Measuring cognitive maintenance in older adults and its association with education and other cognitively stimulating activities

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy

in the Faculty of Humanities

2019

Benjamin D Williams

Faculty of Humanities

List of Contents

Word Count: Approximately 61,500

List of Tables

List of Figures

List of Appendices

Abstract.

Education and other cognitively stimulating activities (CSA) are potentially modifiable factors which may improve cognitive maintenance in later life. As these social exposures are not possible to effectively randomise, inferences from observational data are particularly important. Existing research findings differ regarding whether CSA are associated with cognitive maintenance. This may be explained in part by limitations and implicit assumptions in the most commonly used methods to analyse these associations.

This thesis asks how modifiable social exposures affect cognitive maintenance and examines some of the assumptions underlying standard methods such as growth modelling. The analysis uses the English Longitudinal Study of Ageing (ELSA), a nationally representative longitudinal cohort study of adults aged over 50 living in England. A series of assumptions in standard regression approaches and their implications for the association between education or CSA and cognitive maintenance are examined. Firstly, ELSA's scoring method for memory and executive function is examined using factor analysis. The memory score performs well, but the executive function score does not reflect the data. This leads to incorrect estimation of the association between cognitive maintenance and some important predictors such as age. I then tested for longitudinal measurement invariance (MI) in the cognitive factors and found this did not hold for memory in ELSA using Bayesian approximate MI. This is an advance on conventional tests of MI which had found equivocal results. The assumption that the ELSA sample is drawn from one homogenous population, and that the effect of education on cognitive maintenance is the same across sub-populations, were then tested using growth mixture modelling. A small beneficial effect of higher educational attainment on cognitive maintenance was found in a stable cognition latent class but no association was seen in latent classes with declining cognition.

If CSA participation improves cognitive maintenance, and better cognition increases the likelihood of participation in CSA, this generates time varying confounding affected by prior exposure. Standard growth curves must assume this to be absent. Using inverse probability of treatment weighted marginal structural models to relax this assumption, volunteering and internet use activities were still found to reduce the risk of dementia or cognitive impairment.

This research contributes methodologically to the existing literature by demonstrating how some of the assumptions underpinning the regression models most commonly used to estimate the association between CSA and cognitive maintenance can influence the substantive conclusions drawn. Specifically, it finds that ELSA's executive function index does not represent the data well and Bayesian approximate MI can be used to clarify equivocal conventional tests of longitudinal measurement invariance of the cognitive test factors. Substantively, I find that the effect of education on cognitive maintenance varies somewhat depending on underlying trajectory, and that the association of volunteering and internet use activities with improved cognitive maintenance is robust after time varying confounding is accounted for.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Statement on the Paper Format

By undertaking a paper format thesis, the empirical work contained in this thesis forms three separate journal articles that have each been submitted or prepared for publication. I am the first or the sole author on the four articles. I am responsible for the entire research by developing the research ideas, conducting empirical analyses, drafting and revising the articles. My supervisors are listed as co-authors in recognition of their contributions to my work. Each of my supervisors contributed to the development of my ideas, provided helpful guidance on the analysis, and reviewed and revised the drafts of the papers. All authors have read and approved the final manuscript.

Acknowledgements

The undertaking of this thesis whilst also working clinically, taking clinical exams and starting a new family has made for a wonderful but tiring 5 years. I thank Prof. Tarani Chandola and Prof. Neil Pendleton for their supervision. You have offered me great knowledge, wisdom and immense patience throughout my doctoral studies. I hope that this thesis repays your efforts and the kindness you have shown me.

I thank Prof. Gindo Tampubulon for his regular contribution to my annual reviews. The incisive, constructive and balanced feedback you have given me has been of immense value in improving my work. I thank the ESRC for their sponsorship of the MSc and PhD programme which has enabled me to pursue my academic interests alongside my clinical work.

I thank my children for the joy and encouragement they give me and the understanding they have shown at such a young age. I thank Lianna, my wife, for her love, patience and unwavering practical, emotional and financial support throughout. Without you my life would be poorer in every way.

I thank the Economic and Social Research Council for providing the funding to conduct this work as part of the 1+3 MSc and PhD Programme [ES/J500094/1]. The ELSA data were made available through the UK Data Archive. ELSA was developed by a team of researchers based at the NatCen Social Research, University College London and the Institute for Fiscal Studies. The data were collected by NatCen Social Research. The funding is provided by the National Institute of Ageing in the United States, and a consortium of UK government departments co-ordinated by the Office for National Statistics. The developers and funders of ELSA and the Archive do not bear any responsibility for the analyses or interpretations presented here.

The Author.

Benjamin D Williams is a trainee psychiatrist who has been working on this thesis less than full time alongside core clinical training. Prior to the PhD he undertook the MSc in Social Statistics and Research Methods at the University of Manchester and has a BM,BS from the Peninsula College of Medicine and Dentistry.

Before the MSc he had limited research experience. He had previously spent a few weeks with an epidemiology department as an undergraduate. However, this brief placement generated the interest in the social determinants of health and applied statistics which has ultimately resulted in the thesis you find before you.

1. Introduction

1.1 Overview and motivation

Dementia and cognitive impairment are amongst the greatest challenges in contemporary global health.(Livingston et al., 2017; Prince et al., 2013; Vos et al., 2012) Dementia is now one of the most feared illnesses in western nations.(Cutler, 2015; Wortmann, Andrieu, Mackell, & Knox, 2010) This thesis aims to make its own small contribution to understanding the relationship of social exposures to the maintenance of cognitive function and prevention of dementia. Whilst many studies have addressed this topic, the study of social exposures on cognitive function at a population level poses a range statistical and methodological challenges which leave many questions open.

Newer methods allow the processes underlying cognitive change to be modelled with less restrictive assumptions. This allows some of these open questions about how social exposures effect cognitive maintenance to be answered in new ways. In this thesis, the social exposures of interest are education in earlier life and cognitive stimulating activities in later life. Both are social exposures which stimulate mental activity and may promote better cognitive maintenance. However, as these exposures occur at very different times in the life-course, they require different analytic approaches to ultimately achieve the same ends of estimating their association with cognitive maintenance whilst overcoming modelling restrictions in the current literature.

There continues to be substantial debate about the theoretical framework used to understand underlying changes in cognitive function in later life.(Arenaza-Urquijo & Vemuri, 2018; Cabeza et al., 2018; Stern et al., 2018) Nevertheless, cognitive maintenance, reserve and related concepts form the lens through which cognitive ageing will be viewed. It is important groundwork needed to understand the way social exposures are related to cognitive functioning, and will be reviewed in the section 1.2.

After this review we will then turn to the literature informing the specific social exposures addressed in this thesis. In section 2.1 the literature will be reviewed on how education influences cognitive maintenance in later life. Past research has come to competing conclusions about how this early life exposure affects later life cognitive function.(X. Meng & D'Arcy, 2012; Valenzuela & Sachdev, 2006) I will then review a more specific literature including only those studies which have tested this association using methods to account for population heterogeneity.

Before being able to answer my question about what association education has with cognitive maintenance, I needed to establish how I was going to measure cognitive maintenance. I had to ensure I was using the available data on cognitive function in a valid way. In the English Longitudinal Study of Ageing (ELSA) dataset which I used for this thesis there is a battery of cognitive tests with a pre-specified scoring system.(Steel, Huppert, McWilliams, & Melzer, 2004) The measurement properties of this scoring system had not previously been examined. It was therefore important to establish whether the scoring system accurately reflected the data. The

structure of the data was tested using latent variable modelling. I identified that there were some discrepancies between the scoring system and the factor analysis. The question then became whether this discrepancy was enough to make a substantively important difference. This forms the substance of the first paper presented as part of this thesis.

Having established the factor structure I needed to establish if that structure remained constant over time. This is known as longitudinal measurement invariance (MI).(van de Schoot, Lugtig, & Hox, 2012) Without this MI it is not possible to know whether changes over time are due to changes in cognitive function itself or the performance of the tests used to measure it. In my analysis of ELSA the results from the standard tests for measurement invariance were ambiguous. To solve this, I turned to an alternative Bayesian approximate measurement invariance approach.(van de Schoot et al., 2013) This was able to define how much measurement invariance was present and which parameters were the invariant ones. I was then able to conclude that the property of measurement invariance did not hold for the memory factor I had identified in ELSA. It did hold for an orientation factor, but this factor had a low ceiling and therefore was not likely to be as informative as I would wish for analysing longitudinal change. So, in order to examine change over time I needed to use single cognitive measures separately, rather than combined as either a score or latent variable. These analyses are presented in the second paper in this thesis. I was then able to turn to the substantive question of whether education moderated cognitive decline within latent class of decline and the implications this has for cognitive maintenance. This is the focus of the third paper presented in this thesis.

Population heterogeneity is a situation which arises when within a population there are multiple subpopulations, but that membership of that subpopulation is not directly observable. This is the case in studies of cognitive function in older adults. As the pathology of common dementia's pre-dates the onset of symptoms by many years, within population samples of older adults there is likely to be at least two major clinically relevant sub-populations.(Braak & Del Tredici, 2015) Those with a developing dementia pathology (even if this is unobserved) and those without. As I will show in section 2.2, in studies of cognitive function which have accounted for population heterogeneity nearly all have used education to predict latent class or adjusted observed cognitive scores by education level (figure 1.1.A). This makes implicit assumptions about the underlying causality which do not appear to represent the clinicopathological research.(Brayne et al., 2010; Koepsell et al., 2008; Roe, Xiong, Miller, & Morris, 2007; Serrano-pozo et al., 2013) However, level of education and the cognitive reserve this provides could modify the trajectory someone is on, even if it does not change the likelihood of the underlying pathology. Attending to this translation of clinicopathological research into population research may provide valuable insights into how education contributes to cognitive maintenance in later life. It is made possible by advances in structural equation modelling which have enabled more complex representations of substantive theory and causal structures.

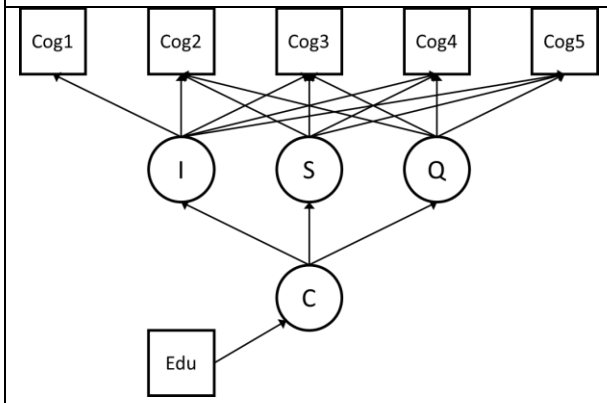
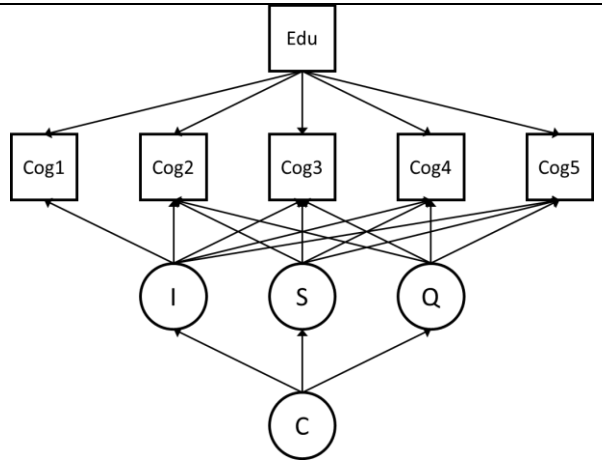
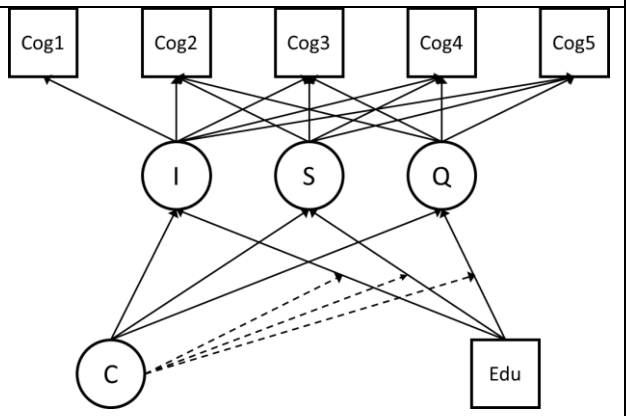
Having tested a more nuanced view of how education in early life is associated with cognitive function, I maintained an emphasis on the representation of causal structures but moved to the maintenance cognitive function in later life. Specifically, in the fourth paper I tested whether or

not engaging in cognitively stimulating activities (CSA) in later life could reduce the risk of meeting the threshold for cognitive impairment or dementia after 10 years of follow-up. CSA are an important potential protective factor against dementia or cognitive impairment.(Sajeev et al., 2016) Whereas the first substantive question utilises developments in structural equation modelling to address population heterogeneity, the second uses developments in marginal structural modelling to address another methodological challenge, time varying confounding influenced by prior exposure.(Daniel, Cousens, De Stavola, Kenward, & Sterne, 2013; Robins, Hernan, & Brumback, 2000) This is a critical question in observational studies of cognitive maintenance because of the potential for bi-directional causality. It has the potential to bias estimates of the effect of exposures over time either towards or away from the null and standard regression approaches with covariate adjustment are unable to account for this complex causal structure.

Marginal structural models (MSMs) were developed as an alternative to standard regression approaches which can account for time varying confounding influenced by prior exposure.(Daniel et al., 2013; Robins et al., 2000) Under strong assumptions MSMs can provide causal estimates from observational data. Despite the advantages they offer, they have seen little application in the study of cognitive ageing. They have not been used to test if time varying confounding explains the association between CSA and cognitive function. Due to the practical and ethical difficulties in performing randomised trials of CSA in older adults, making appropriate inferences from observational studies is of particular importance in this area.

Throughout the thesis I will be using data from the English Longitudinal Survey of Ageing. This is a large cohort study with a sample which was nationally representative of the English population aged over 50 at recruitment.(Steptoe, Breeze, Banks, & Nazroo, 2013) Data on cognitive function and a wide range of medical, behavioural and social covariates is collected in biennial waves. Starting in 2002 I will be using up to wave 5 for the first 3 papers and up to wave 7 for paper 4, due to changes in cognitive tests used in the study at various times.

