). An Historical Framework for Cohort Differences in Intelligence. *Research in Human Development*, *2*(1–2), 43–67.

Scheltens, P., Barkhof, F., Leys, D., Wolters, E. C., Ravid, R., & Kamphorst, W. (1995). Histopathologic correlates of white matter changes on MRI in Alzheimer's disease and normal aging. *Neurology*, *45*(5), 883–888. https://doi.org/10.1212/WNL.45.5.883

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research*, *8*(2), 23–74.

Schlösser, R., Gesierich, T., Kaufmann, B., Vucurevic, G., Hunsche, S., Gawehn, J., & Stoeter, P. (2003). Altered effective connectivity during working memory performance in schizophrenia: A study with fMRI and structural equation modeling. *NeuroImage*, *19*(3), 751–763. https://doi.org/10.1016/S1053-8119(03)00106-X

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred Days of Cognitive Training Enhance Broad Cognitive Abilities in Adulthood: Findings from the COGITO Study. *Frontiers in Aging Neuroscience*, *2*. https://doi.org/10.3389/fnagi.2010.00027

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In *Contemporary intellectual assessment: Theories, tests, and issues, 3rd ed* (pp. 99–144). The Guilford Press.

Schram, M. T., Kostense, P. J., van Dijk, R. A. J. M., Dekker, J. M., Nijpels, G., Bouter, L. M., Heine, R. J., & Stehouwer, C. D. A. (2002). Diabetes, pulse pressure and cardiovascular mortality: The Hoorn Study. *Journal of Hypertension*, *20*(9), 1743–1751.

Schuff, N., Tosun, D., Insel, P. S., Chiang, G. C., Truran, D., Aisen, P. S., Jack, C. R., & Weiner, M. W. (2012). Nonlinear time course of brain volume loss in cognitively normal and impaired elders. *Neurobiology of Aging*, *33*(5), 845–855. https://doi.org/10.1016/j.neurobiolaging.2010.07.012

Scoville, W. B., & Milner, B. (1957). LOSS OF RECENT MEMORY AFTER BILATERAL HIPPOCAMPAL LESIONS. *Journal of Neurology, Neurosurgery, and Psychiatry*, *20*(1), 11–21.

Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, *22*(3), 278–295. https://doi.org/10.1177/0962280210395740

Sexton, C. E., Betts, J. F., Demnitz, N., Dawes, H., Ebmeier, K. P., & Johansen-Berg, H. (2016). A systematic review of MRI studies examining the relationship between physical fitness and activity and the white matter of the ageing brain. *NeuroImage*, *131*, 81–90. https://doi.org/10.1016/j.neuroimage.2015.09.071

Shafto, M. A., Tyler, L. K., Dixon, M., Taylor, J. R., Rowe, J. B., Cusack, R., Calder, A. J., Marslen-Wilson, W. D., Duncan, J., Dalgleish, T., Henson, R. N., Brayne, C., & Matthews, F. E. (2014). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: A cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC Neurology*, *14*, 204. https://doi.org/10.1186/s12883-014-0204-1

Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain: A Journal of Neurology*, *114 (Pt 2)*, 727–741.

Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., Porter, A. C., Tugwell, P., Moher, D., & Bouter, L. M. (2007). Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology*, *7*(1), 10. https://doi.org/10.1186/1471-2288-7-10

Shimony, J. S., Smyser, C. D., Wideman, G., Alexopoulos, D., Hill, J., Harwell, J., Dierker, D., Van Essen, D. C., Inder, T. E., & Neil, J. J. (2016). Comparison of cortical folding measures for evaluation of developing human brain. *NeuroImage*, *125*, 780–790. https://doi.org/10.1016/j.neuroimage.2015.11.001

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*(4), 422–445. https://doi.org/10.1037/1082-989X.7.4.422

Sigurdsson, S., Aspelund, T., Forsberg, L., Fredriksson, J., Kjartansson, O., Oskarsdottir, B., Jonsson, P. V., Eiriksdottir, G., Harris, T. B., Zijdenbos, A., van Buchem, M. A., Launer, L. J., & Gudnason, V. (2012). Brain tissue volumes in the general population of the elderly: The AGES-Reykjavik Study. *NeuroImage*, *59*(4), 3862–3870. https://doi.org/10.1016/j.neuroimage.2011.11.024

Singh, F., Shu, I.-W., Hsu, S.-H., Link, P., Pineda, J. A., & Granholm, E. (2020). Modulation of frontal gamma oscillations improves working memory in schizophrenia. *NeuroImage: Clinical*, *27*, 102339. https://doi.org/10.1016/j.nicl.2020.102339

Small, B. J., Dixon, R. A., McArdle, J. J., & Grimm, K. J. (2012a). Do changes in lifestyle engagement moderate cognitive decline in normal aging? Evidence from the Victoria Longitudinal Study. *Neuropsychology*, *26*(2), 144–155. https://doi.org/10.1037/a0026579

Small, B. J., Dixon, R. A., McArdle, J. J., & Grimm, K. J. (2012b). Do changes in lifestyle engagement moderate cognitive decline in normal aging? Evidence from the Victoria Longitudinal Study. *Neuropsychology*, *26*(2), 144–155. https://doi.org/10.1037/a0026579

Smith, P. J., Blumenthal, J. A., Hoffman, B. M., Cooper, H., Strauman, T. A., Welsh-Bohmer, K., Browndyke, J. N., & Sherwood, A. (2010). Aerobic Exercise and Neurocognitive Performance: A Meta-Analytic Review of Randomized Controlled Trials. *Psychosomatic Medicine*, *72*(3), 239–252. https://doi.org/10.1097/PSY.0b013e3181d14633

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23 Suppl 1*, S208-219. https://doi.org/10.1016/j.neuroimage.2004.07.051

Snitz, B. E., Small, B. J., Wang, T., Chang, C.-C. H., Hughes, T. F., & Ganguli, M. (2015). Do Subjective Memory Complaints Lead or Follow Objective Cognitive Change? A Five-Year Population Study of Temporal Influence. *Journal of the International Neuropsychological Society*, *21*(09), 732–742. https://doi.org/10.1017/S1355617715000922

Song, D., Yu, D. S. F., Li, P. W. C., & Lei, Y. (2018). The effectiveness of physical exercise on cognitive and psychological outcomes in individuals with mild cognitive impairment: A systematic review and meta-analysis. *International Journal of Nursing Studies*, *79*, 155–164. https://doi.org/10.1016/j.ijnurstu.2018.01.002

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*(2), 245–251. https://doi.org/10.1037/0033-2909.87.2.245

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*, *8*(03), 448–460. https://doi.org/10.1017/S1355617702813248

Stern, Y. (2009). Cognitive reserve☆. *Neuropsychologia*, *47*(10), 2015–2028. https://doi.org/10.1016/j.neuropsychologia.2009.03.004

Storsve, A. B., Fjell, A. M., Tamnes, C. K., Westlye, L. T., Overbye, K., Aasland, H. W., & Walhovd, K. B. (2014). Differential longitudinal changes in cortical thickness, surface area and volume across the adult life span: Regions of accelerating and decelerating change. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(25), 8488–8498. https://doi.org/10.1523/JNEUROSCI.0391-14.2014

Streiner, D. L. (2005). Finding Our Way: An Introduction to Path Analysis. *The Canadian Journal of Psychiatry*, *50*(2), 115–122. https://doi.org/10.1177/070674370505000207

Strömmer, J. M., Davis, S. W., Henson, R. N., Tyler, L. K., Cam-CAN, & Campbell, K. L. (2020). Physical Activity Predicts Population-Level Age-Related Differences in Frontal White Matter. *The Journals of Gerontology: Series A*, *75*(2), 236–243. https://doi.org/10.1093/gerona/gly220

Strout, K. A., David, D. J., Dyer, E. J., Gray, R. C., Robnett, R. H., & Howard, E. P. (2016). Behavioral Interventions in Six Dimensions of Wellness That Protect the Cognitive Health of Community-Dwelling Older Adults: A Systematic Review. *Journal of the American Geriatrics Society*, *64*(5), 944–958. https://doi.org/10.1111/jgs.14129

Sullivan, E. V., Marsh, L., Mathalon, D. H., Lim, K. O., & Pfefferbaum, A. (1995). Age-related decline in MRI volumes of temporal lobe gray matter but not hippocampus. *Neurobiology of Aging*, *16*(4), 591–606. https://doi.org/10.1016/0197-4580(95)00074-O

Sun, F. W., Stepanovic, M. R., Andreano, J., Barrett, L. F., Touroutoglou, A., & Dickerson, B. C. (2016). Youthful brains in older adults: Preserved neuroanatomy in the default mode and salience networks contributes to youthful memory in superaging. *Journal of Neuroscience*, *36*(37), 9659–9668.