Figure 1.1. Structural equation models demonstrating the theoretical models for how education may affect cognitive function in summary.

1.1.A Education predicts latent class.	1.1.B Observed scores adjusted for education.	1.1.C Association with education moderated by class.
		

C indicates latent class of change, Edu the level of educational attainment of an individual, Cog1-Cog5 a test of cognitive function on 5 measurement occasions, I is the latent intercept or baseline cognitive function, S is the latent linear rate of change and Q the quadratic term for latent growth. Boxes indicate observed variable's; circles latent variables and lines indicate potential associations with implied direction of causality. Residual variances and covariance's are omitted for simplicity. 1.1.A shows a model where education predicts latent class and, if class is driven by pathological status, assumes that education influences risk of dementia pathology. 1.1.B Represents what is done when cognitive scores are adjusted for education. It shows how this assumes education to affect test performance but not 'true' latent cognitive function or disease status. 1.1.C shows the theoretical model for educations effect on cognitive function used in this analysis. Latent class is independent of education but moderates the effect of education on latent cognitive function.

1.2 Cognitive Reserve, Maintenance and Related Concepts

In the past year three separate groups have proposed differing visions of the key concepts underlying cognitive function in later life.(Cabeza et al., 2018; Stern et al., 2018) These documents represent a shift from the concepts present in the literature when the thesis was first conceptualised and much of the analysis was conducted. However, they work well to refine the concepts important to this thesis, at which time cognitive reserve was the preferred umbrella term. With these extensive expert consensus documents recently published an additional literature review of this specific areas is of limited use. Here I will present summaries of the models proposed by each group and the evidence supporting those models. I will then seek to highlight areas of similarities and differences between them. Two of the reviews discuss research primarily in the context of healthy older adults but the principles apply equally to pre-clinical or clinical states of pathological change as shown in the third review.(Arenaza-Urquijo & Vemuri, 2018) Before this I will briefly summarise the theoretical model under which the thesis was originally developed.

Based on the work of Stern, cognitive reserve was principally divided into passive (brain) reserve and active (cognitive reserve).(Stern, 2002, 2012) Passive reserve was conceptualised in quantitative anatomical terms such as the number of neurones or synapses possessed. It moderated the expression of age or disease related cognitive change by increasing baseline cognition function. It is passive in the sense that it is not associated with rate of decline. Active reserve was viewed as primarily related to functional networks. It moderates the expression of age or disease related changes by modifying the rate of change in cognitive function. This could be either slower or more rapid depending on the theory adopted for a given cognitive function. As will be seen in the discussion below active and passive reserve remain largely consistent with more recent conceptualisations. However, they have now been subdivided into more nuanced models, joined by other complementary processes and moved into broader frames of reference.

One of those alternative frames is Arenaza-Urquijo and Vemuri preference for the terms resistance and resilience.(Arenaza-Urquijo & Vemuri, 2018) These terms are defined in the context of Alzheimer's disease neuropathology (ADP), meaning the presence or absence of significant levels of abnormal tau or amyloid-beta (Aβ) in brain tissue.(Dubois et al., 2014, 2016) They define brain resistance as "avoiding the appearance of ADP" and brain resilience as "an individual's ability to sustain a better-than-expected cognitive performance in relation to the degree of ADP". Making this distinction is important for understanding the role of protective and risk factors for dementia and cognitive impairment and they can be linked to difference phenotypic traits.

An individual with high resistance will have a lower level of ADP than expected for their chronological age and, conversely, an individual with low resistance with have a higher level. Resistance does not relate to how well one performs with a given level of pathological change but the quantity of pathology present. Resistance to Alzheimer's pathology is therefore determined by the relative rates of tau and Aβ deposition and clearance.(Arenaza-Urquijo &

Vemuri, 2018) I would add to this the rate of accumulation of cerebrovascular disease, which has both a direct effect on cognition itself and is also strongly associated with Aβ and tau pathology.(Koncz & Sachdev, 2018) Whilst the exact mechanisms of resistance and resilience may be different between the two pathologies the conceptual framework applies equally well for any dementia pathology (DP).

In Arenaza-Urquijo and Vemuri's terminology brain resistance encompasses neuroprotection, brain maintenance, neural efficiency and neural cognitive reserve. Neuroprotection is a broad term which has been used in several different ways. I use it to refer to any process which either reduces the degree or frequency of pathological insult. For example smoking is thought to contribute to DP via increased oxidative stress.(Durazzo, Mattsson, & Weiner, 2014) Brain maintenance is distinguished from metabolic maintenance and structure maintenance. Brain maintenance is the term used specifically to the brain's response to the primary pathological insult, not its cognitive consequences. An example of brain maintenance would be the differential rates of Aβ clearance in carriers of different apolipoprotein E alleles.(Castellano et al., 2011) Exercise is an example of an exposure which acts via both these pathways, it is neuroprotective in reducing cerebrovascular disease and promotes brain maintenance though the stimulation of growth factors in the hippocampus.(Cotman, Berchtold, & Christie, 2007) Greater neural efficiency contributes to brain resistance because it appears to be related to lower Aβ deposition.(Jagust & Mormino, 2011) However, it also contributes to brain resilience in that more efficient functional networks result in improved cognitive function in the presence or absence of ADP.(Barulli & Stern, 2013; Weiler et al., 2018)

For some authors this places neural efficiency primarily within brain resilience.(Fischer, Wolf, & Fellgiebel, 2019) Similarly, Arenaza-Urquijo and Vemuri include neural cognitive reserve in brain resistance but, as neural cognitive reserve is typically defined, it has conventionally been thought of as a source of cognitive reserve.(Arenaza-Urquijo & Vemuri, 2018; Barulli & Stern, 2013) The conventional definition of cognitive reserve as "differences in cognitive processes as a function of lifetime intellectual activities and other environmental factors that explain differential susceptibility to functional impairment in the presence of pathology or other neurological insult".(Barulli & Stern, 2013) This is clearly very similar to that of brain resilience above. Brain resilience as defined above by Arenaza-Urquijo and Vemuri is, in effect, cognitive reserve as applied in the context of DP only. Considerable confusion is occasioned by the frequent use of the term 'cognitive reserve', or simply 'reserve', to refer to multiple different concepts. I will therefore be using the term brain resilience as the name of this umbrella category with reference to both ageing and disease.

Neural cognitive reserve is one of a range of subordinate categories and mechanisms thought to contribute to brain resilience in total. It is here where we must address the alternative terminologies and differing theoretical models proposed by Cabeza et al. (the "McGill Group") and Stern et al. (the "International Society to Advance Alzheimer's Research and Treatment (ISTAART) group").(Cabeza et al., 2018; Stern et al., 2018) Both groups propose a division of brain resilience into three main sub-categories. The ISTAART group propose categories of

cognitive reserve, brain reserve and brain maintenance.(Stern et al., 2018) The McGill group propose using the terms neurocognitive reserve, maintenance and compensation.

As the area which appears to show greatest agreement, we will start with brain maintenance. The ISTAART group define brain maintenance as “reduced development over time of age-related brain changes and pathology based on genetics or lifestyle”.(Stern et al., 2018) This definition of brain maintenance appears to fall solely within the broader domain of brain resistance. The McGill group drops the specification of brain and defines maintenance more broadly as “the preservation of neural resources, which entails ongoing repair and replenishment of the brain in response to damage incurred at the cellular and molecular levels”.(Cabeza et al., 2018) This definition includes brain resistance to developing disease but also encompasses maintenance processes which attempt to preserve neuronal function in the presence of disease, and are therefore part of brain resilience. This can be further divided into structural and metabolic maintenance. An example of structural brain maintenance is that greater maintenance of white matter tract integrity is associated with better cognitive function in health or disease.(Fischer et al., 2019) An example of metabolic brain maintenance is the protective down-regulation of energy consumption observed with ADP.(Sun, Feng, Liang, Duan, & Lei, 2012) Clarity regarding whether maintenance is in relation to the primary disease process or responses to preserve function in the face of neuronal insults is important for understanding different targets for intervention to improve cognitive function. Despite the use of a singular term by both groups, there is important difference in the theoretical frameworks used.

The difference discussed in most detail in correspondence between the authors themselves is whether neurocognitive reserve should be viewed as a singular concept or a distinction drawn between cognitive reserve and brain reserve.(Cabeza et al., 2019; Stern et al., 2019) The McGill group define neurocognitive reserve as “a cumulative improvement, due to genetic and/or environmental factors, of neural resources that mitigates the effects of neural decline caused by ageing or age-related diseases”.(Cabeza et al., 2018) They further specify that “reserve is hypothesized to result in the accumulation of neural resources before the brain is affected by age-related processes and to take place over a period of years”.(Cabeza et al., 2018) This is a broad definition which emphasises that all cognition is brain based. It does not seek to draw distinctions between brain structure and function.

The ISTAART group split neurocognitive reserve into brain reserve and cognitive reserve. Brain reserve is defined as “neurobiological capital (numbers of neurons, synapses, etc.)”.(Stern et al., 2018) This structural differences in reserve were previously known as passive reserve. Brain reserve is passive in the sense that, whist it may be increased over the life-course, it is fixed at a given point in time. It moderates the expression of DP by increasing the time to the clinical expression of cognitive or functional impairment without affecting the underlying disease or ageing process. In a longitudinal study of cognitive function greater levels of brain reserve would be seen to increase baseline functioning but not to affect the rate of decline.

Cognitive reserve is defined as “the adaptability (i.e., efficiency, capacity, flexibility) of cognitive processes that helps to explain differential susceptibility of cognitive abilities or day-to-day

function to brain ageing, pathology, or insult".(Stern et al., 2018) This is a theoretical construct for the sum of additive and emergent effects resulting from "networks of brain regions associated with performing a task and the pattern of interactions between these networks".(Stern et al., 2018) This was previously associated with the term active reserve and emphasises dynamic function capacity to respond to pathological or age-related changes. Unlike brain reserve, cognitive reserve is predicted to moderate cognitive decline as the brain responds to advancing pathological changes. However, different neural implementations of cognitive reserve lead to differing potential trajectories of cognitive change over time.(Lenehan, Summers, Saunders, Summers, & Vickers, 2015)

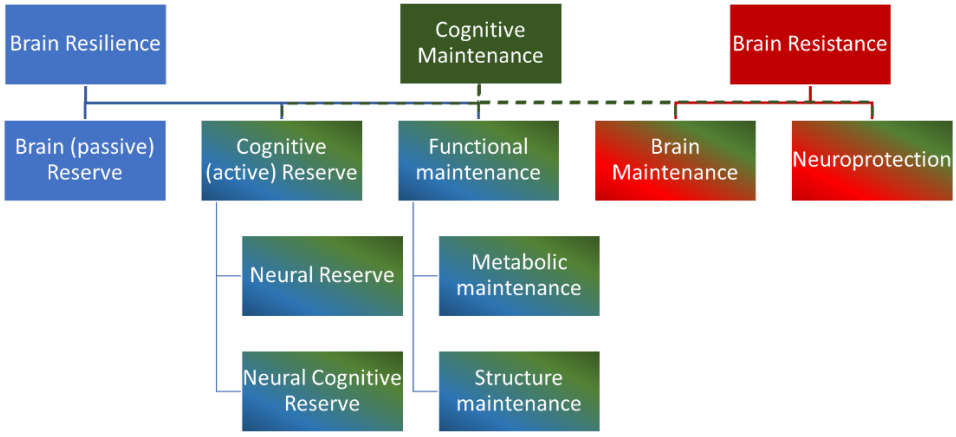
The two main versions of cognitive reserve are neural cognitive reserve and neural compensation reserve. Neural cognitive reserve relates to the efficiency, capacity and flexibility in selection of primary networks responsible for performing a cognitive task.(Barulli & Stern, 2013) Neural compensation reserve relates to the use of secondary networks recruited to perform tasks after failure in the primary networks. If an exposure, education in our case, contributes to neural cognitive reserve then this would be anticipated to slow cognitive decline.(Weiler et al., 2018) The greater efficiency in the primary networks compensates for ageing or pathological change due to greater redundancy in the primary network. If an exposure contributes to neural compensation then it enables the recruitment of secondary networks to compensate for damaged primary networks.(Colangeli et al., 2016; Serra et al., 2017) So, if an exposure contributes to neural compensation reserve then it would be expected that observed cognitive decline may initially be slower but would then accelerate rapidly as the secondary networks are also overcome by the disease process.(Lenehan et al., 2015; Serra et al., 2017)

In Stern and colleagues previous work and under the ISTAART group framework the neural compensation they include in cognitive reserve has significant overlap with the McGill groups definition of neural compensation.(Barulli & Stern, 2013; Cabeza et al., 2018; Stern et al., 2018) However, the McGill group view compensation not on a life-course timescale but as the neural resource which can be activated on task specific timescale.(Cabeza et al., 2019) The ISTAART group, on the other hand, emphasise compensation as a persistent alteration in patterns of activation which would be seen completing the same task many times on differing occasions. This difference in emphasis between a predominantly neuropsychological and clinical or social gerontological perspective, rather than a difference in interpretation of the research literature, appears to be behind the reasons for the differences in classification. As such, I will be using the ISTAART group classification of compensation as a part of cognitive reserve.

There is considerable overlap between the active forms of reserve, neural reserve and neural compensation and maintenance. Reserve reflects differences in the use of available neural resources and maintenance the preservation of those resources. Neither is able to operate without the other. As stated in Cabeza et al. "if education augments reserve by increasing synaptic density , this can attenuate age-related cognitive decline if the new synapses are preserved via maintenance".(Cabeza et al., 2018) In longitudinal analysis without access to detailed measures of brain structure, metabolic function and pathological burden it is not directly

possible to assess maintenance of brain structures. I will therefore be using the term cognitive maintenance as an over-arching category to refer to longitudinal cognitive change, bridging brain maintenance and longitudinal aspects of cognitive reserve. Using the term cognitive maintenance, I will not be referring to any specific neural or disease process but the common end result of these processes which is observable through change in measurements of cognitive function. Essentially this captures all active or dynamic processes and not passive cognitive reserve (figure 1.2).