Swain, R. A., Harris, A. B., Wiener, E. C., Dutka, M. V., Morris, H. D., Theien, B. E., Konda, S., Engberg, K., Lauterbur, P. C., & Greenough, W. T. (2003). Prolonged exercise induces angiogenesis and increases cerebral blood volume in primary motor cortex of the rat. *Neuroscience*, *117*(4), 1037–1046. https://doi.org/10.1016/S0306-4522(02)00664-4

Swan, G. E., DeCarli, C., Miller, B. L., Reed, T., Wolf, P. A., Jack, L. M., & Carmelli, D. (1998). Association of midlife blood pressure to late-life cognitive decline and brain morphology. *Neurology*, *51*(4), 986–993. https://doi.org/10.1212/WNL.51.4.986

Szabo, A. N., McAuley, E., Erickson, K. I., Voss, M., Prakash, R. S., Mailey, E. L., Wójcicki, T. R., White, S. M., Gothe, N., Olson, E. A., & Kramer, A. F. (2011). Cardiorespiratory fitness, hippocampal volume, and frequency of forgetting in older adults. *Neuropsychology*, *25*(5), 545–553. https://doi.org/10.1037/a0022733

Szuhany, K. L., Bugatti, M., & Otto, M. W. (2015). A meta-analytic review of the effects of exercise on brain-derived neurotrophic factor. *Journal of Psychiatric Research*, *60*, 56–64. https://doi.org/10.1016/j.jpsychires.2014.10.003

Taylor, J. R., Williams, N., Cusack, R., Auer, T., Shafto, M. A., Dixon, M., Tyler, L. K., Cam-CAN, & Henson, R. N. (2017). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage*, *144*, 262–269. https://doi.org/10.1016/j.neuroimage.2015.09.018

Taylor-Robinson, D. C., Maayan, N., Soares-Weiser, K., Donegan, S., & Garner, P. (2012). Deworming drugs for soil-transmitted intestinal worms in children: Effects on nutritional indicators, haemoglobin and school performance. *Cochrane Database of Systematic Reviews*, *11*. https://doi.org/10.1002/14651858.CD000371.pub5

Taylor-Robinson, D. C., Maayan, N., Soares-Weiser, K., Donegan, S., & Garner, P. (2015). Deworming drugs for soil-transmitted intestinal worms in children: Effects on nutritional indicators, haemoglobin, and school performance. *The Cochrane Database of Systematic Reviews*, *7*, CD000371. https://doi.org/10.1002/14651858.CD000371.pub6

Teixeira, C. V. L., Gobbi, L. T. B., Corazza, D. I., Stella, F., Costa, J. L. R., & Gobbi, S. (2012). Non-pharmacological interventions on cognitive functions in older people with mild cognitive impairment (MCI). *Archives of Gerontology and Geriatrics*, *54*(1), 175–180. https://doi.org/10.1016/j.archger.2011.02.014

Thorvaldsson, V., MacDonald, S. W. S., Fratiglioni, L., Winblad, B., Kivipelto, M., Laukka, E. J., Skoog, I., Sacuiu, S., Guo, X., Östling, S., Börjesson-Hanson, A., Gustafson, D., Johansson, B., & Bäckman, L. (2011). Onset and Rate of Cognitive Change Before Dementia Diagnosis: Findings From Two Swedish Population-Based Longitudinal Studies. *Journal of the International Neuropsychological Society*, *17*(1), 154–162. https://doi.org/10.1017/S1355617710001372

Tóth-Király, I., Bõthe, B., Rigó, A., & Orosz, G. (2017). An Illustration of the Exploratory Structural Equation Modeling (ESEM) Framework on the Passion Scale. *Frontiers in Psychology*, *8*, 1968. https://doi.org/10.3389/fpsyg.2017.01968

Troadec, J.-D., Marien, M., Darios, F., Hartmann, A., Ruberg, M., Colpaert, F., & Michel, P. P. (2001). Noradrenaline provides long-term protection to dopaminergic neurons by reducing oxidative stress. *Journal of Neurochemistry*, *79*(1), 200–210. https://doi.org/10.1046/j.1471-4159.2001.00556.x

Tsang, R., Colley, L., & Lynd, L. D. (2009). Inadequate statistical power to detect clinically significant differences in adverse event rates in randomized controlled trials. *Journal of Clinical Epidemiology*, *62*(6), 609–616. https://doi.org/10.1016/j.jclinepi.2008.08.005

Tu, Y.-K., Gunnell, D., & Gilthorpe, M. S. (2008). Simpson's Paradox, Lord's Paradox, and Suppression Effects are the same phenomenon – the reversal paradox. *Emerging Themes in Epidemiology*, *5*(1), 2. https://doi.org/10.1186/1742-7622-5-2

Tucker-Drob, E. M. (2011). Global and Domain-Specific Changes in Cognition throughout Adulthood. *Developmental Psychology*, *47*(2), 331–343. https://doi.org/10.1037/a0021361

Tucker-Drob, E. M., Brandmaier, A. M., & Lindenberger, U. (2019). Coupled Cognitive Changes in Adulthood: A Meta-analysis. *Psychological Bulletin*, *145*(3), 273–301. https://doi.org/10.1037/bul0000179

Ullman, J. B., & Bentler, P. M. (2012). Structural Equation Modeling. In *Handbook of Psychology, Second Edition*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118133880.hop202023

Usami, S., Hayes, T., & McArdle, J. J. (2016). Inferring Longitudinal Relationships Between Variables: Model Selection Between the Latent Change Score and Autoregressive Cross-Lagged Factor Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(3), 331–342. https://doi.org/10.1080/10705511.2015.1066680

Valenzuela, M. J., & Sachdev, P. (2006). Brain reserve and dementia: A systematic review. *Psychological Medicine*, *36*(4), 441–454. https://doi.org/10.1017/S0033291705006264

Valenzuela, M. J., & Sachdev, P. (2007). Assessment of complex mental activity across the lifespan: Development of the Lifetime of Experiences Questionnaire (LEQ). *Psychological Medicine*, *37*(07), 1015. https://doi.org/10.1017/S003329170600938X

van der Meer, D., Frei, O., Kaufmann, T., Chen, C.-H., Thompson, W. K., O'Connell, K. S., Monereo Sánchez, J., Linden, D. E. J., Westlye, L. T., Dale, A. M., & Andreassen, O. A. (2020). Quantifying the Polygenic Architecture of the Human Cerebral Cortex: Extensive Genetic Overlap between Cortical Thickness and Surface Area. *Cerebral Cortex*, *30*(10), 5597–5603. https://doi.org/10.1093/cercor/bhaa146

van Harmelen, A.-L., Kievit, R. A., Ioannidis, K., Neufeld, S., Jones, P. B., Bullmore, E., Dolan, R., NSPN Consortium, Fonagy, P., & Goodyer, I. (2017). Adolescent friendships predict later resilient functioning across psychosocial domains in a healthy community cohort. *Psychological Medicine*, *47*(13), 2312–2322. https://doi.org/10.1017/S0033291717000836

van Praag, H., Kempermann, G., & Gage, F. H. (2000). Neural consequences of environmental enrichment. *Nature Reviews. Neuroscience*, *1*(3), 191–198. https://doi.org/10.1038/35044558

Vandenbroucke, J. P. (2015). Preregistration: When shall we start the real discussion? *European Journal of Public Health*, *25*(4), 555–556. https://doi.org/10.1093/eurpub/ckv118

Verstynen, T. D., Lynch, B., Miller, D. L., Voss, M. W., Prakash, R. S., Chaddock, L., Basak, C., Szabo, A., Olson, E. A., Wojcicki, T. R., Fanning, J., Gothe, N. P., McAuley, E., Kramer, A. F., & Erickson, K. I. (2012). Caudate Nucleus Volume Mediates the Link between Cardiorespiratory Fitness and Cognitive Flexibility in Older Adults. *Journal of Aging Research*, *2012*, e939285. https://doi.org/10.1155/2012/939285

Voelcker-Rehage, C., Godde, B., & Staudinger, U. M. (2011). Cardiovascular and Coordination Training Differentially Improve Cognitive Performance and Neural Processing in Older Adults. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00026

Voelcker-Rehage, C., & Niemann, C. (2013). Structural and functional brain changes related to different types of physical activity across the life span. *Neuroscience & Biobehavioral Reviews*, *37*(9, Part B), 2268–2295. https://doi.org/10.1016/j.neubiorev.2013.01.028

Voss, M. W., Vivar, C., Kramer, A. F., & van Praag, H. (2013). Bridging animal and human models of exercise-induced brain plasticity. *Trends in Cognitive Sciences*, *17*(10), 525–544. https://doi.org/10.1016/j.tics.2013.08.001

Waldman, D. A. (2013). Interdisciplinary research is the key. *Frontiers in Human Neuroscience*, *7*, 562. https://doi.org/10.3389/fnhum.2013.00562

Walhovd, K. B., Fjell, A. M., Giedd, J., Dale, A. M., & Brown, T. T. (2017). Through Thick and Thin: A Need to Reconcile Contradictory Results on Trajectories in Human Cortical Development. *Cerebral Cortex (New York, N.Y.: 1991)*, *27*(2), 1472–1481. https://doi.org/10.1093/cercor/bhv301