Figure 1.2 Hierarchy of terminology used to describe subordinate categories underlying observed cognitive function.†



† The colours used indicate which terms are included under which of the 3 main umbrella terms. Those with blue are subsumed under brain resilience, those with red under brain resistance and those with green under cognitive maintenance.

1.3 Research Questions

The aim of this thesis was to use new statistical methods to develop our understanding of whether education and cognitively stimulating activities may improve cognitive maintenance. Due to these exposures occurring at different points in the life course they required different statistical approaches. During the process of answering whether education is associated with changes in cognitive maintenance, additional methodological questions presented themselves as necessary hurdles to be overcome. The first two research questions below therefore closely inform how the third was approached. The fourth question addressing the cognitively stimulating activities develops the major argument of the thesis with respect to how new methodologies can provide new insights into cognitive maintenance by addressing limitations of the existing literature.

The primary and secondary research questions I will address in this thesis are:

1) When analysing variables associated with cognitive maintenance does analysing growth with a sum score or factor score lead to different substantive conclusions?

    a. What is the factor structure of the cognitive tests in the English Longitudinal Study of Ageing?

    b. How does this factor structure compare to the pre-specified cognitive index scores?

2) When conventional tests for longitudinal measurement invariance based on the comparative fit index provide inconclusive results can Bayesian approximate measurement invariance be used as a suitable alternative?

    a. Do the cognitive function latent factors in ELSA show longitudinal measurement invariance?

3) Is education associated with cognitive maintenance and does this association vary by latent class of decline?

4) Does exposure to cognitively stimulating activities in later life reduce risk of dementia or cognitive impairment once time-varying confounding affected by past exposure is accounted for?

    a. Does exposure to cognitively stimulating activities in later life improve cognitive function once time-varying confounding affected by past exposure is accounted for?

The fourth research question directly addresses a fundamental challenge in population studies of cognitive maintenance, time-varying confounding affected by past exposure. In this analysis there is an emphasis on the prevention of cognitive impairment and dementia as one of key outcomes of improved maintenance. The secondary outcome asks the same question but using a continuous, rather than dichotomous, outcome to provide a supplementary view on whether cognitive maintenance is affected more broadly than the prevention of impairment alone.

2. Education, cognitively stimulating activities and cognitive function

This chapter introduces the literature on education, other cognitively stimulating activities and cognitive maintenance. It also reviews the existing literature using the methodologies of growth mixture modelling and marginal structural models which are important for relaxing particular assumptions in standard regression analyses. This chapter focusses on practical and theoretical aspects of their use in the existing literature, the statistical motivation for the methods will be described in chapter 3. In this chapter section 2.1.1 and 2.1.2 will review the literature on education and cognitive function and dementia risk. Section 2.1.3 focussed on the issue of population heterogeneity and reviews the literature on the association between education and cognitive function which has used growth mixtures models to account for this. 2.1.4 gives a brief review of literature on the relationships between dementia pathology, education and cognitive function. Section 2.2 reviews literature on cognitively stimulating activities. This section starts with an overview in 2.2.1 before moving on to discuss the literature on general cognitive enrichment in 2.2.2 and specific cognitive activities in 2.2.3. The literature reviewed in section 2.1.1 and 2.1.2 dates from the beginning of the thesis in 2014-15 and reflects the state of the evidence at that time. The reviews in sections 2.1.3 onwards are focussed more narrowly and were conducted up to 2018 and revised in 2019.

2.1 Education and Cognition

Dementia and mild cognitive impairment are highly prevalent in older adults globally.(Alexander et al., 2015; Prince et al., 2013) Both dementia and mild cognitive impairment have a high cost in terms of individual suffering, caregiver burden, healthcare and long-term care.(Alzheimer's Association, 2014; Emerson et al., 2017; P.-J. Lin & Neumann, 2013; P. J. Lin, Zhong, Fillit, Chen, & Neumann, 2016; Paradise et al., 2015; Prince et al., 2013) Given the scale of the problem it is important that modifiable factors are identified which can improve cognitive maintenance. Improving cognitive maintenance may then either reduce the probability of developing impairment or slow the progression of impairment. Higher levels of education is one modifiable factor which has attracted considerable interest for both reducing dementia/cognitive impairment risk and better maintenance of cognitive function in healthy old age.(Livingston et al., 2017) This discussion will attempt to discuss dementia and cognitive ageing separately, however there will inevitably be some overlap.

2.1.1 Systematic Reviews of Education and Dementia Risk

Several systematic reviews have included much of the literature on the association between education and dementia risk.(Beydoun et al., 2014; X. Meng & D'Arcy, 2012; Sharp & Gatz, 2011) I will focus on the reviews which were most recently published in 2014-2015. This includes two focussed reviews and one including education amongst a range of other modifiable risk factors.

Meng and D'Arcy in their 2012 paper synthesise a large number of epidemiological and non-epidemiological studies into a comprehensive review of education as related to dementia risk.(X. Meng & D'Arcy, 2012) This review views education as a proxy for cognitive reserve as described by Stern and described in section 1.2.(Stern, 2012) They view higher levels of education as resulting in changes in brain structure and processing. These delay the onset of clinically observable impairment but are not related to progression of the underlying pathology of dementia. This means that their conclusions are specifically with respect to the clinical syndrome of dementia and not pathological change.

A dichotomous exposure variable of high or low education was created for the meta-analysis. This was done on a study by study basis without a fixed set of reference categories for the exact amount of education received or attained. This meant that a wide range of studies could be included. The trade-off is that the low and high education groups used are quite heterogenous. It implies that a consistent dose response relationship is assumed across studies with different ordinal levels of exposure to education. They conclude from their fixed effects meta-regression of cohort studies that those with 'low' education have a pooled odds ratio (OR) of 1.88 (1.51-2.34) of incident dementia (all types) compared with those with high education. In sub-analyses the OR is similar for Alzheimer's disease (AD) at 1.82 (1.36-2.44) but considerably higher for vascular dementia at 2.75 (2.20-3.45). They conclude from their meta-analysis that higher education reduces both incidence and prevalence of dementia.

In contrast to the finding that high education is protective in terms of incidence of dementia, in their qualitative review they find that 70% of studies report more rapid cognitive decline amongst more highly educated dementia sufferers. They find that 5/8 studies report earlier age of diagnosis in more highly educated individuals, with no studies finding later age of onset in more educated participants. On the other hand, they find that cognitive scores at initial presentation are higher in more highly educated participants and that 10/14 pathological studies find higher levels of pathology in the more highly educated. They conclude overall that their qualitative analysis supports the idea that education provides cognitive reserve, as described in section 1.2, and moderates the clinical expression of disease.

At face value these conclusions seem almost contradictory. The finding of more rapid decline and higher pathological loads in those with higher education at equivalent levels of function to those in the low education group suggests they are presenting at a later stage in their illness as defined pathologically. However, the authors also find that more highly educated individuals have earlier age of onset. Therefore, more highly educated individuals must either experience a more rapid progression of pathology or develop the pathology substantially earlier than those with less education. Moreover, the overall incidence and prevalence are reduced by higher education. Taken together the findings of the Meng and D'Arcy systematic review would suggest that higher education reduces one's risk of dementia but makes it earlier or more severe if suffered. This may be partially explained by a case detection bias. Those with greater education but also greater genetic risk may be identified as cases at earlier ages with more aggressive disease.(S. J. van der Lee et al., 2018) Those with greater education but lower

genetic risk may not be reaching the levels of cognitive impairment required to be detected. Nevertheless, Meng and D'Arcy's finding of earlier onset and more severe disease being associated with greater education does not correspond well with conventional theories of cognitive reserve.

However, those findings are what one would expect if individuals with higher education had higher passive cognitive reserve, and therefore presented with the clinical syndrome of dementia at a later stage in pathology, but were more likely to access services for diagnosis and disclose problems in diagnostic interviews compared to those with lower education. This is particularly likely given that the samples in the studies with positive findings were drawn from patients at dementia clinics, of which several were in countries without comprehensive medical provision, rather than from general population samples.(Bowler, Munoz, Merskey, & Hachinski, 1998; Brodaty et al., 2014; Kokmen, Beard, Brien, & Roccu, 1996) As dementia tends to progress more rapidly with earlier onset, it may also be the case that education is protective against later onset dementia but not early onset dementia.(Jacobs et al., 1994) These explanations could all be potential explanations explain the surprising combination of earlier onset but lower incidence and prevalence.

In their systematic review published 1 year earlier, Sharp and Gatz focus only on studies of incidence and prevalence.(Sharp & Gatz, 2011) They report that the majority of studies they included report lower prevalence and lower incidence amongst more highly educated participants. Sharp and Gatz felt that the heterogeneity between studies was too high to conduct a meta-analysis. This contrasts with Meng and D'Arcy who did perform a meta-analysis but relied upon a dichotomous low-high education variable to do so. They identify that studies from more economically developed regions were considerably more likely to find an association between education and dementia than those from developing regions.

However, a considerable number of prevalence and incidence studies either did not adjust for any other variables or only for age and gender. This raises the distinct possibility that the observed associations might be affected by unadjusted confounding and provides little information about potential causal mechanisms. Indeed, the authors recognise this and propose a lifespan developmental model in which education is both the result and cause of multiple factors across the lifecourse which may ultimately effect dementia risk. They propose that cognitive reserve should be considered in relationship to the ability meet the demands of one's environment, which may lead to differing effects of education across populations, cohorts, ethnicity and gender. Whilst this is an attractive hypothesis, few studies have the necessary methodological complexity to begin to disentangle these effects.

Whereas the prior two reviews focussed specifically on education, a systematic review by Beydoun and colleagues examined education as one of several modifiable risk factors for dementia or cognitive impairment.(Beydoun et al., 2014) They identified 25 cross sectional studies of dementia prevalence which met their inclusion criteria, of which only 1 found no association between education and either prevalent vascular dementia (VaD) or prevalent AD.

Although these cross-sectional studies were included, their analysis primarily focussed on cohort studies which examined the association between education and incident dementia or cognitive impairment. They included 27 cohort studies. 18 (66.7%) of these studies did find an association between greater education and lower risk of dementia or cognitive impairment in the population. Only 4 of the studies they included found no association with any dementia outcome. The others found the association only in women (2 studies), in APOE4 negative individuals only (2 studies), with VaD but not AD (1 study), with baseline cognitive function but not decline (1 study), or that the association was not significant once IQ was accounted for (1 study). There was significant heterogeneity in the effect size estimates. Only 4 were ultimately included in their meta-analysis, primarily because of non-comparable data on covariates. In their pooled analysis of those studies the risk ratio for incident AD for high (defined as ≥8 years) versus low (defined as <8 years) education was 1.99 with a 95% confidence interval of 1.3 to 3.04.

Overall, there is a strong consensus amongst the systematic reviews to date that higher levels of education is associated with reduced risk of being diagnosed with dementia. It is worth noting that these studies focussed on the diagnosis as a clinical syndrome. As it is possible to have sub-clinical levels of impairment despite the presence of considerable quantities of dementia pathology, these findings do not mean that education is associated with lower risk of dementia pathology. Additional limitations common to these reviews include frequently inadequate adjustment for confounding factors (such as the time of diagnosis and the availability and accessibility of dementia diagnoses services) and the potential for ascertainment or measurement bias. This means that there remains considerably uncertainty about the mechanism(s) responsible for the observation of this association. In order to consider this more fully it will be helpful to turn to look at those reviews studying the association between education and normal cognitive ageing.

In the period since these reviews until the point of writing prior to the deadline for this thesis, perhaps the most influential publication on dementia risk was the report from the Lancet commission on dementia prevention, intervention and care.(Livingston et al., 2017) Their estimate of the effect of education is smaller than the meta-analyses above with an estimate of a relative risk of 1.59 (95% CI 1.26–2.01) for having no time in secondary education. Due to the global prevalence of low education, they find that it has the highest population attributable fraction of all the modifiable risk factors they review. A point of difference with the other reviews above is that they remain considerably less certain about the effects of education above this basic level.

2.1.2 Systematic Reviews of Education and Cognitive Ageing.

In Beydoun and colleagues' systematic review discussed above, the effect of education on cognitive function is also examined in a subsection of studies. Of the 11 cross sectional studies which met their inclusion criteria higher education was positively associated with better cognitive function in all of them. Of 15 longitudinal studies, 4 did not find a significant association between

education and cognitive decline. Due to the broad scope of the review, they provide little detail about the nature or strength of the associations found.

The overall conclusion that education is related to less cognitive decline was also the conclusion of Valenzeula and Sachdev's systematic review from 2006.(Valenzuela & Sachdev, 2006) They performed a non-parametric meta-analysis of 13 longitudinal studies, which reports a large effect of education on slowing cognitive decline.(Valenzuela & Sachdev, 2006) However, their approach to meta-analysis rests on strong assumptions about the equivalence of effect sizes and the fact that all studies are equally weighted regardless of size or quality.

It comes as something of a surprise then that the systematic review by Plassman and colleagues from 2009 found that the evidence regarding cognitive decline and ageing to be 'inconsistent'.(Plassman, Williams Jr, Burke, Holsinger, & Benjamin, 2010) This difference will partially result from different availability of papers at the time of analysis. However, Plassman and colleagues also employed the Agency for Healthcare Research and Quality criteria to assess methodological quality which should have led to the inclusion only of more methodologically rigorous research. Moreover, as table 2.1 demonstrates, only a small proportion of the studies reviewed are included in both reviews despite only 3 of the papers reviewed in Beydoun et al. being published after the search period of Plassman et al. One potential explanation for the differing findings is that no association is found more often in studies with 3 or more measurement occasions and thus superior modelling of change over time.(Wilson, Hebert, Scherr, Barnes, & Leon, 2009) It is noticeable from table 2.1 that Plassman et al. included several more papers with 3 measurement occasions.

Another proposed mechanism is that studies which have identified a different rate of decline due to education have used cognitive tests which are prone to ceiling effects and would therefore mask decline in more able individuals.(Karlamangla et al., 2009) Indeed, the higher quality studies using better instruments do seem less likely to report an effect of education on the rate of change. Perhaps the differing conclusions say more about the process of systematic review than the effect of education on rates of cognitive decline. Nevertheless, I am more inclined to agree with the conclusion of the older systematic review that the evidence does indeed appear inconsistent with regards to the effect of education on cognitive change over time in older adults.

Table 2.1 Papers included in the systematic reviews of cognitive ageing and education by Beydoun et al. and Plassman et al.

Study	Country (project)	Age at recruitment	Sample size	Years	Sampling	Education	Analysis Type	Covariates	Main Result
Studies Shared Between Beydoun et al. and Plassman et al.									
(Alvarado, Zunzunegui, Del Ser, & Béland, 2002)	Spain	65+	557	ON REQUEST					
(Kalmijn et al., 1997)	Netherlands (Zutphen Elderly Study)	Not spec.	390	1990,1993	Not described	≤6 years, >6 years	Logistic Regression	ApoE4 status, age, baseline function	Higher odds of decline for low educated ε4 negative individuals
(S. Lee, Kawachi, Berkman, & Grodstein, 2003)	USA (Nurses Health Study)	70-79	15594	1995-2000, 2001-2002	Not described here.	Postgraduate Degree, Degree, Nursing Diploma	Logistic Regression, Linear Regression	Husbands education, Paternal occupation, area income, multiple others	Higher odds of low baseline score and change score for Diploma nurses.
(Sunmin Lee, Buring, Cook, & Grodstein, 2006)	USA (Women's Health Study)	≥66	5573	1998, 2000	Follow-on study from RCT of aspirin and vitamin E	6 categories from Vocational nurse to PhD	Linear Regression, Logistic regression	Focus on income, also traditional	Education predicts baseline function and rate of decline,

								cardiovascular risk factors	income strong predictor of baseline function.
(Wilson et al., 2009)*†	USA (Chicago Health and Ageing Project)	≥65	6533	1993-2007	Stratified Random Sample	Years of education (continuous)	Mixed-effects model	Race, occupation, chronic disease	Higher levels of education not associated with rate of cognitive decline.
(Yaffe et al., 2009) †‡	USA (Health ABC)	70-79	2509	1997-2004/5	Simple random sample from selected ZIP codes	Literacy <9th grade or ≥ 9th grade, Education > highschool or ≥ highschool	Random effects to identify individual rates of change. Multinomial logistic regression to predict decline category.	Demographics, self-rated health and cardiovascular risk factors and diseases, various biomarkers	Education associated with lower odds of being in declining groups, strong effect for literacy.
Studies in Beydoun et al. only									
Study	Country (project)	Age at recruitment	Sample size	Years	Sampling	Education	Analysis Type	Covariates	Main Result
(Aevarsson & Skoog, 2000)	Sweden	85-88	494	1986-1990	Whole population of	≤6 years vs >6 years.	Logistic regression to	None included.	Higher baseline MMSE score in

					Göteborg aged 85		predict dementia.		more educated, slower decline in more highly educated women.
(Castro-Costa et al., 2011) †	Brazil (Bambui Study)	≥60	1606	1997-2007	Whole population of Bambui aged ≥60	≤3, 4-7 or ≥8 years of school	Mixed effects modelling of MMSE results	Age and gender only. Interactions estimated separately.	Women and those with higher level of education declined faster.
(H Christensen et al., 1997) ‡	Australia (PATH Through Life)	≥70	540	1990-1994	Simple random sample of Canberra and Queanbeyan residents	≤9 years, 10-13 years or ≥14 years schooling.	Linear regression on wave 2 score for 8 separate cognitive tests.	Wave 1 score, age, gender, activity level, disability, health and change in health.	Education protective against decline in crystallised intelligence, not other cognitive domains.
(Helen Christensen et al., 2009)* ‡	Australia (PATH through life)	60-64	416	2001/2002 to 2005/2006	Simple random sample of Canberra and Queanbeyan residents	From 4-12, 13, 14-15 and ≥16 years.	Linear regression on the difference between time 1 and time 2 scores.	Age, gender and initial cognitive test scores.	Education was not associated with rates of cognitive decline (nor were white matter hyperintensities, atrophy or

									intracranial volume).
(Evans et al., 1993) ‡	USA	≥65	2273	1982/1983 to 1986/1986	Individuals in age range from East Boston, Massachusetts	Unclear, represented both as ordinal and continuous variable.	Linear Regression of normalised rank of change in cognitive scores	Occupation, income, birthplace, language, age, gender and cardiovascular diseases.	Lower levels of education associated with faster rates of cognitive decline.
(Lykestos, Chen, & Anthony, 1999)	USA (Baltimore Epidemiologic catchment area study)	≥18	1488	1981 – 1993/1996	Random sample within area	0-8, 9-11, 12 (GED), 13-15 or ≥16 years.	Change in MMSE between waves 2&3 adjusted for wave 1 performance	Age, gender, ethnicity and baseline score.	Fewer years of education (and non-white ethnicity) were associated with greater decline.
(Marengoni, Fratiglioni, Bandinelli, & Ferrucci, 2011)	Italy (inCHIANTI)	≥60		1998/2000 – 2001/2003	Total population within defined area		Logistic Regression and Cox Proportional Hazard of cognitive impairment not dementia	Demographics, job stress and physical demand, cardiovascular diseases, diabetes, apolipoprotein E (APOE)genotype,	Lower level of education associated with higher odds of CIND.

							(defined as ≥1 S.D decline in MMSE)	smoking, alcohol consumption, depressive symptoms, and C-reactive protein	
(Schmand et al., 1997) †	The Netherlands (Longitudinal Ageing Study Amsterdam and Amsterdam Study of the Elderly)	65-84 & 65-85	1774 & 1950	1992/1993 & 1990 - 1994	Clustered random sample	9 point ordinal scale analysed in separately in lower and higher strata & 6 point scale	Two way ANOVA & ANOVA with repeated measures	Age and Gender	Fewer years of education associated with faster and earlier decline in cognitive function.
(Seeman et al., 2005) †	USA (MacArthur Studies of Successful Ageing)	70-79	895	1988/1989 - 1995	high functioning subsample of community cohorts	0-8 years, 9-11 years, 12 years (completed high school) and ≥13 years	Summed score from multiple cognitive tests or <7 on SPMSQ by Generalised estimating equations.	Age, gender, physical activity, smoking diabetes and stroke (selected on significance from wider range)	More years of education associated with better cognitive performance, but APOE allele associated with poorer cognition

									in those with more education.
(White et al., 1994) ‡	USA (Established Populations for Epidemiologic Study of the Elderly)	≥65	10294 at baseline, total used in analysis not specified	1981/1983 – 1987/1989	full population of target areas and 1 stratified sample	Education dichotomised into ≥8 or ≥9 years of education	Multiple Logistic Regression for odds of incident cognitive impairment	Age, Gender, Occupation, stroke and baseline score	Lower level of education associated with incident cognitive impairment.
(van Hooren et al., 2007)	The Netherlands (Maastricht Ageing Study)	64-81	578	1992-2004	Stratified random sample (older subsample used for analysis)	Ordinal scale of primary education, lower secondary education and higher secondary or greater.	Multivariate ANOVA for standardised scores in multiple cognitive domains.	Age and gender	Education associated with higher cognitive function in multiple domains.
(Zahodne et al., 2011b)	Australia (Victoria Longitudinal Study)	55-94	1023	1986/87 - 2003	Not described nor referenced description	Education considered primarily as continuous variable of	Linear Growth Curve Modelling of multiple cognitive domains	Age and gender	Education related to baseline cognitive performance but not change over time

						years of education			
Studies in Plassman et al. only									
Study	Country	Age at recruitment	Sample size	Years	Sampling	Education	Analysis Type	Covariates	Main Result
(Helen Christensen et al., 2008)*	Australia (PATH)	60-64	2021	2001-2002, 2005-2006	Stratified random sample	Years of education grouped as 0-12, 13, 14-15 and ≥16.	Multivariate ANOVA	Genotype, head injury, cardiovascular risk, intelligence.	No effect of education or intelligence, 2/6 domains interaction between education and APOE genotype.
(Karlamangla et al., 2009)* †	USA (AHEAD)	≥69	6476	1993-2002	Multistage Probability Sampling	Years of education grouped as <8, 8-11, 12-14 and >14.	Mixed Effects	Demographics and socioeconomic variables	Education not associated with decline but income is.
(Koster et al., 2005) ‡	USA (Health ABC)	70-79	2574	1997, 2001	Simple random sample from selected ZIP codes	Years of education grouped as <12, 12 and >12.	Logistic Regression, odds of cognitive decline	Demographic and socioeconomic variables, cardiovascular	Education and income were associated with lower odds of cognitive decline.

								disease and risk, biomarkers.	
(Manly, Schupf, Tang, & Stern, 2005) *†	USA (WHICAP)	≥67	1002	1997-2004	Stratified Random Sample	High (>12 years) or low (<12 years)	Generalised Estimating Equations	Age, gender, literacy, ethnicity.	Education itself is not associated with slower decline but literacy is.
(Shadlen et al., 2005) *†	USA (ACT Study)	≥65	2140	1994-2000	Random sample from a group health co-operative	Years of education (continuous)	Generalised Estimating Equations	APOE genotype, demographics, diabetes, depression, cardiovascular disease.	Educational alone no effect, but significantly less decline in APOE ε4 homozygotes with high education than low education.
(Tervo et al., 2004)	Finland (Kuopio)	60-76	747	1997/1998, 2000-2002	Random sample from Kuopio	Years of education (continuous)	Logistic Regression	Demographics, cardiovascular disease and risk factors	Odds of incident mild cognitive impairment reduced with more years of education.

(Tyas et al., 2007) †	USA (Nun Study)	≥75	470	1991-2002	All members of Sisters of Notre Dame born pre-1917	≤ High school, undergraduate degree, postgraduate degree	Polytomous Logistic Regression	Age, education, APOE status	More education reduces risk of mild cognitive impairment but not progression to dementia.
(Winnock et al., 2002)	France (PAQUID)	≥65	600	1988-1998	Random sample from Southwest France	Completion vs. non-completion of primary school	Random effects linear regression.	Age, gender, time in study, APOE status	No difference in rate of change over time by education.

* No or minimal independent effect of education on rate of change.

† More than 2 measurement occasions used in analysis (not including adjustment for baseline performance).

‡ Duplicate inclusion of the study sample in another referenced analysis.

2.1.3 Population heterogeneity and growth mixture models

There is a general consensus that education reduces an individual's risk of being diagnosed with dementia, as shown in the reviews above.(Beydoun et al., 2014; Livingston et al., 2017; X. Meng & D'Arcy, 2012; Sharp & Gatz, 2011) However, uncertainty remains regarding how education has this effect. Much of this relates to differences in how dementia and cognitive ageing are defined differently in neuropsychological and pathological terms. In other words, should you consider someone as having dementia if they have a certain set of pathological brain changes or when they reach a threshold level in cognitive function?

The distinction between Alzheimer's disease pathology (ADP) and Alzheimer's disease (AD) can now be drawn in a setting where pathophysiological markers are available.(Dubois et al., 2014) The discrepancy between pathology and function creates difficulties in research where these biomarkers are not available. When comparing an individual with poorer cognitive function to another with better cognitive function it becomes complex to determine if they have more pathology, less cognitive reserve which acts independently of pathology or less cognitive reserve which moderates the effect of the pathology. Moreover, even if biomarkers are available there remains a considerable degree of uncertainty with regards to the natural history of the progression of ADP in the early stages of disease.(Dubois et al., 2016) Cohort studies or analysis of clinical and registry data may provide important insights on dementia risk, normal cognitive ageing in healthy individuals or those with diagnosed dementia or cognitive impairment but lack the pathophysiological markers. This is the case in the English Longitudinal Study of Ageing which is the source of data used in this thesis.

In clinical or epidemiology studies those with low baseline cognitive function and a degree of age-related cognitive decline may be diagnosed as having dementia, whilst in those with high baseline function and ADP may go undiagnosed. In studies on dementia incidence or prevalence, this will lead to upwardly biased risk in the less educated. On the other hand, if those same individuals are then excluded from studies on normal cognitive ageing, this will downwardly bias the apparent effect of higher education. Whilst, the scientific consensus is that AD is a distinct pathological process which is not on a continuum with normal ageing, the potential for significant biases in diagnosis with respect to underlying pathology is problematic for ageing research.(Nelson et al., 2012, 2011) This is further complicated by the presence of other diseases causing cognitive impairment, especially small vessel cerebrovascular disease (SVD) which is both a cause of cognitive impairment alone and highly prevalent in subjects both with and without AD.(Slavin, McManus, & Stott, 2012)

In the general population, a high proportion of older adults will have ADP or SVD even if they are not overtly cognitively impaired.(Braak & Del Tredici, 2015; Dubois et al., 2016; Riley et al., 2011) It is beyond the scope of this thesis to provide a detailed discussion of many nuanced ways old age may be defined or who may be classified as an older adult. For the purposes of this thesis the terms old age and older adult will be used to refer to adults aged 60 or over. Whilst this is in many ways chronologically very young to qualify as 'old age', around the age of

60 there is a point of inflection after which the prevalence of ADP and other dementia pathologies start to rise exponentially.(Nelson et al., 2011)

A population sample of older adults will contain at least two sub-populations, those with ADP pathology and those without. In samples or databases without pathophysiological markers of ADP it is not directly possible to observe who is in which population. The presence of distinct but not directly observable (latent) sub-populations is known as population heterogeneity.(Muthen, 2004) This is important for population studies to take consider because there is evidence that the effect of education on cognitive ageing may vary depending on ADP status.(Colangeli et al., 2016; Serra et al., 2017) If education provides neural compensation, then those in the latent sub-population with ADP may decline more rapidly with higher levels of education, but this effect would not be seen in healthy old age.(Y.-N. Song et al., 2018) Probably the most common method used to account for population heterogeneity in longitudinal studies is growth mixture modelling.(Muthen, 2004) This provides a flexible modelling framework which can be specified to test whether education contributes to brain resistance or resilience through brain reserve, neural reserve or neural compensation. Figure 2.1. uses structural equation models to demonstrate the different potential forms of reserve by adding additional detail to that shown figure 1.1. Figure 2.1 includes 3 graphics showing different plausible sets of causal assumptions linking education, latent disease status, cognitive reserve and latent cognitive function and observed cognitive tests.