Walhovd, K. B., Fjell, A. M., Reinvang, I., Lundervold, A., Dale, A. M., Eilertsen, D. E., Quinn, B. T., Salat, D., Makris, N., & Fischl, B. (2005). Effects of age on volumes of cortex, white matter and subcortical structures. *Neurobiology of Aging*, *26*(9), 1261–1270. https://doi.org/10.1016/j.neurobiolaging.2005.05.020

Walhovd, K. B., Fjell, A. M., Westerhausen, R., Nyberg, L., Ebmeier, K. P., Lindenberger, U., Bartrés-Faz, D., Baaré, W. F. C., Siebner, H. R., Henson, R., Drevon, C. A., Strømstad Knudsen, G. P., Ljøsne, I. B., Penninx, B. W. J. H., Ghisletta, P., Rogeberg, O., Tyler, L., Bertram, L., & Lifebrain Consortium. (2018). Healthy minds 0-100 years: Optimising the use of European brain imaging cohorts ("Lifebrain"). *European Psychiatry: The Journal of the Association of European Psychiatrists*, *50*, 47–56. https://doi.org/10.1016/j.eurpsy.2017.12.006

Walhovd, K. B., Krogsrud, S. K., Amlien, I. K., Bartsch, H., Bjørnerud, A., Due-Tønnessen, P., Grydeland, H., Hagler, D. J., Håberg, A. K., Kremen, W. S., Ferschmann, L., Nyberg, L., Panizzon, M. S., Rohani, D. A., Skranes, J., Storsve, A. B., Sølsnes, A. E., Tamnes, C. K., Thompson, W. K., … Fjell, A. M. (2016). Neurodevelopmental origins of lifespan changes in brain and cognition. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(33), 9357–9362. https://doi.org/10.1073/pnas.1524259113

Walhovd, K. B., Storsve, A. B., Westlye, L. T., Drevon, C. A., & Fjell, A. M. (2014). Blood markers of fatty acids and vitamin D, cardiovascular measures, body mass index, and physical activity relate to longitudinal cortical thinning in normal aging. *Neurobiology of Aging*, *35*(5), 1055–1064. https://doi.org/10.1016/j.neurobiolaging.2013.11.011

Wang, J.-J., & Kaufman, A. S. (1993). Changes in Fluid and Crystallized Intelligence Across the 20- to 90-Year Age Range on the K-Bit. *Journal of Psychoeducational Assessment*, *11*(1), 29–37. https://doi.org/10.1177/073428299301100104

Wang, Y., Hao, L., Zhang, Y., Zuo, C., & Wang, D. (2019). Entorhinal cortex volume, thickness, surface area and curvature trajectories over the adult lifespan. *Psychiatry Research: Neuroimaging*, *292*, 47–53. https://doi.org/10.1016/j.pscychresns.2019.09.002

Wechsler, C. J. (1999). *Wechsler Memory Scale, Third UK Edition*. London: Harcourt Assessment.

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

Wenger, E., Brozzoli, C., Lindenberger, U., & Lövdén, M. (2017). Expansion and Renormalization of Human Brain Structure During Skill Acquisition. *Trends in Cognitive Sciences*, *21*(12), 930–939. https://doi.org/10.1016/j.tics.2017.09.008

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial Invariance within Longitudinal Structural Equation Models: Measuring the Same Construct across Time. *Child Development Perspectives*, *4*(1), 10–18. https://doi.org/10.1111/j.1750-8606.2009.00110.x

Wierenga, L. M., Langen, M., Oranje, B., & Durston, S. (2014). Unique developmental trajectories of cortical thickness and surface area. *NeuroImage*, *87*, 120–126. https://doi.org/10.1016/j.neuroimage.2013.11.010

Williams, V. J., Hayes, J. P., Forman, D. E., Salat, D. H., Sperling, R. A., Verfaellie, M., & Hayes, S. M. (2017). Cardiorespiratory fitness is differentially associated with cortical thickness in young and older adults. *NeuroImage*, *146*, 1084–1092. https://doi.org/10.1016/j.neuroimage.2016.10.033

Winkler, A. M., Greve, D. N., Bjuland, K. J., Nichols, T. E., Sabuncu, M. R., Håberg, A. K., Skranes, J., & Rimol, L. M. (2018). Joint Analysis of Cortical Area and Thickness as a Replacement for the Analysis of the Volume of the Cerebral Cortex. *Cerebral Cortex (New York, NY)*, *28*(2), 738–749. https://doi.org/10.1093/cercor/bhx308

Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., Duggirala, R., & Glahn, D. C. (2010). Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *NeuroImage*, *53*(3), 1135–1146. https://doi.org/10.1016/j.neuroimage.2009.12.028

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and Psychological Measurement*, *76*(6), 913–934. https://doi.org/10.1177/0013164413495237

Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics*, *141*(2), 1281–1301. https://doi.org/10.1016/j.jeconom.2007.02.002

World Health Organization (Ed.). (2015). *World report on ageing and health*. World Health Organization.

Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 219–240, 269–281). Lawrence Erlbaum Associates Publishers.

Wu, C.-Y., Hu, H.-Y., Chou, Y.-J., Huang, N., Chou, Y.-C., & Li, C.-P. (2015). High Blood Pressure and All-Cause and Cardiovascular Disease Mortalities in Community-Dwelling Older Adults. *Medicine*, *94*(47), e2160. https://doi.org/10.1097/MD.0000000000002160

Yankner, B. A., Lu, T., & Loerch, P. (2008). The Aging Brain. *Annual Review of Pathology: Mechanisms of Disease*, *3*(1), 41–66. https://doi.org/10.1146/annurev.pathmechdis.2.010506.092044

Yoo, W., Mayberry, R., Bae, S., Singh, K., (Peter) He, Q., & Lillard, J. W. (2014). A Study of Effects of MultiCollinearity in the Multivariable Analysis. *International Journal of Applied Science and Technology*, *4*(5), 9–19.

Zaslavsky, O., Cochrane, B. B., Woods, N. F., LaCroix, A. Z., Liu, J., Herting, J. R., Goveas, J. S., Johnson, K. C., Kuller, L. H., Martin, L. W., Michael, Y. L., Robinson, J. G., Stefanick, M., & Tinker, L. F. (2014). Trajectories of positive aging: Observations from the women's health initiative study. *International Psychogeriatrics*, *26*(08), 1351–1362. https://doi.org/10.1017/S1041610214000593

Zeileis, A. (2004). Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software*, *11*(1), 1–17. https://doi.org/10.18637/jss.v011.i10

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860

Zhang, Z., Hamagami, F., Grimm, K. J., & McArdle, J. J. (2015). Using R Package RAMpath for Tracing SEM Path Diagrams and Conducting Complex Longitudinal Data Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 132–147. https://doi.org/10.1080/10705511.2014.935257

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361–370.

Appendix A Chapter 3, additional material

A.1 Additional whole brain results

Correlation		Cam-CAN		LCBC	
		R	*p*	R	*p*
Age	Volume	-.62	<.0001	-.64	<.0001
	Thickness	-.6	<.0001	-.78	<.0001
	Area	-.36	<.0001	-.34	<.0001
Fluid Intelligence	Volume	.56	<.0001	.41	<.0001
	Thickness	.42	<.0001	.45	<.0001
	Area	.39	<.0001	.28	<.0001
Age-residualized FldIn	Volume	0.2	<.0001	.15	<.0001
	Thickness	.039	.33	.077	.0009
	Area	0.21	<.0001	.13	<.0001

Table 5-2 Comparing whole brain correlations in Cam-CAN and LCBC data

Metric	Model	R-Squared	F-Statistic	*p*	BIC
Cortical Volume	Linear	0.38	399.7	<0.001	5270.861
	Quadratic *	0.39	206.2	<0.001	5269.154
Cortical Thickness	Linear *	0.36	366.2	<0.001	5291.885
	Quadratic	0.37	184.2	<0.001	5296.544
Surface Area	Linear *	0.13	96.94	<0.001	5491.733
	Quadratic	0.13	48.56	<0.001	5497.909
Thickinthehead	Linear *	0.71	1538	<0.001	4796.678
	Quadratic	0.71	768.2	<0.001	4802.762
Curvature	Linear	0.60	955.2	<0.001	4996.165
	Quadratic *	0.63	532.4	<0.001	4959.439
Sulcal Depth	Linear *	0.14	106.2	<0.001	5483.685
	Quadratic	0.14	53.17	<0.001	5489.911

Grey Matter Volume (SPM)	Linear *	0.30	269.4	<0.001	5356.77
	Quadratic	0.30	135.2	<0.001	5361.933
Fractal Dimensionality	Linear *	0.42	467.6	<0.001	5230.34
	Quadratic	0.42	234.1	<0.001	5235.915

Table 5-3: Comparing linear and quadratic model fit for the metric-age correlations in Cam-CAN. The best fitting model (with lower BIC) is marked with *.