Figure 2.2 graphically demonstrates the predictions of different theories of cognitive reserve in terms of predicted change over time in observed cognitive test scores. Graphs 1-3 demonstrate the hypothetical relationships between education and cognitive maintenance under the assumptions of brain (passive), neural compensation and neural cognitive reserve in the single class case. For simplicity only two categories of education, high and low, are present. Graphs 4-12 demonstrate the potential combinations of these 3 theories in a two-class setting. The stable class represents probable healthy cognitive ageing and the declining class probable cases of dementia. There is one graph for each combination of reserve in either class. All of these graphs demonstrate plausible hypotheses which could be identified in a growth mixture model where the effect of education is moderated by class of latent trajectory.

Figure 2.1 Structural equation models showing hypothetical casual pathways through which education could affect cognition with respect to theories of cognitive reserve.

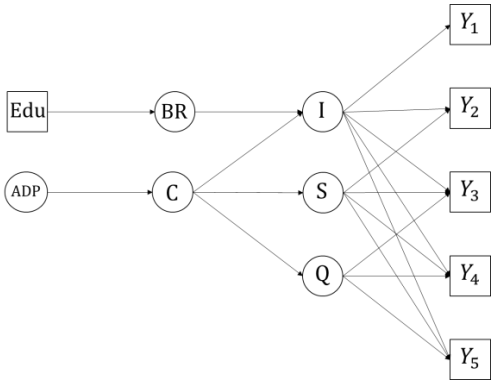
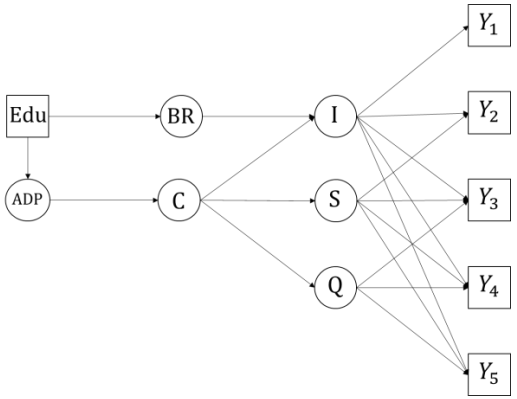
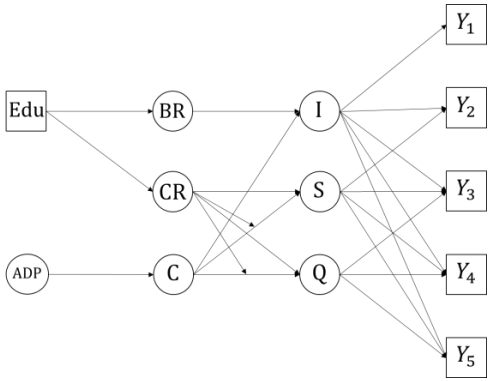
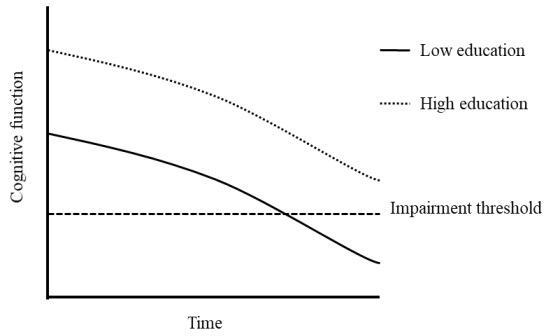
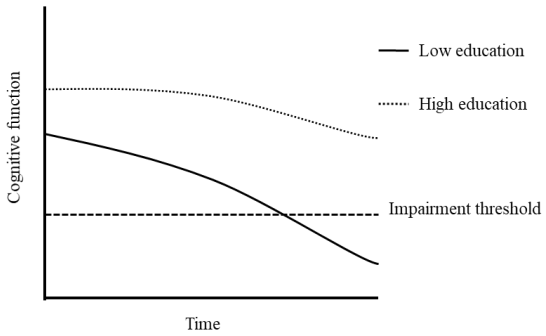
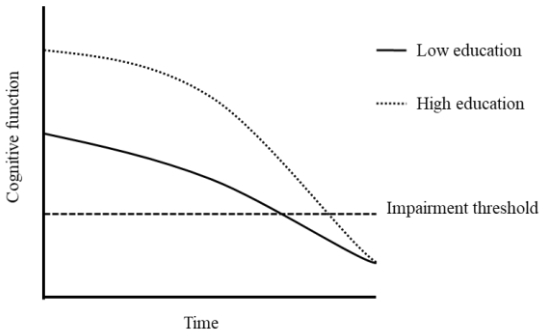
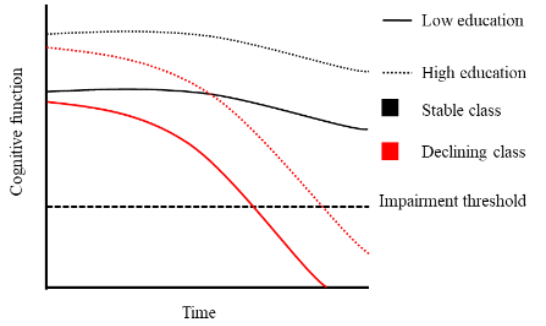
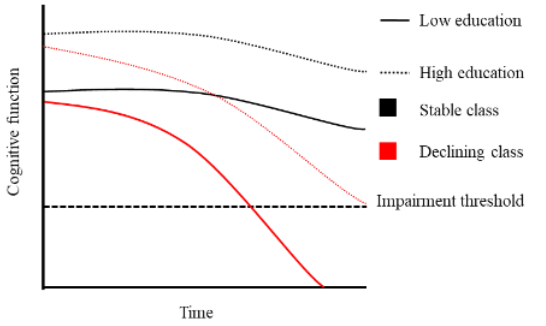
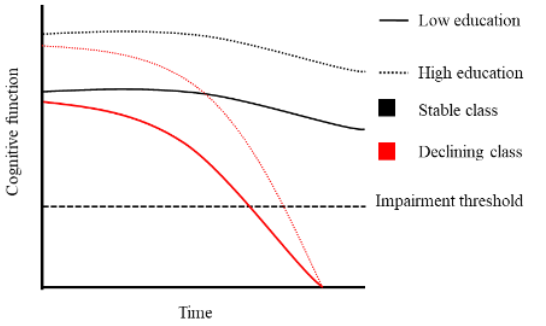
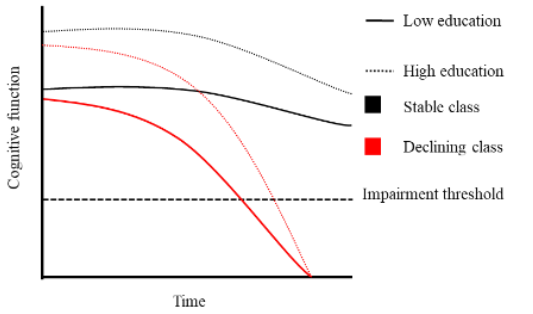
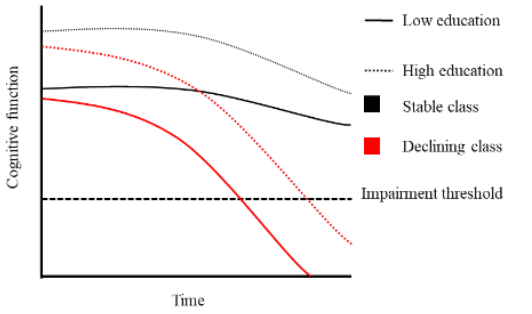
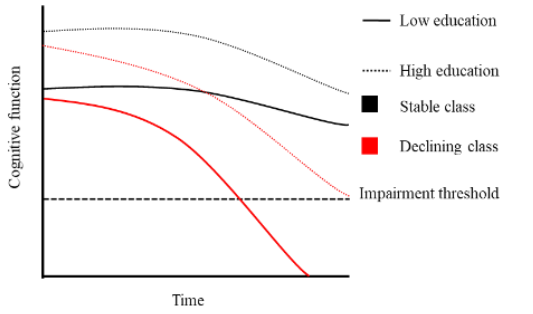
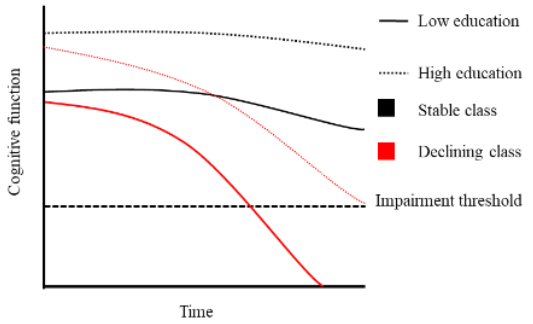
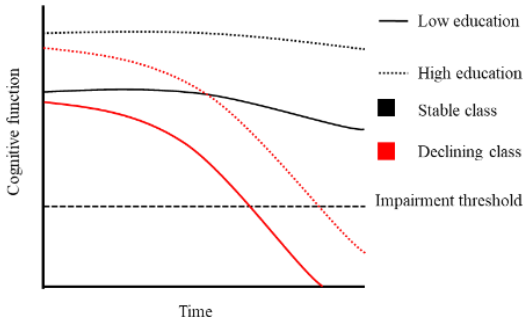
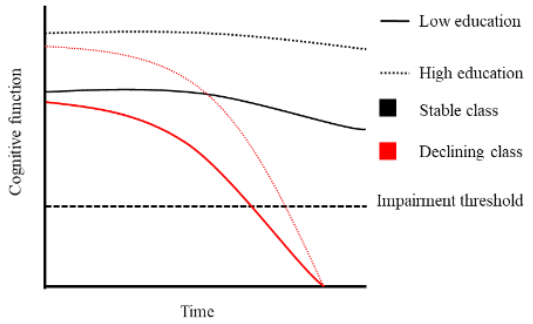
1. Education contributes to brain resilience by affecting brain reserve only, independent of disease status.	2. Education contributes to brain resilience via brain reserve and resistance by reducing ADP build-up.	3. Education contributes to brain resilience by affecting both brain reserve and cognitive reserve.
		
Squares represent observed variables and circles unobserved variables. Variables are: Y_1-Y_5 - a measure of cognitive function on 5 occasions, I - latent intercept (baseline cognitive function), S - latent slope (rate of change), Q - latent quadratic growth term (non-linear rate of change), C - latent class of cognitive trajectory, BR - brain reserve, Edu – educational attainment, ADP – Alzheimer's disease pathology, CR- cognitive reserve.		

Figure 2.2 Trajectories of cognitive decline due to age or disease under different theoretical assumptions (adapted from Lenehan et al. 2015.)

<table>
<tr>
<td>1. Brain (passive) reserve single class
 
</td>
<td>2. Neural Cognitive reserve single class
 
</td>
<td>3. Neural compensation reserve single class
 
</td>
</tr>
<tr>
<td>4. Brain (passive) reserve in both of two classes
 
</td>
<td>5. Brain (passive) reserve in a stable class and neural cognitive reserve in a declining class
 
</td>
<td>6. Brain (passive) reserve in a stable class and neural compensation reserve in a declining class
 
</td>
</tr>
</table>

7. Neural compensation in both of two classes 	8. Neural compensation in a stable class and brain reserve in a declining class 	9. Neural compensation in a stable class and neural cognitive reserve in a declining class 
10. Neural cognitive reserve in both of two classes 	11. Neural cognitive reserve in a stable class and brain reserve in a declining class 	12. Neural cognitive reserve in a stable class and neural compensation in a declining class 

Further detail on the statistical methodology and motivation for growth mixture modelling is provided in the methods section. In 2014 when this project was started there were few studies which had taken this approach to cognitive ageing. Several years on, there is now a considerably larger body of literature which has used a latent class growth model or growth mixture modelling to study the association between education and cognitive ageing.(Baker et al., 2017; Ding et al., 2019; Downer, Chen, Raji, & Markides, 2017; Hayden et al., 2011; Hochstetler et al., 2016; Seonjoo Lee et al., 2018; Marioni et al., 2014; Min, 2018; Muniz-Terrera, Brayne, & Matthews, 2010; Olaya, Bobak, Haro, & Demakakos, 2017; Pietrzak et al., 2014; Royall, Palmer, Chiodo, & Polk, 2014; Small & Bäckman, 2007; Zahodne, Wall, et al., 2015) All to date have used frequentist estimation of single indicator growth and most are descriptive of the classes identified.

The first paper examining cognition using a GMM was written in 2005 primarily as a methodological paper in which cognition is a motivating example.(Proust & Jacqmin-Gadda, 2005) They analysed Mini-Mental State Examination (MMSE) scores from the French PAQUID study of 1392 community dwelling adults aged ≥65 from two French districts who were followed up on 5 occasions from 1988 to 1998. They found two latent classes of cognitive change. The first class was a stable class and the second a class with quadratic decline over time. They found that participants diagnosed with dementia but classified in the stable class had lower levels of education. This is suggestive of those individuals having low levels of pathology but poor baseline functioning.

A research team comprising many of the same investigators has more recently published a further analysis with an additional 10 years of follow-up of the PAQUID cohort.(Marioni et al., 2014) They also extended their previous work by incorporating joint modelling of mortality risk. In this study they were able to subdivide their previously identified classes into stable groups with low and high baseline and declining groups with delayed or immediate decline. Both decliner groups had higher mortality rates (the immediate higher than the delayed) than the two stable groups (the high-performance stable group having the best survival overall). With regards to educational attainment they found that higher levels of education were a strong predictor of not belonging to the immediate or slow decline groups.

The second paper to be published on cognition using GMM was Small and Backman's 2007 paper.(Small & Bäckman, 2007) As with the paper by Proust and Jacqmin-Gadda the primary outcome modelled is MMSE score. They used data from the Kungsholmen project in Sweden on 457 adults aged ≥75 at recruitment who were either dementia free at the end of the three year follow-up period or had incident dementia at the last measurement occasion. They assumed a two-class model a priori. Their analysis focussed on whether decline was linear or quadratic and the relationship between class as identified by the GMM and diagnosis (assumed to be accurate). The best fitting model they identified was one in which both classes had quadratic decline with the poor performance group having both lower intercepts and more rapidly accelerating decline over time. They found that their 'false positive' group (those classified in the poor cognitive trajectory who did not receive a dementia diagnosis) had fewer

years of education than the ‘false negative group’ (those who received a dementia diagnosis but were classified as higher functioning).

Muniz-Terrera and colleagues used GMM to analyse data from the Cambridge City over 75 Cohort Study (CC75C).(Muniz-Terrera et al., 2010) They also used MMSE score as their measure of cognitive performance. They identified three groups, a class with high baseline performance with slight linear decline, a group with poor baseline performance and accelerating decline and a small group with poor baseline performance with rapid linear decline. In contrast to the previous two studies, they identified similar levels of education in each class. Furthermore, education only affected cognitive performance in the higher performing group where having had a shorter education was associated with lower baseline performance and faster average decline. Their class of poor performers with accelerating decline was unusually large (54% of the sample or n=1102 individuals). Dropout and death were modelling through the inclusion of a logistic regression model dependent on observed variables which may account for some of the differences with the earlier two studies.

Xie, Mayo and Koski studied trajectories of change in MMSE score in 187 outpatients identified as having mild cognitive impairment over a 3 and a half year period.(Xie, Mayo, & Koski, 2011) They identified 5 separate classes of which 4 were largely ordinal separations but one small class of 4.2% of participants demonstrated considerably more rapid decline from a moderate baseline. They found that, although poorer performing groups were progressively less likely to have >12 years of education, education did not significantly affect the probability of group assignment. As this study is based upon a small opportunistic clinical sample, its results do lack generalisability to the broader population of older adults.

A similar difficulty with generalisation is present in Hayden et al.’s GMM analysis of cognitive change in the Religious Orders Study.(Hayden et al., 2011) However, it does have the significant advantage of a comprehensive battery of cognitive tests to draw upon as well as a subsample who underwent post-mortem examination for dementia pathology. In this study three classes were identified with progressively more rapid slopes of decline over time. There was no particular difference in educational level between the classes, but the Religious Order Study participants are a highly educated cohort. Analysis of the subsample who underwent autopsy found a medium sized effect for amyloid load, and large effect for tangle density, on class membership probabilities. This confirmed that the classification of decline trajectories did correspond to underlying pathological burden of disease.

Leoutsakos and colleagues put GMM to an interesting use in the analysis of a trial on the effect of non-steroidal anti-inflammatory drugs and cognition.(Leoutsakos, Muthen, Breitner, & Lyketsos, 2012) Whilst their concern was primarily whether the treatment effect differed within class rather than social determinants of cognitive ageing, it is interesting to note that as with Hayden *et al.* and Terrera *et al.* they identified three classes separated by increasing speed of decline.

Royal and colleagues undertook a GMM of cognitive ageing with the specific intention of identifying a latent class which represents an ageing specific (disease free) phenotype.(Royall

et al., 2014) They analysed data from a sample of 547 retirees living in a single American retirement community evaluated over a 3 year period. The data on cognition includes a range of cognitive tests which the researchers had previously identified as having minimal retest effects. They analysed each test separately and identified 2 or 3 latent classes depending on the test. For most tests the classes primarily represented an ordinal difference in initial performance with modest differences in slope. However, in keeping with previous findings for the MMSE there were 3 distinct classes with an 11% subgroup who had considerably more rapid decline than the high-performance group.

Pietrzak and colleagues examined data from 333 ≥60 year old adults who were cognitively normal at baseline over a 54 month period.(Pietrzak et al., 2014) They identified 3 classes which were high stable performance, low to average performance with slight decline and poor baseline with rapid decline classes. In their bivariate analysis low levels of education was associated with higher risk of belonging to the slight decline group. There was not a significant bivariate association between education and membership of the rapid decline class, but there were only 13 participants in this group, so the confidence intervals were very wide. They did not include education in their multivariate model predicting class membership.

Zahodne and colleagues used GMM to analyse cognitive trajectories from 2593 participants in the Washington Heights Inwood Columbia Ageing Project over 8 years of follow-up.(Zahodne, Wall, et al., 2015) They identified 4 classes. These were stable-high, stable-low, gradual decline and rapid decline. Rather than testing for an effect of education on latent class or on slope within latent class, their outcome was cognitive function adjusted for educational attainment and age. This implicitly either assumes that education contributes to brain resistance through passive brain reserve, or that the authors are attempting to correct for the effect of education biasing results with respect to underlying pathology as discussed earlier in this section. This method of adjustment allows for neither brain resistance nor cognitive reserve to be identified. Another notable feature of this analysis is that they had a sub-sample with structural MRI scan data. Using this they were able to validate their gradual decline and rapid decline latent classes as having greater hippocampal atrophy and thinner cortices. This lends important support to the validity of the GMM approach for correctly identifying an individual's disease status.

Hochstetler and colleagues analysed trajectories of MMSE score from a mix of healthy controls and patients with mild cognitive impairment or Alzheimer's disease (n=1192 in total) participating in the Alzheimer's disease neuroimaging initiative.(Hochstetler et al., 2016) They jointly modelled outcomes on the functional activities questionnaire and Alzheimer's disease assessment scale cognitive subscale. They identified 3 classes. These were high-stable, intermediate-moderate decline and low-rapid decline classes. Age, higher alcohol consumption, APOE status and amyloid positivity on positron emission tomography were strongly predictive of membership of the 2 poorer performing classes. They used education as a predictor of class membership but there was no significant association. Of 325 health controls in this study, 2 were placed in the intermediate class and none were in the rapid decline class. This supports the clinical validity of classes identified using GMM. A further analysis of the Alzheimer's

disease neuroimaging initiative data was presented by Ding and colleagues once additional follow-up was available.(Ding et al., 2019) They used data from older adults who were initially diagnosed as cognitively healthy (n=219) and with mild cognitive impairment (n=372) over 9 years of follow-up and up to 12 measurement occasions. In the initially healthy older adults, they identified 6 latent classes and 5 latent classes in those with MCI. Education was used as a predictor of latent class and more education was associated with membership of the higher performance stable classes but not with membership of a rapid-curvilinear decline class.

Using the English longitudinal study of ageing (ELSA), which is the data I will also be basing my analysis on, two studies have been published giving group trajectories for cognitive function.(Olaya et al., 2017; G. Tampubolon, Nazroo, & Pendleton, 2017) Olaya and colleagues used growth-based trajectory models, closely related to GMM, to analyse 10 word immediate and delayed recall trajectories over 6 occasions and 10 years in n=9515 ELSA participants aged 50-79.(Olaya et al., 2017) They split the cohort by age at 65. In both age groups they found a very similar 4 class structure with low-decline, low-stable, medium-stable and high-stable groups of cognitive performance. Education was a very strong predictor of group membership. Tampubolon and colleagues in their analysis of the ELSA data focussed on the ability of latent class membership to predict the distal outcome of probable dementia at the end of the observation period.(G. Tampubolon et al., 2017) 4 latent classes are reported which essentially follow an ordinal scale in baseline performance with the worse performance classes also showing moderately faster decline. The effect of education either within or between class is not reported.

Downer and colleagues used data from 1336 participants of the Hispanic Established Population for the Epidemiologic Study of the Elderly who were observed during four Waves from 2004–2005 to 2012–2013. They separately examined 3 cognitive domains of memory, global cognition and non-memory tasks. In each of these three areas they identified 3 latent classes. The latent classes of cognitive function were high stable, intermediate-slight decline and low-rapid decline. There was a 95% rate of agreement between the classes for the different cognitive domains. Years of formal education was used to predict latent class membership. More years of education was strongly associated with greater likelihood of membership of the high-stable latent class.

Baker and colleagues performed a retrospective cohort study on 3441 patients from a large mental health trust in the UK who had had 3 or more MMSE conducted during their clinical care.(Baker et al., 2017) The identified 6 trajectories of change over time. 5 of these trajectories had ordinal levels of baseline and change in MMSE with only a single qualitatively different intermediate-rapid decline class. The study is focussed on the association of neuropsychiatric symptoms with class of cognitive trajectory and the effect of education is not reported.

Using the Korean Longitudinal Study of Ageing, Min analysed data from 2445 adults aged 60 or above without diagnosed dementia.(Min, 2018) Using a Korean version of the mini-mental state examination as their outcome over a period of 6 years, they had 2 latent classes. One high-

stable and one low-decline class. Higher levels of education strongly predicted membership of the high-stable class in a dose-response fashion.

Lee and colleagues combined data from several north American ageing study cohorts, including several of those already mentioned.(Seonjoo Lee et al., 2018) In their pooled sample of 13037 adults aged 72-85 with between 2 and 15 years of follow-up they identified 2 latent classes of change in episodic memory function. They found that higher levels of education were associated with membership of the non-declining class across most of the cohorts included.

These studies have all used GMM or a closely related methodology to analyse cognition in a population of elders but there are substantial differences between them. The majority of analyses identified 3 or 4 classes of cognitive decline. In all cases there was no more than 3 qualitatively different classes showing either stability, linear decline or accelerating decline. Additional classes tended to have differences in baseline performance but not change over time. Education was predominantly associated with membership of higher performing classes.

All analyses used sum scores of standardised cognitive batteries or single cognitive tests (delayed recall for example). None of these analyses used factor analysis to identify if their standardised battery was an adequate fit to the data. This meant they were unable to use factor scores longitudinally (known as multiple indicator growth) or test their measurement instruments for measurement invariance. Summed MMSE score was the most common outcome measure. This represents a significant limitation in the literature to date as significant untested assumptions are made about the measurement of cognition and the invariance of those measures over time. The MMSE in particular is also known to have a relatively strong ceiling effect.(Franco-marina et al., 2019) Moreover, few of the studies incorporated explicit management of non-random missingness. Of those studies which included education, few have focussed on the effect of education on class membership or trajectory within class. This is important because, whilst several studies interpret their results in terms of cognitive reserve, by making education a predictor of class this implicitly assumes a mechanism of brain resistance rather than reserve. This does not reflect the findings of clinicopathological studies which suggest that higher levels of education are not associated with differences in ADP. However, as shown by Hayden et al. and Pietrzak et al., the primary determinant of latent class membership is underlying ADP and SVD pathology.(Hayden et al., 2011; Pietrzak et al., 2014)

2.1.4 Dementia Pathology and Education

As stated in the previous section, the majority of the studies using a longitudinal mixture modelling approach to cognitive ageing have used education to predict class membership. This choice could mean that the authors are focussing purely on dementia as a level of impairment or clinical syndrome without reference to underlying pathological status. Alternatively, given that the primary driver of latent class is underlying ADP/SVD, it means that the primary hypothesis that tested is whether education contributes to brain resistance. This approach assumes that the mechanism through which education provides its effect is by reducing ADP/SVD. However,

this assumption is not necessarily supported by the clinicopathological literature.(Bennett, Arnold, Valenzuela, Brayne, & Schneider, 2014; Bennett et al., 2003; Brayne et al., 2010; Del Ser, Hachinski, Merskey, & Munoz, 1999; Farfel et al., 2013; Koepsell et al., 2008; Serrano-pozo et al., 2013)

Bennett and colleagues evaluated whether education modified the association between ADP at autopsy and cognitive function in 130 participants of the Religious Orders Study.(Bennett et al., 2003) They found no association between ADP and education. The did find strong evidence that education moderated the association of diffuse and neuritic plaques with cognitive function. This effect was strongest for a decrease in the negative effect of plaques on episodic memory and perceptual speed. Education did not moderate the effect of neurofibrillary tangles on measured cognitive function. This suggests that education may provide greater resilience in earlier disease stages when amyloid-β plaques are most closely associated with cognitive function than neurofibrillary tangles. The ECLipSE study harmonised 3 population cohort studies with post-mortem brain donation from 872 participants, of which 56% were demented at death.(Brayne et al., 2010) They found that more years in education were related to increase brain weight, consistent with theories of brain reserve. However, education was not independently associated with either ADP or SVD. The found that greater education was associated with a lower risk of dementia in a dose dependent fashion but that this was independent of ADP/SVD. Del Ser and colleagues early study included 87 individuals with Alzheimer's disease, Lewy body dementia or both.(Del Ser et al., 1999) They found no difference in the quantitative level of ADP, but did find higher levels of SVD in individuals with lower levels of education.

Farfel and colleagues performed a cross-sectional study of 675 individuals with low levels of formal education at autopsy.(Farfel et al., 2013) They found that, compared to a group with no formal education, individuals with more than 1 year of formal education had no significant differences in ADP, lacunar infarcts or cortical Lewy bodies. They did find that education was associated with SVD. They found that education moderated the effect of lacunar infarcts but not other pathological changes on informant reported cognitive function. Koepsell and colleagues performed a large autopsy study on 2051 participants drawn from 27 Alzheimer's disease centres across the USA.(Koepsell et al., 2008) As a clinical sample recruited from dementia assessment centres, only 13.7% of the sample was non-demented. They found no correlation between education and severity of ADP. However, they did find that the effect of education on cognitive function was attenuated at more advanced stages of ADP. This would support the neural compensation theory of cognitive reserve. Further analysis of this cohort by Serrano-Pozo and colleagues confirmed that the effect of education on cognitive function was independent of ADP and SVD.(Serrano-pozo et al., 2013) This also indicates that the effect of education on cognitive function is not mediated via brain resistance.

Taken together these studies make a comprehensive case that education does not contribute substantially towards brain resistance. This would suggest that if the underlying latent trajectory classes of cognitive function are believed to be primarily driven by pathology, then it is more

substantively coherent to model education as a moderator of the association between latent trajectory class and rate of cognitive decline.

2.2 Cognitively Stimulating Activities Literature Review

2.2.1 Cognitive Stimulating Activities

The ‘use it or lose it’ hypothesis posits that activation of neuronal activity is an important part of maintaining brain function into later life.(Swaab, 1991) Neuronal activation increases the potential for damage, such as DNA damage from increased oxidative stress, which is one of the primary drivers of cell ageing. However, the stimulation of neurons also triggers protective mechanisms, such as DNA repair.(Q. Bin Zhu, Bao, & Swaab, 2019) Within normal physiological limits the balance tends to favour repair when compared to less stimulated neurons. This may contribute to cognitive maintenance by reducing promoting Alzheimer’s disease pathology (ADP) clearance or preserving neuronal integrity despite the presence of ADP or cerebrovascular disease. Environmental stimulation is one mechanism by which this protective level of stimulation may be achieved. Cognitively stimulating activities (CSA) are an important potential source of this stimulation which makes them a promising modifiable factor for cognitive maintenance. They have received attention from academics and are also the behaviour most frequently identified by adults in western countries as reducing their risk of dementia.(Friedman et al., 2015; Sajeev et al., 2016; Yates, Ziser, Spector, & Orrell, 2016)

The Global Council on Brain Health has defined CSA as ‘mentally engaging activities or exercises that challenge a person’s ability to think’.(Global Council On Brain Health, 2017) This is the definition which will be used in this thesis. However, the breadth of the definition does create problems for researchers trying to operationalise the measurement of CSA. The list of possible activities which could be included under this definition is vast. The frequency, intensity and duration of those activities is also highly variable. Comparison across studies and across culture-bound activities is difficult when the types of activities undertaken by older adults can differ substantially.