A.2 Additional regional results

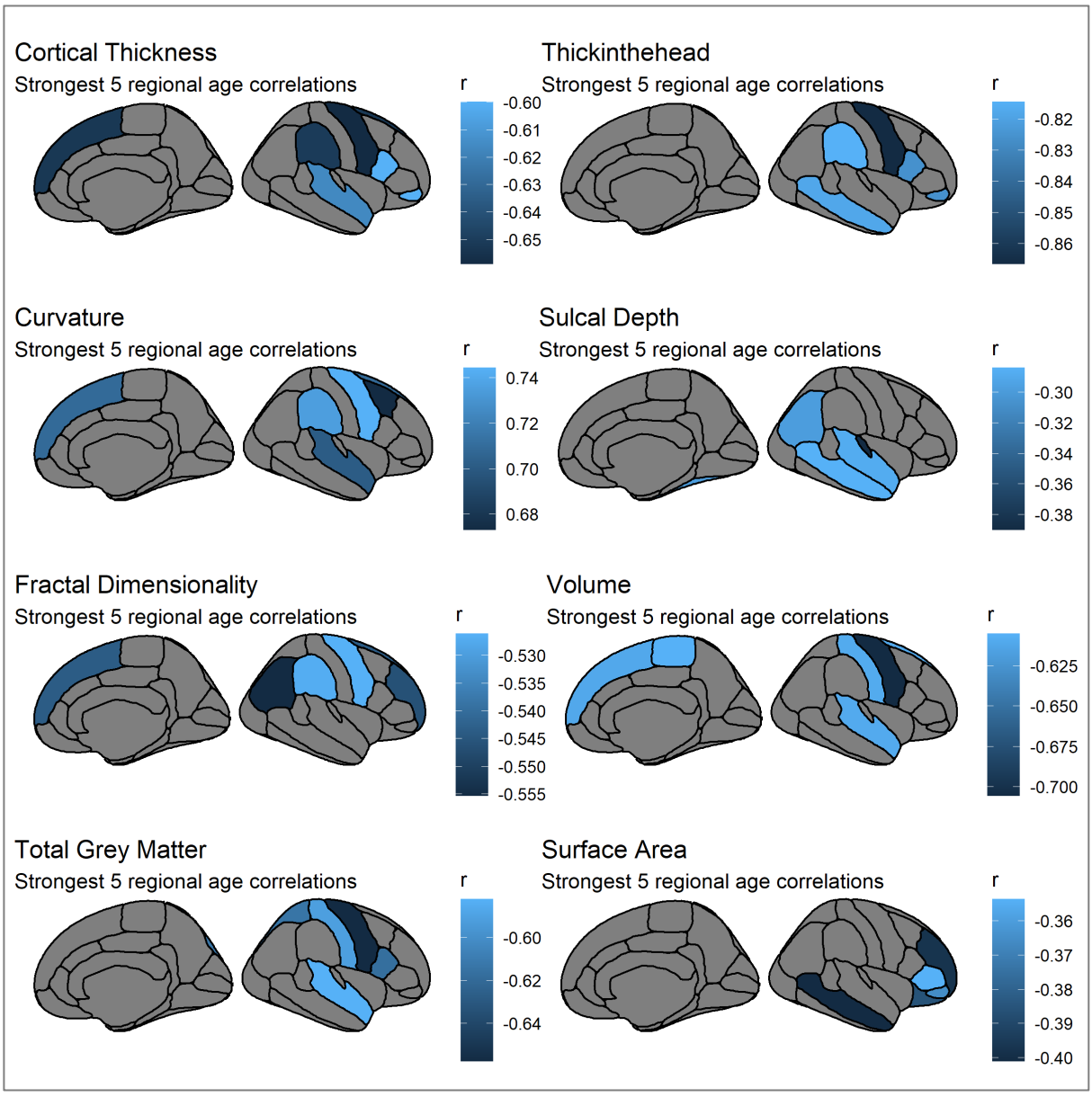


Figure 5-1: regions most strongly associated with age. Shows a large variability, with volume showing pre-frontal age effects while, for instance, sulcal depth effects are focused in the temporal lobes.

	Fractal Dim.		Curvature		Thickness		Thickinthhead		Volume		TGM		Depth		Area	
ROI	**r**	***p***	**r**	***p***	**r**	***p***	**r**	***p***	**r**	***p***	**r**	***p***	**r**	***p***	**r**	***p***
bankssts	NA	NA	NA	NA	NA	NA	-0.783	<.001	-0.496	<.001	NA	NA	NA	NA	NA	NA
caudal anterior cingulate	-0.285	<.001	0.589	<.001	-0.353	<.001	-0.649	<.001	-0.366	<.001	-0.443	<.001	0.131	0.002	-0.216	<.001
caudal middle frontal	-0.479	<.001	0.673	<.001	-0.569	<.001	-0.772	<.001	-0.501	<.001	-0.557	<.001	-0.118	0.005	-0.233	<.001
corpus callosum	-0.207	<.001	0.451	<.001	-0.239	<.001	-0.609	<.001	-0.361	<.001	-0.536	<.001	-0.033	0.494	-0.191	<.001
cuneus	-0.037	0.349	-0.036	0.365	-0.159	<.001	-0.357	<.001	0.016	0.685	-0.35	<.001	-0.008	0.894	-0.009	0.843
entorhinal	NA	NA	NA	NA	NA	NA	-0.313	<.001	-0.264	<.001	NA	NA	NA	NA	NA	NA
fusiform	-0.38	<.001	0.492	<.001	-0.305	<.001	-0.683	<.001	-0.374	<.001	-0.461	<.001	-0.298	<.001	-0.306	<.001
inferior parietal	-0.555	<.001	0.671	<.001	-0.585	<.001	-0.747	<.001	-0.558	<.001	-0.524	<.001	-0.298	<.001	-0.347	<.001
inferior temporal	-0.27	<.001	0.475	<.001	-0.209	<.001	-0.646	<.001	-0.293	<.001	-0.431	<.001	-0.216	<.001	-0.268	<.001
insula	-0.242	<.001	0.63	<.001	-0.536	<.001	-0.71	<.001	-0.423	<.001	-0.49	<.001	0.018	0.769	-0.004	0.929
isthmus cingulate	-0.352	<.001	0.523	<.001	-0.423	<.001	-0.76	<.001	-0.405	<.001	-0.387	<.001	-0.048	0.303	-0.165	<.001
lateral occipital	-0.414	<.001	0.519	<.001	-0.297	<.001	-0.647	<.001	-0.329	<.001	-0.467	<.001	-0.176	<.001	-0.254	<.001
lateral orbitofrontal	-0.389	<.001	0.396	<.001	-0.2	<.001	-0.627	<.001	-0.491	<.001	-0.502	<.001	-0.226	<.001	-0.386	<.001
lingual	-0.256	<.001	0.506	<.001	-0.321	<.001	-0.65	<.001	-0.343	<.001	-0.567	<.001	-0.14	0.001	-0.201	<.001
medial orbitofrontal	-0.239	<.001	0.335	<.001	-0.296	<.001	-0.55	<.001	-0.443	<.001	-0.541	<.001	-0.002	0.972	-0.226	<.001
middle temporal	-0.446	<.001	0.637	<.001	-0.534	<.001	-0.818	<.001	-0.544	<.001	-0.513	<.001	-0.284	<.001	-0.401	<.001
paracentral	-0.463	<.001	0.459	<.001	-0.578	<.001	-0.663	<.001	-0.605	<.001	-0.564	<.001	-0.039	0.409	-0.161	<.001
parahippocampal	-0.116	0.003	0.226	<.001	-0.149	<.001	-0.45	<.001	-0.354	<.001	-0.433	<.001	-0.062	0.17	-0.232	<.001
pars opercularis	-0.487	<.001	0.637	<.001	-0.6	<.001	-0.826	<.001	-0.597	<.001	-0.617	<.001	-0.063	0.17	-0.333	<.001
pars orbitalis	-0.397	<.001	0.252	<.001	-0.6	<.001	-0.826	<.001	-0.459	<.001	-0.523	<.001	-0.077	0.085	-0.365	<.001
pars triangularis	-0.508	<.001	0.564	<.001	-0.581	<.001	-0.797	<.001	-0.599	<.001	-0.525	<.001	-0.124	0.003	-0.354	<.001
pericalcarine	-0.118	0.003	0.487	<.001	-0.049	0.213	-0.604	<.001	-0.389	<.001	-0.46	<.001	-0.015	0.775	-0.055	0.178
postcentral	-0.469	<.001	0.632	<.001	-0.494	<.001	-0.773	<.001	-0.609	<.001	-0.591	<.001	-0.218	<.001	-0.055	0.178
posterior cingulate	-0.45	<.001	0.595	<.001	-0.459	<.001	-0.706	<.001	-0.529	<.001	-0.522	<.001	0.169	<.001	-0.341	<.001
precentral	-0.526	<.001	0.744	<.001	-0.659	<.001	-0.867	<.001	-0.706	<.001	-0.658	<.001	-0.105	0.014	-0.205	<.001
precuneus	-0.487	<.001	0.663	<.001	-0.559	<.001	-0.731	<.001	-0.526	<.001	-0.408	<.001	0.052	0.259	-0.266	<.001