This has led many studies to use a composite score comprised of many different activities.(Sajeev et al., 2016; Yates et al., 2016) This approach presupposes a degree exchangeability between different activities. It is only possible to sum different activities if the assumption is made that each activity has a similar effect size on cognitive maintenance. This essentially views CSA as part of a cognitively enriched lifestyle rather than as specific activities which may have different effects.

Typically composite scores of cognitive activity have been modelled as a linear or proportional odds ordinal scale.(Yates et al., 2016) This introduces an additional assumption regarding the additive effect of each cognitive activity which has to be correctly specified as linear, quadratic, or otherwise. Any interactions modelled this way must also be exchangeable. The advantage of a composite score approach is primarily a large reduction in dimensionality. As many activities could be classified as CSA, the number of potential interactions is very high. This reduces statistical power substantially, increases the likelihood of empty cell problems, and increases the risk of chance findings.

A further methodological difficult is that if CSA do improve cognitive maintenance, they may also influence propensity to further CSA, thus creating bi-directional causality (figure 2.3). This requires specific modelling approaches which will be given fuller treatment in the statistical analysis section and the marginal structural modelling section of the literature review.

Randomised controlled trials (RCTs) are the most common way of avoiding this problem but are beyond the scope of this thesis. Adequate RCTs are very difficult to conduct because of the long follow-up time needed, the aforementioned lack of agreement in operationalising CSAs and the practical and ethical implications of trying to randomised people's leisure time. Nonetheless, it is worth noting the limitations of the evidence regarding trials of targeted (mostly computerised) cognitive interventions. They have been found to be effective at improving the specific abilities tested in healthy old age, mild cognitive impairment and dementia. (García-Casal et al., 2017; Hill et al., 2017; Karbach & Verhaeghen, 2014; Lampit, Hallock, & Valenzuela, 2014) However, there is much more limited evidence for the effect of transfer to benefits to un-related cognitive tasks, on daily functioning or when used at home in more externally valid environments.(García-Casal et al., 2017; Guye & von Bastian, 2017; Kane et al., 2017; Lampit et al., 2014; Melby-Lervåg & Hulme, 2016) This makes understanding the effect of non-targeted CSA which are already accessible to older adults particularly important.

There are many ways which CSA in later life could be subdivided. For the present discussion, a broad distinction between leisure and non-leisure activities will be made. The literature discussed will be primarily based upon two systematic reviews focussed on leisure activities, one on retirement, another on volunteering and one further on general computer use specifically.(Guiney & Machado, 2018; Liapis & Harding, 2017; A. Meng, Nexø, & Borg, 2017; Sajeev et al., 2016; Yates et al., 2016) These reviews contain limited discussion of cognitive maintenance or the mechanisms by which observed associations may occur. Taking a broad view of cognitive maintenance as including mechanisms consistent with both brain resistance and resilience, then reduced risk of dementia or cognitive impairment is a valid and important surrogate for improved maintenance for activities undertaken in later life.

2.2.2 Cognitive enrichment.

The systematic review by Sajeev and colleagues operationalised CSA as information seeking and/or processing and focussed on the outcome of dementia risk.(Sajeev et al., 2016) A meta-analysis was not conducted because of the variation in how CSA was defined between studies. They included 12 papers from 11 cohort studies with a combined total of 13939 participants with a mean follow-up time of between 2.5 and 6.1 years. They only included studies with a clinical diagnosis of dementia. All studies reported on activities undertaken during the study period as their primary exposures of interest. CSA activities were combined in composite scores of either frequency (9), number of activities (3) or time spent (1). All studies estimated a lower risk of dementia with greater activity participation, this reached statistical significance in 9 of the studies. In studies which differentiated between low, moderate or high participation or between

stimulating or passive activities, high participation and stimulating rather than passive activities were associated with lower risk.

A major focus in this review was on a method for the estimation of the degree of reverse causation or unmeasured confounding which would be required to explain any observed associations. They calculated bias corrected estimates for a range of hypothetical unobserved confounders and from this were able to establish how large or common an unobserved confounder would need to be. For this analysis they focussed on a single study as an exemplar of the approach.(Akbaraly et al., 2009) Under this calculation a strong harmful unobserved confounder would need to be between 40-80% more prevalent in the low activity group to explain the observed association, whilst a weak harmful unobserved confounder would not explain the association under any of the simulated scenarios. This suggests that, whilst some unobserved confounding is likely, it is highly improbable that there is no association.

In their calculations for reverse causation, they estimated relatively realistic scenarios, such as 10% or 25% prevalence of unobserved pre-clinical dementia in the high CSA group. Under these scenarios, prevalence in the low CSA groups would have to be between 8-25% higher than in the high CSA group to totally explain the association. Whilst relatively unlikely, this is by no means impossible. This suggests that reverse causation is likely to be a greater problem than unobserved confounding. Methods to account for reverse causation are therefore particularly important to address in an analysis of the contribution of CSA to cognitive maintenance. The way in which the marginal structural models used in this thesis address this problem will be developed further in the methods section.

In addition to Sajeev et al., Yates and colleagues conducted a systematic review of the association of cognitively stimulating leisure activities with either risk of dementia or cognitive impairment.(Yates et al., 2016) They focussed primarily on this association and did not perform the additional bias analyses conducted by Sajeev et. al. They included 19 studies, of which 7 were longitudinal and 2 were case control studies of clinical samples with dementia. 15 of these studies were taken forward to meta-analysis. 11 found statistically significant reductions in the probability of dementia or cognitive impairment. Due a split in studies reporting risk ratio, odds ratio or hazard ratio of either cognitive impairment or dementia, they performed 5 separate meta-analyses. Of these the pooled hazard ratio (2 studies), risk ratio (3 studies) and odds ratio (2 studies) for dementia found significant reductions, as did the pooled odds ratio for cognitive impairment (5 studies). The overall reduction was estimates to be in the 40-50% range for higher levels of CSA. The pooled hazard ratio for cognitive impairment was not significant (3 studies).

The authors of the systematic review noted than in many of the studies, CSA was distilled from a range of activity scoring systems into an ordinal scale of low, moderate and high. The studies included are noted to consider the possibility of reverse causation in their discussions. However, the only step identified to reduce this risk was the exclusion of participants with dementia at baseline. The review authors note that further steps to address this problem are required in future research. They also comment that it is important for future studies to examine whether

specific activities, and general computer use in particular, are associated with reduced risk of dementia or cognitive impairment. In this thesis I will be attempting to both address the issue of reverse causation and examining the effect of specific CSAs including general computer use.

Despite the very consistent finding from these reviews of a positive association between CSA and dementia risk, some notable studies with different analytic approaches highlight problems with reverse causation.(M. J. Aartsen, Smits, van Tilburg, Knipscheer, & Deeg, 2002; Gow, Corley, Starr, & Deary, 2012) In their analysis of the Lothian cohort, Gow and colleagues made use of access to a confounder which is not typically accessible, childhood intelligence.(Gow et al., 2012) After adjusting for childhood intelligence, they found that the association between CSA and function was no longer significant. Another alternative method which has been employed is to utilise cross-lagged models. Examples of this type of analysis have found that, in general, past cognition is more strongly predictive of future cognitive activity, although this effect may differ depending on baseline levels of literacy.(M. J. Aartsen et al., 2002; Lifshitz-Vahav, Shrira, & Bodner, 2017)

Several studies have also used data from the English Study of Ageing to examine the association between CSA and cognitive function. Haslam, Cruwys and Haslam in their analysis of ELSA waves 3-5 focussed on the issue of group social engagement as opposed to individual social engagement or loneliness.(Leopold, Engelhardt, & Engelhartdt, 2013) They found that high levels of group engagement were associated with considerably better cognitive function and that this association was greater with advancing age. The relevance of this analysis to the current discussion is that many of their group activities, such as group memberships, activities and hobbies or attendance at cultural events, can also be viewed as CSA. Contrastingly, the individual level social participation, such as relationship quality and frequency of social contact would not typically be included under the umbrella of CSA. Their analysis also was unable to account for problems of reverse causation. Given that psychological symptoms of dementia include apathy, depression and social withdrawal, this is a substantial limitation.(Ismail et al., 2016) It is also not possible to determine whether the association of equal effects across activities is suitable.

Other analyses of ELSA have come to contrasting conclusions. Shankar and colleagues found that loneliness and social isolation (which also means participation in fewer CSA) were both associated with poorer cognitive function.(Shankar, Hamer, McMunn, & Steptoe, 2013a) In total contrast to Leopold *et al*, Rafnsson and colleagues found that in ELSA social isolation was not associated with dementia risk but that more intimate relationships were protective.(Rafnsson, Orrell, Orsi, Hogervorst, & Steptoe, 2017) Rafnsson and colleagues used Cox proportional hazards regression rather than multilevel modelling. This suggests that the association between social and cognitively stimulating activities and cognitive maintenance may be quite sensitive to model specification.

2.2.3 Specific Cognitively Stimulating Activities

The association between general or ‘meaningful’ computer use and cognitive function was the subject of the systematic review conducted by Liapis and Harding.(Liapis & Harding, 2017) They defined meaningful use as ‘technology selected specifically for or by the participant for the purpose of leisure, social connection or activities of daily living and intended to be either useful or pleasurable.’. This is the kind of self-directed computer use which is recorded in ELSA. It has been given separate attention to the two previous systematic reviews because computer use was not considered a CSA by the majority of the studies included in those reviews. Nonetheless, it falls within the definition of CSA used in this thesis.

They included 9 studies in their review.(Liapis & Harding, 2017) 5 of these studies were small studies (n<20) of the feasibility of implementing compute use with older adults and were not powered to assess for effects on dementia risk. The remaining 4 were from population cohorts which tested for an association between computer use and cognitive function, impairment or dementia. All found an association between greater computer use and reduced risk of cognitive impairment. Of particular note for this thesis are the analyses by Xavier, D’Orsi and colleague which utilise the data from ELSA.(d’Orsi et al., 2017; André J. Xavier et al., 2014) The 2014 paper included in the review by Liapis et al., focussed on cognitive decline whereas the 2017 paper focussed on dementia diagnosis.(d’Orsi et al., 2017; Liapis & Harding, 2017) The 2014 paper found a modest but statistically significant improvement of 3.07% in delayed recall of a 10 word list for current internet users in a generalised estimating equations model.(André J. Xavier et al., 2014) This effect was not modified by gender, age or wealth. The 2017 paper used participant self-reported or informant reported dementia as their primary outcome over 5 waves of data analysis.(d’Orsi et al., 2017) Using Cox proportional hazards regression they found an approximately 40% reduction in the hazard of dementia in those who were internet users. These studies are informative, but neither of these analyses make use of the validated cognitive testing for dementia and cognitive impairment available in wave 7 of ELSA. Additionally, their methods do not address the problem of time dependent confounding affected by past exposure which, as previously mentioned, which is a key methodological challenge in studies of cognitive maintenance.

Alongside internet use, another activity within the definition of CSA which will be used in my analysis is volunteering. Guiney and Machado both review the literature on the association between volunteering and cognition as well as develop a theoretical framework for the association.(Guiney & Machado, 2018) This framework posits that volunteering is a complex multifactorial exposure or intervention which increases cognitive activity, social activity and physical activity. This in turn improves cognitive maintenance and mental health more broadly. Both in turn lead to improvements in cognitive functioning.

In their literature review they identify 15 articles testing for an association between cognitive functioning and volunteering.(Guiney & Machado, 2018) Seven cross sectional studies of older adults consistently found that volunteering was associated with better cognitive functioning. Five longitudinal studies found that baseline volunteering had mixed results with 2 finding modest protective effects against cognitive decline and 2 others finding protective effects which neared

statistical significance. 2 longitudinal studies which accounted for change in behaviour over time were included in the review. Both studies found that those who maintained their participation in volunteering also had improved maintenance of their cognitive function. They identified only a small single RCT in institutionalised adults who already had moderate dementia. This study found no significant effect on rate of decline. Of those longitudinal studies included the analysis of the health and retirement study by Infurna and colleagues has subsequently been undated by Proulx and colleagues to include more detailed information about the amount of time spent volunteering.(Infurna, Okun, & Grimm, 2016; Proulx, Curl, & Ermer, 2018) The analyses are not directly comparable because of differences in the outcomes used (risk of impairment vs overall cognitive scores) but both appear to confirm an association of moderately improved cognitive maintenance in old adults who volunteer. Guiney and Machado's literature review did not include any analyses from ELSA and I have not been unable to locate any studies specifically aimed at the association between volunteering and cognitive function.

Alongside volunteering, working into later life is another potential source of stimulation which could contribute to improve cognitive maintenance.(A. Meng et al., 2017) In older adults this is usually viewed as later retirement. Meng and colleagues systematic review found only weak evidence that later retirement was related to better maintenance of crystallised cognitive abilities and conflicting evidence regarding fluid cognitive abilities.(A. Meng et al., 2017) Due to the exponential rise of dementia being most notable 5-10 years after retirement ages in most countries, a substantial follow-up period or very large cohort of young-old adults is needed to test this association. Starke and colleagues used a mixed effects model to examine the association between retirement age and episodic memory in ELSA.(Starke, Seidler, Hegewald, Klimova, & Palmer, 2019) They found that there was no significant difference in the rate of change in episodic memory after retirement.

As with the other areas discussed above, none of the included studies used methods to account for time-dependent confounding influence by past exposure. This is still relevant regarding retirement decisions for older adults who were becoming cognitively impaired at young-old ages. However, due to the lower incidence of cognitive impairment around retirement age than in later decades of life, this is perhaps slightly less problematic for retirement than for other CSA.