rostral anterior cingulate	-0.316	<.001	0.379	<.001	-0.248	<.001	-0.597	<.001	-0.36	<.001	-0.53	<.001	-0.017	0.769	-0.227	<.001
rostral middle frontal	-0.546	<.001	0.597	<.001	-0.512	<.001	-0.674	<.001	-0.583	<.001	-0.56	<.001	-0.265	<.001	-0.398	<.001
superior frontal	-0.544	<.001	0.709	<.001	-0.653	<.001	-0.759	<.001	-0.611	<.001	-0.523	<.001	0.001	0.972	-0.313	<.001
superior parietal	-0.514	<.001	0.595	<.001	-0.491	<.001	-0.62	<.001	-0.562	<.001	-0.614	<.001	-0.071	0.114	-0.298	<.001
superior temporal	-0.446	<.001	0.701	<.001	-0.616	<.001	-0.62	<.001	-0.609	<.001	-0.582	<.001	-0.288	<.001	-0.332	<.001
supramarginal	-0.527	<.001	0.735	<.001	-0.651	<.001	-0.814	<.001	-0.532	<.001	-0.529	<.001	-0.131	0.002	-0.266	<.001
temporal pole	NA	NA	NA	NA	NA	NA	-0.469	<.001	-0.065	0.1	NA	NA	NA	NA	NA	NA
transverse temporal	-0.441	<.001	0.554	<.001	-0.403	<.001	-0.772	<.001	-0.523	<.001	-0.555	<.001	-0.39	<.001	-0.251	<.001

Table 5-4: Regional age correlations in Cam-CAN. All p-values are FDR corrected at alpha = 0.05.

	Fractal Dim.		Curvature		Thickness		Thickinthhead		Volume		TGM		Depth		Area	
ROI	**r**	***p***	**r**	***p***	**r**	***p***	**r**	***p***	**r**	***p***	**r**	***p***	**r**	***p***	**r**	***p***
bankssts	NA	NA	NA	NA	NA	NA	0.5674	<.001	0.444	0	NA	NA	NA	NA	NA	NA
caudal anterior cingulate	0.2698	<.001	-0.417	<.001	0.1704	<.001	0.4147	<.001	0.3295	<.001	0.4097	<.001	-0.0778	0.0635	0.2379	<.001
caudal middle frontal	0.3771	<.001	-0.471	<.001	0.4006	<.001	0.5299	<.001	0.4446	<.001	0.4864	<.001	0.1249	0.0033	0.2622	<.001
corpus callosum	0.215	<.001	-0.2576	<.001	0.1928	<.001	0.3885	<.001	0.346	<.001	0.4828	<.001	0.0748	0.072	0.2375	<.001
cuneus	0.1092	0.0064	-0.0358	0.371	0.1683	<.001	0.3039	<.001	0.0715	0.0733	0.3672	<.001	0.0789	0.0622	0.0963	0.0158
entorhinal	NA	NA	NA	NA	NA	NA	0.193	<.001	0.2084	<.001	NA	NA	NA	NA	NA	NA
fusiform	0.3402	<.001	-0.3918	<.001	0.228	0	0.492	<.001	0.3851	<.001	0.4307	<.001	0.2932	<.001	0.3356	<.001
inferior parietal	0.4256	<.001	-0.4731	<.001	0.4092	0	0.5132	<.001	0.4706	<.001	0.46	<.001	0.2483	<.001	0.3238	<.001
inferior temporal	0.2339	<.001	-0.3343	<.001	0.1356	<.001	0.4609	<.001	0.3132	<.001	0.4296	<.001	0.2008	<.001	0.2904	<.001
insula	0.2481	<.001	-0.4877	<.001	0.4297	0	0.5285	<.001	0.4425	<.001	0.4822	<.001	0.0743	0.072	0.1121	0.0051
isthmus cingulate	0.3602	<.001	-0.3941	<.001	0.2673	0	0.5197	<.001	0.4097	<.001	0.4044	<.001	0.0888	0.0367	0.2608	<.001
lateral occipital	0.3459	<.001	-0.3806	<.001	0.2138	0	0.4512	<.001	0.333	<.001	0.4508	<.001	0.1978	<.001	0.2771	<.001
lateral orbitofrontal	0.3265	<.001	-0.3219	<.001	0.1324	0.001	0.4521	<.001	0.4797	<.001	0.4834	<.001	0.1785	<.001	0.4013	<.001
lingual	0.2588	<.001	-0.395	<.001	0.2754	0	0.4362	<.001	0.3567	<.001	0.5064	<.001	0.1386	0.0012	0.2481	<.001
medial orbitofrontal	0.2589	<.001	-0.2022	<.001	0.206	0	0.3608	<.001	0.409	<.001	0.4988	<.001	0.1313	0.0022	0.2646	<.001
middle temporal	0.3497	<.001	-0.4748	<.001	0.3663	0	0.5803	<.001	0.474	<.001	0.4836	<.001	0.2195	<.001	0.3872	<.001
paracentral	0.3923	<.001	-0.3062	<.001	0.4319	0	0.4234	<.001	0.5066	<.001	0.4851	<.001	0.0934	0.0298	0.2189	<.001
parahippocampal	0.0971	0.015	-0.2133	<.001	0.0967	0.0158	0.3032	<.001	0.3136	<.001	0.4221	<.001	0.096	0.0262	0.2378	<.001
pars opercularis	0.3566	<.001	-0.4324	<.001	0.4052	0	0.5762	<.001	0.4932	<.001	0.5221	<.001	0.1024	0.0176	0.3045	<.001
pars orbitalis	0.3343	<.001	-0.1666	<.001	0.4052	0	0.5762	<.001	0.4301	<.001	0.4901	<.001	0.0484	0.2331	0.3523	<.001
pars triangularis	0.4014	<.001	-0.4317	<.001	0.3947	<.001	0.5717	<.001	0.5055	<.001	0.5126	<.001	0.1776	<.001	0.3417	<.001
pericalcarine	0.1584	<.001	-0.3028	<.001	0.0817	0.0407	0.3864	<.001	0.3562	<.001	0.4512	<.001	0.1093	0.0111	0.1201	0.0027
postcentral	0.3729	<.001	-0.3886	<.001	0.3695	<.001	0.54	<.001	0.5282	<.001	0.5172	<.001	0.2284	<.001	0.1201	0.0027
posterior cingulate	0.3972	<.001	-0.4792	<.001	0.27	<.001	0.4584	<.001	0.484	<.001	0.4692	<.001	-0.0528	0.1995	0.3768	<.001
precentral	0.4311	<.001	-0.5043	<.001	0.5033	<.001	0.6068	<.001	0.5988	<.001	0.5486	<.001	0.1299	0.0023	0.2782	<.001
precuneus	0.4034	<.001	-0.4873	<.001	0.4185	<.001	0.4999	<.001	0.4739	<.001	0.3961	<.001	0.0232	0.5615	0.296	<.001

rostral anterior cingulate	0.2956	<.001	-0.3167	<.001	0.1279	0.0014	0.4072	<.001	0.3444	<.001	0.5039	<.001	0.0574	0.1668	0.2567	<.001
rostral middle frontal	0.4321	<.001	-0.4144	<.001	0.333	<.001	0.4428	<.001	0.5018	<.001	0.5364	<.001	0.249	<.001	0.3703	<.001
superior frontal	0.4081	<.001	-0.4722	<.001	0.4384	<.001	0.5096	<.001	0.5343	<.001	0.4952	<.001	0.0893	0.0367	0.343	<.001
superior parietal	0.392	<.001	-0.4086	<.001	0.3547	<.001	0.4109	<.001	0.4642	<.001	0.5267	<.001	0.084	0.0477	0.2766	<.001
superior temporal	0.3721	<.001	-0.5254	<.001	0.4633	<.001	0.4109	<.001	0.5332	<.001	0.5393	<.001	0.2944	<.001	0.3448	<.001
supramarginal	0.4267	<.001	-0.5117	<.001	0.4574	<.001	0.5694	<.001	0.4669	<.001	0.4851	<.001	0.2104	<.001	0.2766	<.001
temporal pole	NA	NA	NA	NA	NA	NA	0.3933	<.001	0.1189	0.0029	NA	NA	NA	NA	NA	NA
transverse temporal	0.4103	<.001	-0.4615	<.001	0.3054	<.001	0.5635	<.001	0.4785	<.001	0.5139	<.001	0.3651	<.001	0.283	<.001

Table 5-5: Regional fluid intelligence correlations in Cam-CAN. All p-values are FDR corrected at alpha = 0.05.

	Fractal Dim.		Curvature		Thickness		Thickinthhead		Volume		TGM		Depth		Area	
ROI	**r**	***p***	**r**	***p***	**r**	***p***	**r**	***p***	**r**	***p***	**r**	***p***	**r**	***p***	**r**	***p***
bankssts	NA	NA	NA	NA	NA	NA	0.0464	0.7974	0.1451	<0.05	NA	NA	NA	NA	NA	NA
caudal anterior cingulate	0.1172	0.0071	-0.0261	0.8393	-0.0568	0.5147	0.0107	0.8935	0.1166	0.0039	0.1592	<0.05	0.0295	0.4757	0.1229	0.0024
caudal middle frontal	0.1048	0.0123	-0.0134	0.9842	0.0432	0.602	0.0303	0.8031	0.1667	<0.05	0.1471	<0.05	0.0924	0.0374	0.1657	<0.05
corpus callosum	0.1153	0.0071	0.0681	0.3866	0.0553	0.5147	0.0031	0.9507	0.1504	<0.05	0.1203	<0.05	0.0568	0.1715	0.1548	<0.05
cuneus	0.094	0.0249	-0.0273	0.8393	0.039	0.602	0.0725	0.6909	0.0858	0.0316	0.1504	<0.05	0.0866	0.0491	0.1114	0.0056
entorhinal	NA	NA	NA	NA	NA	NA	-0.0025	0.9507	0.1114	0.0057	NA	NA	NA	NA	NA	NA
fusiform	0.1311	0.004	-0.0691	0.3866	0.029	0.6909	0.0312	0.8031	0.19	<0.05	0.1711	<0.05	0.103	0.0222	0.1878	0
inferior parietal	0.0929	0.0253	-0.0094	0.9842	0.0337	0.6181	0.0219	0.8935	0.143	<0.05	0.1683	<0.05	0.0739	0.0902	0.1392	<0.05
inferior temporal	0.0773	0.0568	-0.0068	0.9842	0.0207	0.7455	0.0467	0.7974	0.1625	<0.05	0.1638	<0.05	0.1029	0.0222	0.1504	<0.05
insula	0.1365	0.0032	-0.0634	0.3866	0.1065	0.2351	0.0814	0.6909	0.2199	<0.05	0.195	<0.05	0.1255	0.0052	0.1579	<0.05
isthmus cingulate	0.1782	<0.05	-0.0477	0.4806	0.0139	0.7787	0.0256	0.8881	0.1898	<0.05	0.1937	<0.05	0.0629	0.1434	0.1956	0
lateral occipital	0.1047	0.0123	-0.0067	0.9842	0.0195	0.7455	0.0229	0.8935	0.1566	<0.05	0.1941	<0.05	0.1244	0.0052	0.1569	<0.05
lateral orbitofrontal	0.1215	0.0071	-0.0813	0.3866	0.0415	0.602	0.0654	0.6909	0.2129	<0.05	0.219	<0.05	0.0613	0.1434	0.1971	0
lingual	0.1367	0.0032	-0.0562	0.4117	0.0921	0.325	0.0316	0.8031	0.1713	<0.05	0.1418	<0.05	0.0447	0.2813	0.1506	<0.05
medial orbitofrontal	0.1477	0.0021	0.0305	0.8121	0.0341	0.6181	0.0113	0.8935	0.1634	<0.05	0.1834	<0.05	0.1904	0	0.1619	<0.05
middle temporal	0.0895	0.0298	-0.0535	0.4307	0.0377	0.602	0.0494	0.7974	0.1596	<0.05	0.158	<0.05	0.062	0.1434	0.1697	<0.05
paracentral	0.1171	0.0071	-0.0154	0.9842	0.06	0.5147	0.009	0.9025	0.1445	<0.05	0.1381	<0.05	0.1009	0.0222	0.1484	<0.05
parahippocampal	0.0472	0.2376	-0.0684	0.3866	0.0156	0.7727	0.0037	0.9507	0.1075	0.0075	0.1613	<0.05	0.0626	0.1434	0.1099	0.006
pars opercularis	0.0729	0.0702	-<0.05	0.9842	0.0231	0.7455	0.0372	0.7974	0.1335	0.001	0.1245	0.002	0.1008	0.0222	0.1258	0.002
pars orbitalis	0.1144	0.0071	0.003	0.9842	0.0231	0.7455	0.0372	0.7974	0.1926	<0.05	0.2176	<0.05	0.0241	0.5474	0.1737	0
pars triangularis	0.1166	0.0071	-0.0571	0.4117	0.0374	0.602	0.056	0.7974	0.1519	<0.05	0.2034	<0.05	0.1347	0.0037	0.1539	<0.05
pericalcarine	0.1115	0.0081	0.0694	0.3866	0.0607	0.5147	-0.0171	0.8935	0.1275	0.0017	0.166	<0.05	0.1321	0.004	0.1113	0.0056
postcentral	0.0824	0.0451	0.0496	0.476	0.0492	0.602	0.0415	0.7974	0.1488	<0.05	0.138	<0.05	0.1241	0.0052	0.1113	0.0056
posterior cingulate	0.1405	0.0032	-0.1061	0.2342	-0.0092	0.8456	0.0173	0.8935	0.1852	<0.05	0.1733	<0.05	0.0898	0.0419	0.202	0
precentral	0.1151	0.0071	-0.0146	0.9842	0.0778	0.3985	0.0532	0.7974	0.1731	<0.05	0.1482	<0.05	0.1022	0.0222	0.1955	0
precuneus	0.116	0.0071	-0.0423	0.5621	0.0687	0.5147	0.0315	0.8031	0.1649	<0.05	0.1831	<0.05	0.0791	0.0736	0.1589	<0.05

rostral anterior cingulate	0.1306	0.004	-0.0643	0.3866	0.0155	0.7727	0.0429	0.7974	0.1442	<0.05	0.1765	<0.05	0.0709	0.102	0.1286	0.0016
rostral middle frontal	0.1155	0.0071	-0.0024	0.9842	0.0019	0.9625	0.0126	0.8935	0.1746	<0.05	0.2204	<0.05	0.1347	0.0037	0.1671	<0.05
superior frontal	0.0926	0.0253	0.0068	0.9842	0.0197	0.7455	0.0198	0.8935	0.1833	<0.05	0.2126	<0.05	0.1392	0.0036	0.1958	0
superior parietal	0.0794	0.052	-0.0048	0.9842	0.0433	0.602	0.0114	0.8935	0.1177	0.0037	0.117	0.0035	0.074	0.0902	0.1054	0.0082
superior temporal	0.1117	0.0081	-0.059	0.4117	0.0787	0.3985	0.0114	0.8935	0.1808	<0.05	0.1569	<0.05	0.1452	0.0027	0.1891	0
supramarginal	0.117	0.0071	-0.0045	0.9842	0.0453	0.602	0.0408	0.7974	0.1592	<0.05	0.1725	<0.05	0.1519	0.0021	0.1475	<0.05
temporal pole	NA	NA	NA	NA	NA	NA	0.0795	0.6909	0.0942	0.0188	NA	NA	NA	NA	NA	NA
transverse temporal	0.1582	0.0011	-0.0969	0.2342	0.0612	0.5147	0.0695	0.6909	0.1677	<0.05	0.1158	0.0037	0.1282	0.005	0.1478	<0.05

Table 5-6: Regional age-residualized fluid intelligence correlations in Cam-CAN. All p-values are FDR corrected at alpha = 0.05.

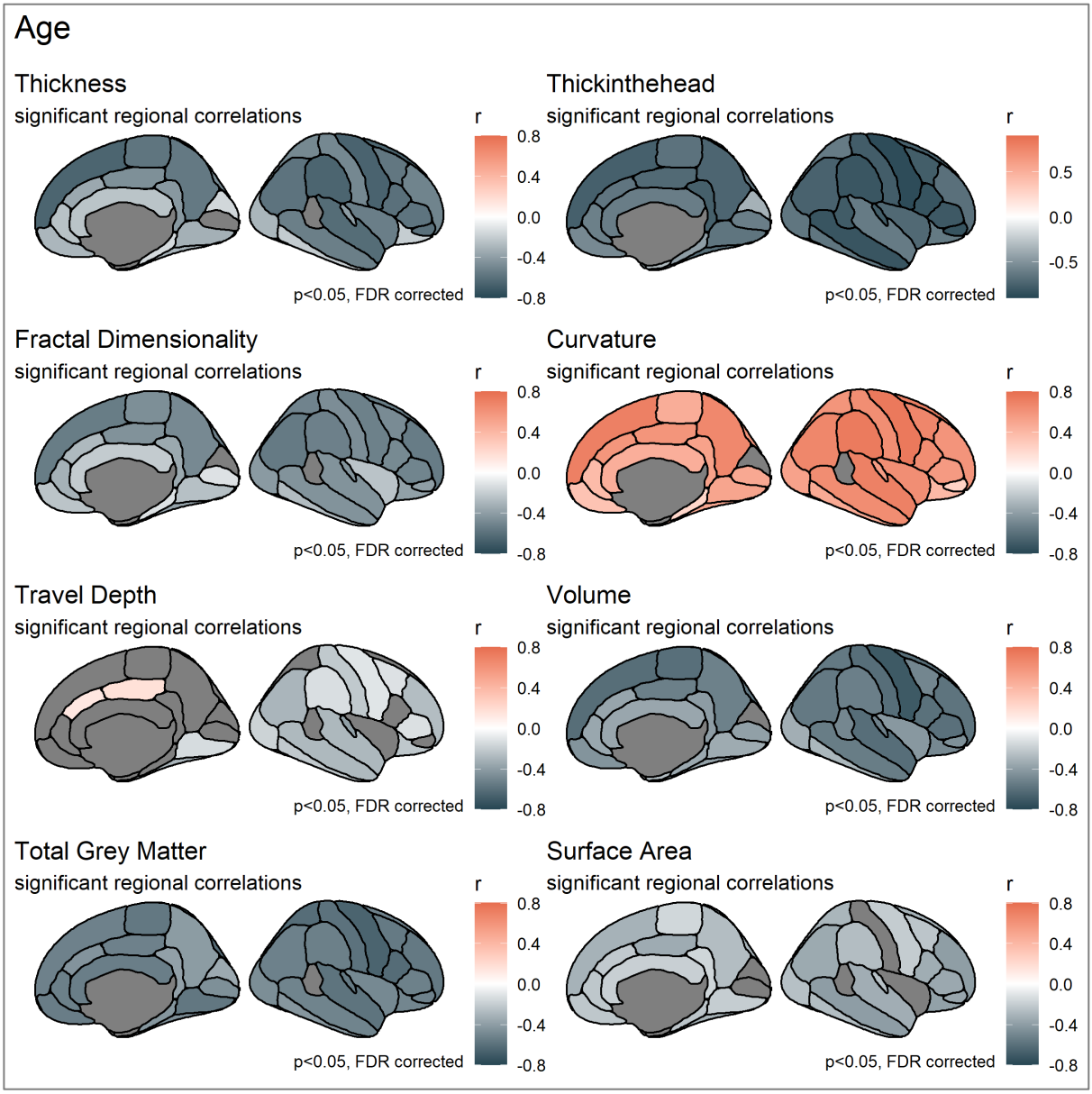


Figure 5-2: Significant regional age correlation for each metric. FDR corrected at alpha = 0.05.

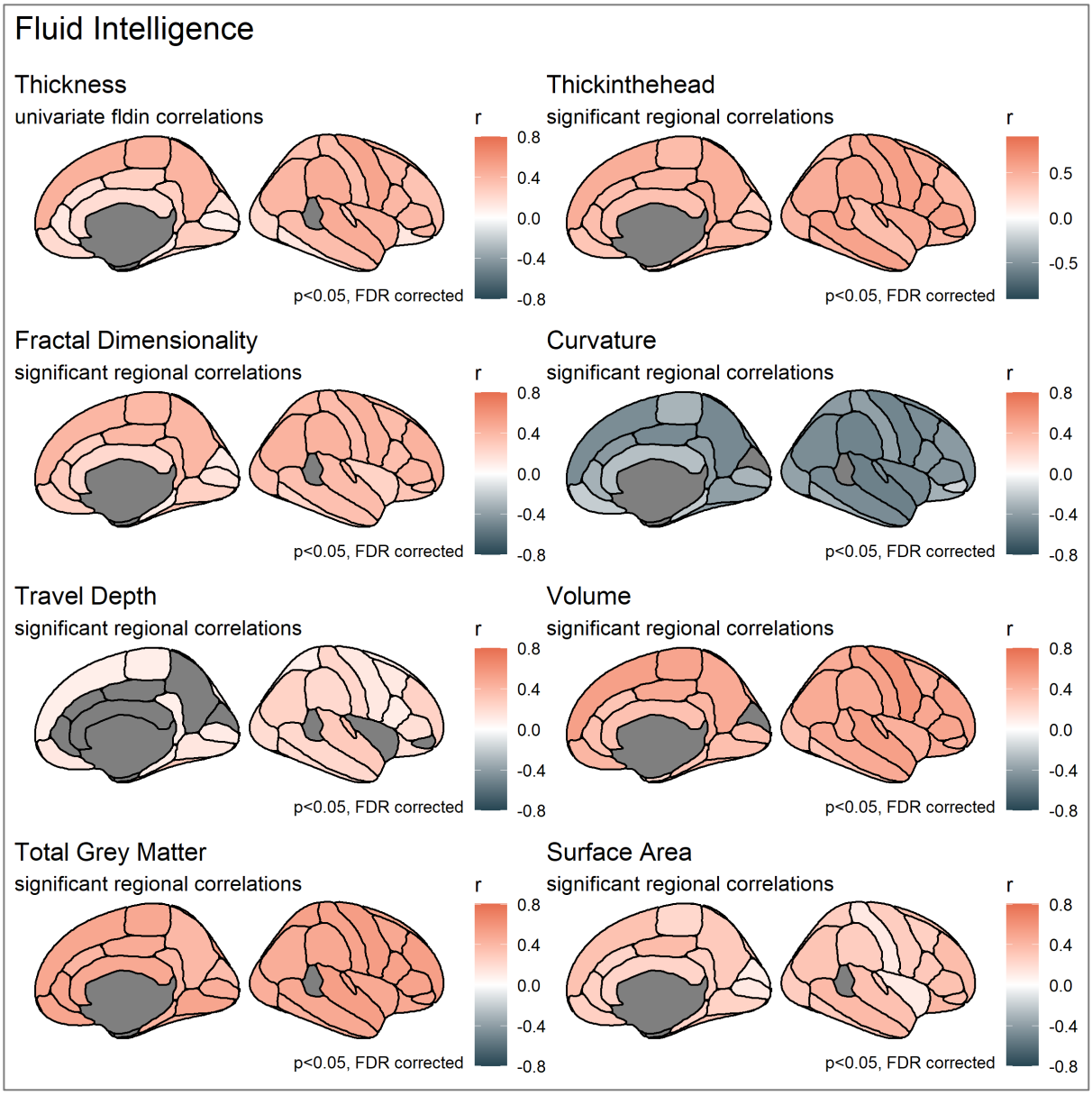


Figure 5-3: Significant regional fluid intelligence correlation for each metric. FDR corrected at alpha = 0.05.

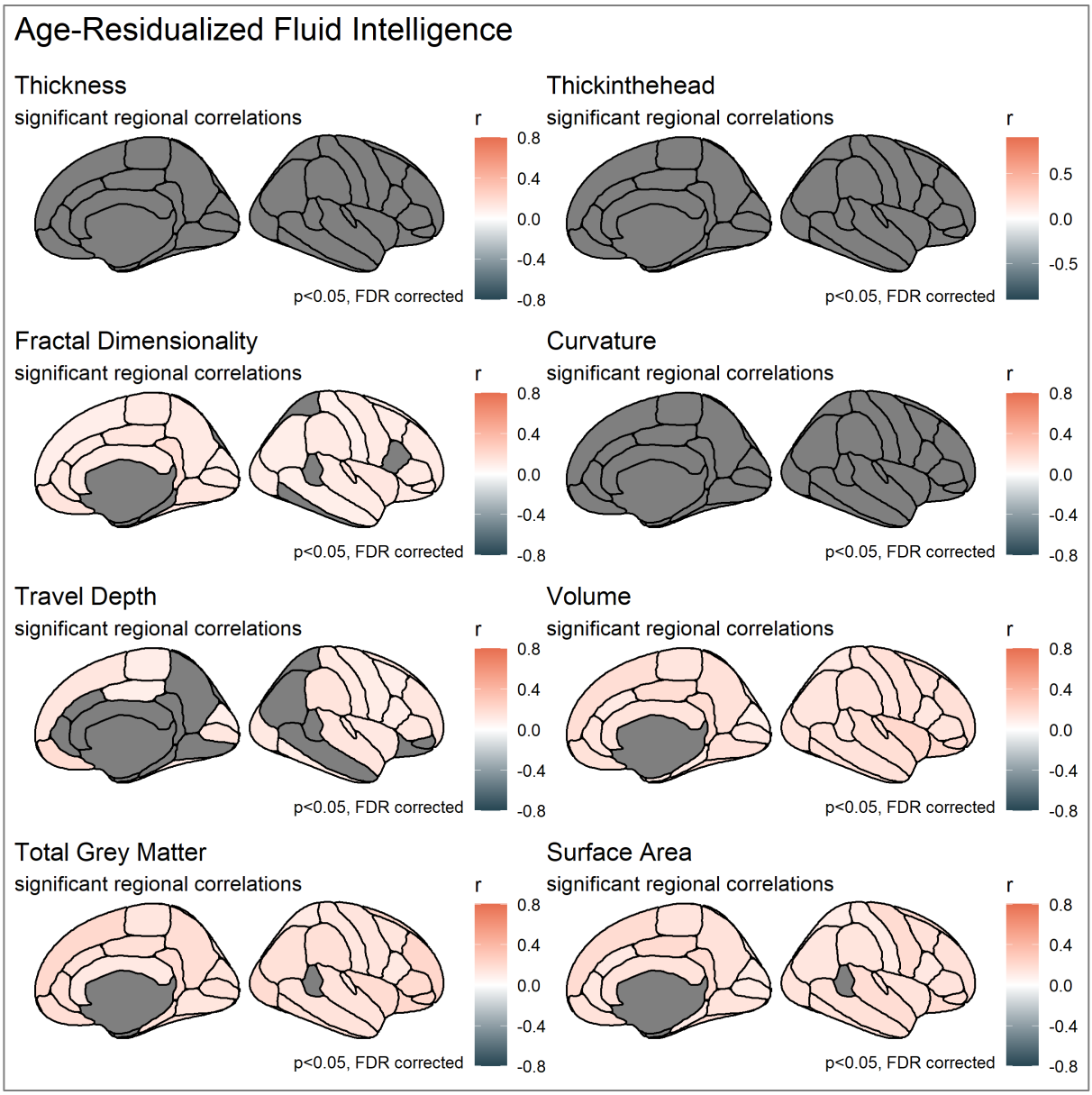


Figure 5-4: Significant regional age-residualized fluid intelligence correlation for each metric. FDR corrected at alpha = 0.05.

A.3 Additional longitudinal results

First, to assess whether Cattell test type (online versus pen/paper) made a difference, we tested for metric invariance and scalar invariance in the wave two cognitive data. This led to negligible drops in model fit (ΔCFI = 0.008 and 0.004 for metric and scalar invariance, respectively, Cheung & Rensvold, 2002), suggesting that assuming pencil and paper vs computer-based testing had equal measurement properties did not adversely affect the measurement of fluid intelligence. For all further analysis, this grouping factor was therefore ignored. Second,to ensure comparability of cognitive scores across Time 1 and Time 2, we tested for longitudinal measurement invariance (Widaman, Ferrer & Conger, 2010). We found that imposing invariance did not meaningfully decrease model fit (ΔCFI = 0.002; Cheung & Rensvold, 2002), suggesting longitudinal measurement invariance is tenable, and we were able to proceed to interpret change scores in the latent factor. Following the above inspections, we used Latent Change Score Models (LCSM) to examine morphometric and cognitive change over time.

	Time	*N*	Mean	Minimum	Maximum	*SD*	Skewness	Excess kurtosis
Age	T1	261	54.97	19.25	89	18.17	-0.02	-1.16
	T2	261	56.32	21.25	91.58	18.2	-0.03	-1.18
Cattell (sum score)	T1	215	32.50	12	44	6.06	-0.39	-0.10
	T2	215	30.42	10	44	6.65	-0.76	0.80
Surface Area	T1	261	2527.43	1896.25	3299.01	256.81	0.22	-0.22
	T2	261	2521.75	1898.46	3297.51	255.73	0.23	-0.21
Cortical Thickness	T1	261	2.61	2.28	2.89	0.1	-0.19	0.45
	T2	261	2.6	2.29	2.91	0.1	-0.19	0.3
Volume	T1	261	7175.41	5417.15	9412.12	822.25	0.44	-0.05
	T2	261	7124.88	5342.85	9311.37	824.73	0.42	-0.05

Table 5-7: Cam-CAN raw scores and descriptive statistics for age, Cattell and longitudinal brain structure metrics

Cam-CAN		Model Fit Indices					
Metric	**Model**	**χ2**	***p***	**RMSEA [90 % CI]**	**CFI**	**SRMR**	**Yuan-Bentler scaling factor**
Thickness	**FIML**	5.275	0.072	0.039 [0.000, 0.072]	0.992	0.026	0.763
Surface Area	**FIML**	4.228	0.121	0.033, [0.000, 0.079]	0.997	0.015	0.721
Volume	**FIML**	3.655	0.161	0.028 [0.000, 0.065]	0.995	0.014	1.468

Table 5-8: Second order latent change score model fit indices Cam-CAN.

	Model Fit Indices					
Metric	**χ^2**	***p***	**RMSEA [90 % CI]**	**CFI**	**SRMR**	**Yuan-Bentler scaling factor**
Thickness	13.605	0.001	0.090 [0.050, 0.135]	0.993	0.038	1.070
Surface Area	2.418	0.298	0.033, [0.000, 0.079]	0.999	0.007	1.091
Volume	47.648	0.000	0.178 [0.133, 0.227]	0.975	0.034	0.845

Table 5-9: Second order latent change score model fit indices LCBC.

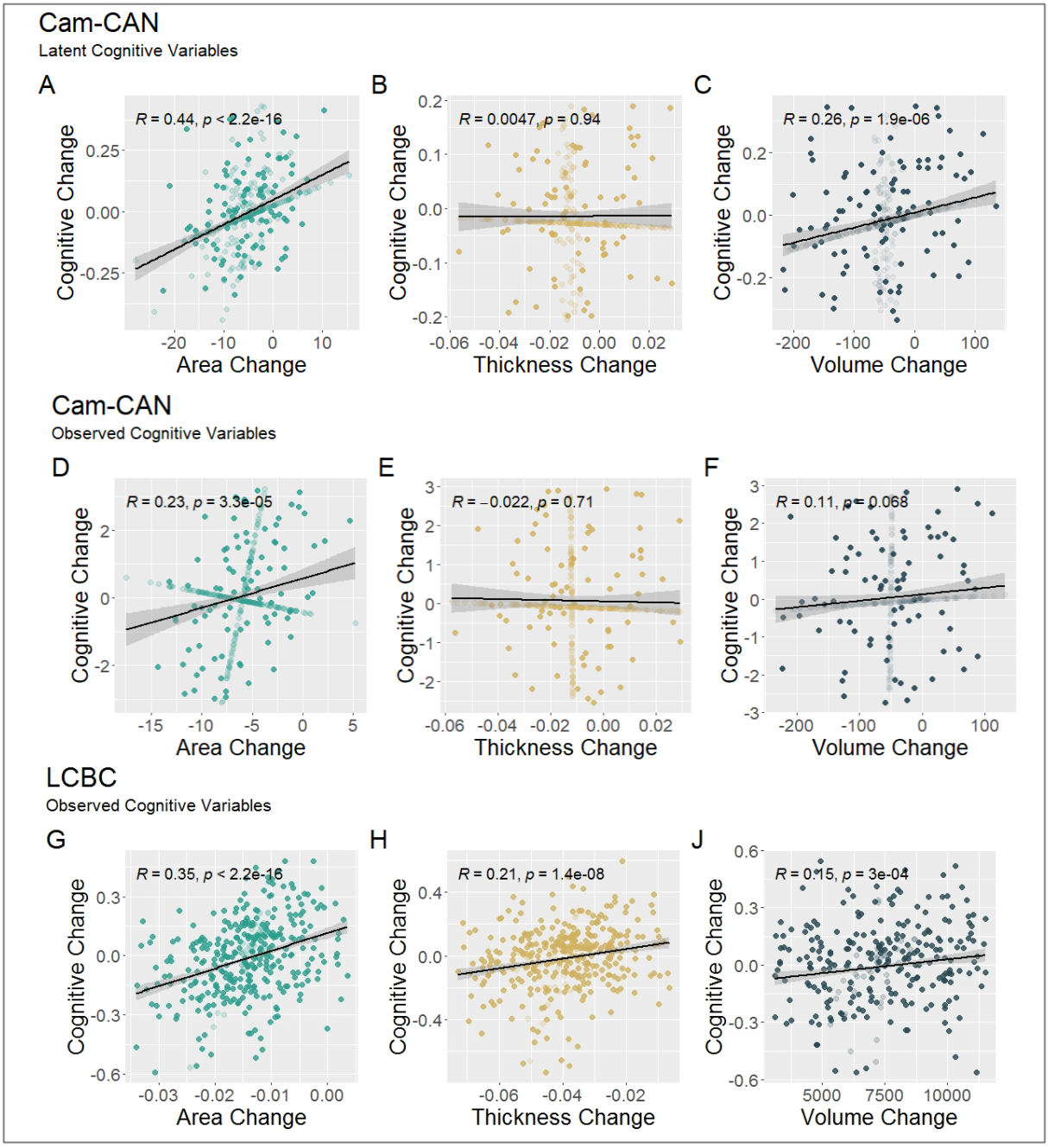


Figure 5-5: correlations of cognitive change and neural change in Cam-CAN (A-F) and LCBC (G-J). Shows that change in surface area is most strongly associated with cognitive change. Models A-C include latent cognitive variables, which were not possible to derive from the LCBC data, where we used observed cognitive scores instead. To compare like-for-like models, we include Cam-CAN observed variable models here, too (D-F). Note that the shaded dots are the models' missingness estimates.