The other CSA from ELSA which I will use are newspaper reading, membership of a social club and membership of 'education, arts or music group or evening classes'. Limited research has been conducted on these specific areas.(Sprague et al., 2019) In the analysis of ELSA data by Haslam, Cruwys and Haslam discussed in section 2.2.2, they formed part of two of their measures of group social engagement.(Leopold et al., 2013) Whilst they found that group engagement was strongly associated with improved cognitive performance this enables us to say little about these specific activities.

Educational interventions have been attempted in a few small intervention studies.(de Medeiros, Mosby, Hanley, Pedraza, & Brandt, 2011; Iizuka et al., 2019; Sprague et al., 2019) De Medeiros and colleagues conducted a randomised trial which recruited 51 non-demented adults aged 67-

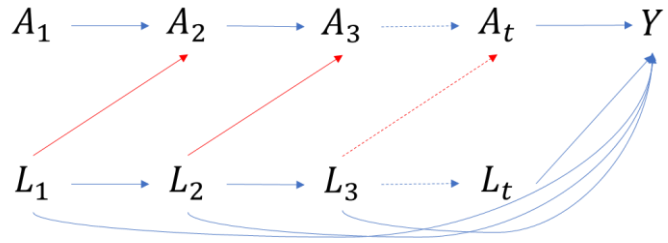
96 and allocated them to either an autobiographical writing class, a reminiscence group or a no intervention arm.(de Medeiros et al., 2011) They found no significant difference between the study arms after 34 weeks. However, with only 51 participants and a brief (for the timescale of cognitive ageing) follow-up, their power to detect differences was limited. Other intervention studies have examined art classes in healthy old adults with mixed results.(Iizuka et al., 2019) This would correspond to the types of classes ELSA respondents may be attending.

Reading has frequently been included in total enrichment scores, but some of the population research literature has analysed reading as an individual activity. Varghese *et al.*, performed a very highly cited analysis of the 469 participants aged 75-85 from the Bronx ageing study.(Verghese et al., 2003) They found that reading, as well as writing, dancing, playing musical instruments and playing board games, was associated with reduced hazard of incident dementia over 5 years of follow-up. However, in their primary analysis which was a fully adjusted cox regression they used a composite score of total cognitive activity. Zhu *et al.* analysed of 7 years of data on 6586 adults aged 65-015 in the Chinese Longitudinal Healthy Longevity Survey.(X. Zhu, Qiu, Zeng, & Li, 2017) They found a modestly protective effect of reading with a hazard ratio of 0.91 (95% confidence interval 0.84-0.99). Of relevance to the other CSA in this thesis, they also found no association between organised social activities and cognitive function. There have also been a couple of relevant small intervention studies, though these suffer from the lack of power and follow-up as discussed with similar trials previously. A book club was used as a control group by Shatil in their randomised controlled trial.(Shatil, 2013) They found no change in cognitive function after 4 months of intervention. Suzuki et al, randomised 58 older adults with mild cognitive impairment to either a picture book reading intervention or a series of lectures on staying healthy.(Suzuki et al., 2014) They found improvements in episodic memory, attention and executive function in the treatment arm.

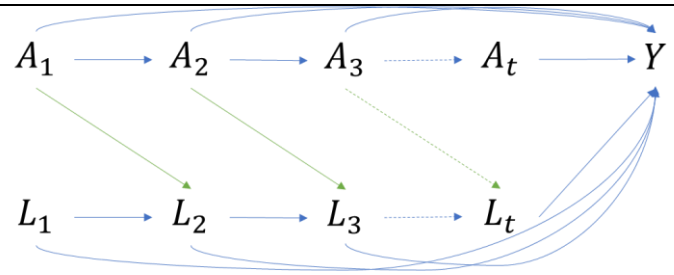
Taken as a whole, this literature indicates the potential for specific CSA to improve cognitive maintenance and reduce dementia risk. It also demonstrates problems conducting adequate randomised controlled trials on specific CSA and the need for methods of observational analysis which may be able to account for reverse causation in the form of time varying confounding affected by past exposure.

Figure 2.3 Causal diagrams showing potential longitudinal confounding structures in the association between cognitively stimulating activities, cognitive function and dementia diagnosis.

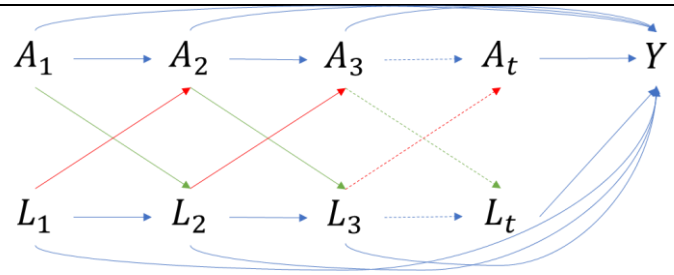
(1)



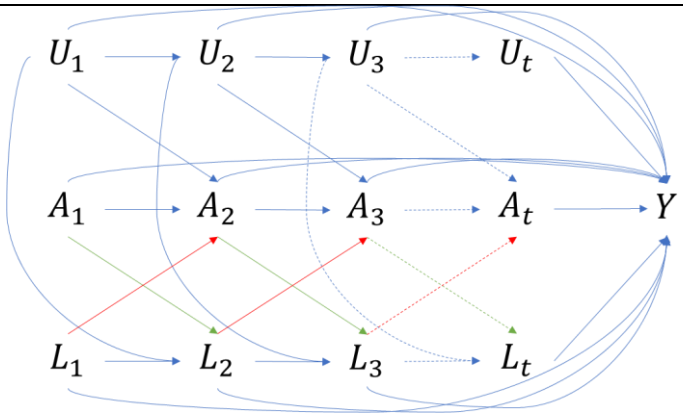
(2)



(3)



(4)



A_t is the exposure. L_t is a time varying covariate. U_t is an unmeasured confounder. Y is the outcome (cognitive impairment or dementia in this case).

(1) Shows time dependent confounding of the association between Y and A_t by $L_{(t-1)}$. This can be adjusted for using standard regression models. (2) Shows mediation of the effect of A_t by $L_{(t+1)}$. Alternatively, A_t can be seen as a confounder of $L_{(t+1)}$ and Y. If the quantity of interest is the effect of A_t on Y then standard regression adjustment with block indirect effects mediated by $L_{(t-1)}$. (3) Shows time dependent confounding affected by prior exposure. Here A_t has a causal effect on $L_{(t+1)}$ and L_t has a causal effect on $A_{(t+1)}$.

Conditioning on L_t with block the confounding of $A_{(t+1)}$ and Y, but will also block the effect of $A_{(t-1)}$ mediated via L_t. If the exposure A_t is cognitively stimulating activities, L_t cognitive function and Y is a diagnosis of dementia then much of the effect of earlier A_t CSA is likely to be mediated by future cognitive function $L_{(t+1)}$ and blocking this effect would not be substantively coherent. Hence standard regression is unable to produce unbiased effects if this is the underlying causal structure. (4) Illustrates the presence of unmeasured confounders.

3. Methods

This chapter introduces the statistical methods used in this thesis. Section 3.1 reviews methods for modelling trajectories of cognitive function over time and its relationship to predictor variables. Section 3.2 discusses models for informative missingness which can be applied alongside the methods in section 3.1 to account for missing data. Section 3.3 introduces the concept of longitudinal measurement invariance, from both frequentist and Bayesian perspectives. Section 3.4 introduces marginal structural models and their estimation using inverse probability of treatment and censoring weighting.

3.1 Modelling change in cognitive function over time

3.1.1 Growth curve models

As described in the previous section population heterogeneity is a major challenge in the study of cognitive ageing. Growth mixture models (GMMs), or the related latent class growth models (identical except for the lack of a random intercept) are one of the primary methods of identifying heterogeneity in change over time.

GMMs are, in essence, a combination of latent class analysis and latent growth curve models. The models estimate change over time and cluster similar trajectories into latent sub-populations. In this case latent classes of cognitive decline or stability which are likely to represent unobserved disease states. They are generalisations of conventional hierarchical models and it is from this starting point they will be described. This section draws extensively on the excellent chapters by Muthen, Bollen and Curran and the initial notation is based upon that used by Steele (Bollen & Curran, 2006; Muthen, 2004; Steele, 2014). First, random effects models will be described so that the parallel with equivalent latent growth curve models can be demonstrated.

Let our sample contain *n* individuals whose cognitive function has been measured on up to T occasions. Let y_{ti} be the score on a cognitive test for individual *i* (*i*=1,…,*N*) on occasion *t* (*t*=1,2,…,T). Let λ_{ti} be the metric of time for individual *i* which may vary between individuals on the same measurement occasion. Including this allows variation in the timings of measurements, though if all subjects are observed at the same time then this reduces from λ_{ti} to λ_i. So, in a random intercepts model which is specified as:

1. $$y_{ti} = \beta_{0i} + \beta_1 \lambda_{ti} + e_{ti}$$

$$\beta_{0i} = \beta_0 + u_{0i}$$

Here β_0 represents the mean intercept and β_1 the change in *y* per unit time, known as the growth rate. e_{ti} is the occasion specific error and u_{0i} the individual specific error or random effect. Both error terms are assumed to follow a normal distribution with mean 0 and variance

σ_e^2 and σ_{u0}^2 respectively. This is a random intercept model meaning individuals starting values are allowed to vary but the growth rate is held constant over all individuals.

These models can then be extended into random slopes models which allows the growth rate to vary across individuals. This can be expressed in the following format:

2. $$y_{ti} = \beta_{0i} + \beta_{1i}\lambda_{ti} + e_{ti}$$

$$\beta_{0i} = \beta_0 + u_{0i}$$

$$\beta_{1i} = \beta_1 + u_{1i}$$

The term β_{1i} now captures both the mean effect of a unit change in λ_{ti} on *y* and the individual error u_{1i} in this slope. So the growth rate for any given individual is now $\beta_1 + u_{1i}$. The random errors u_{0i} and u_{1i} are assumed to follow a bivariate normal distribution with mean 0, variance σ_{u0}^2 and σ_{u1}^2 and covariance σ_{u01}. It is possible to add non-linear growth functions or splines but they are not directly relevant to the current discussion and for simplicity these will not be outlined at this stage.

One can then add time invariant or baseline covariate x_1, such as gender or baseline age, which may be expressed as:

3. $$y_{ti} = \beta_{0i} + \beta_{0i}\lambda_{ti} + \beta_2 x_{1i} + e_{ti}$$

$$\beta_{0i} = \beta_0 + u_{0i}$$

$$\beta_{1i} = \beta_1 + u_{1i}$$

An alternative way of including the baseline covariate is to include it in the level 2 section of the model. This is more similar to the notation used in the structural equation modelling literature. This is expressed as:

4. $$y_{ti} = \beta_{0i} + \beta_{0i}\lambda_{ti} + e_{ti}$$

$$\beta_{0i} = \beta_0 + \beta_{02}x_{2i} + u_{0i}$$

$$\beta_{1i} = \beta_1 + \beta_{12}x_{2i} + u_{1i}$$

This is all expressed in the format known variously as a multilevel model, random effects model or growth curve model. These have been a popular way to measure cognitive change. Latent growth curve models are an alternative way of representing change over time. In the latent variable framework the observed values of an individual's trajectory are thought of as measurements used to represent an underlying and unobserved latent trajectory which gives rise to those measurements. Whilst, in general, the multilevel modelling approach is more computationally efficient and latent growth curve modelling more flexible, it should be noted that they produce comparable results when used to estimate the same set of relationships and assumptions.(Chou, Bentler, & Pentz, 1998; Curran, 2003)

A latent growth curve model with a linear growth assumption can be expressed as:

5. $$y_{ti} = \eta_{0i}\lambda_{0t} + \eta_{1i}\lambda_{1t} + \epsilon_{ti}$$

$$\eta_{0i} = \nu_0 + \zeta_{0i}$$

$$\eta_{1i} = \nu_1 + \zeta_{1i}$$

The similarities with equation 2 are quite clear. Instead of the first line representing the level 1 portion of a multilevel model, it now represents the measurement part of a latent growth curve (LGC). Similarly, the second and third lines here represent the structural part of a LGC which are broadly equivalent to latent intercepts and slopes.

The β's representing coefficients have been replaced with η_{0i} representing a latent variable for the intercept and η_{1i} which represents a latent variable for growth trajectory. The term ϵ_{ti} represents the measurement error variances, these are assumed to be homogenous over time. The terms ν_0 and ν_1 represent the intercepts for the corresponding latent variables and ζ_{0i} and ζ_{1i} their residuals. The loading factors for the latent variables are λ_{0i} and λ_{1t}. For the intercept all factor loadings are fixed to 1 so λ_{0t}=1. For the growth trajectory the loading is done such that it represents the appropriate measure of time, so for linear growth $\lambda_{1t} = t - 1$. Additional terms can be added to allow non-linear change over time.

One can then expand this model in a similar fashion to equation 4 by adding time invariant covariates:

6. $$y_{ti} = \eta_{0i}\lambda_{0t} + \eta_{1i}\lambda_{1t} + \epsilon_{ti}$$

$$\eta_{0i} = \nu_{01} + \nu_{02}x_{2i} + \zeta_{0i}$$

$$\eta_{1i} = \nu_{11} + \nu_{12}x_{2i} + \zeta_{1i}$$

Time-varying covariates can also be added. The time varying covariates directly influence y_{ti} in the structural part of the model. They can be allocated random slopes if this is substantively of interest. In this model used for this thesis the substantive interest was on influences on longitudinal trajectories rather than occasion specific measurements. Moreover, the focus is on the total effect of education and therefore only additional confounders, not mediators following education were included. The fact that the time varying covariates have a separate effect on both latent intercept and change over time is key for testing what type of brain resilience education provides. If there is only a significant effect on the intercept then this provides support for education providing brain reserve but not neural or neural compensation reserve. If there is an effect on latent growth then the direction of this effect will differentiate between neural reserve or neural compensation.

3.1.2 Growth Mixture Models

The advantage of an SEM approach is that this model can then be extended in a variety of ways which are different to the extensions possible in the multilevel modelling framework. One of the most straightforward SEM extensions is that either outcomes or covariates can

themselves be latent variables. This has particular potential for cognitive function because it is measured using a combination of tests and is not directly observable. Another important extension emphasised previously is the ability to model trajectories for multiple groups with different covariance structures.

The growth curve models above account for individual differences in baseline and change over time by allowing intercept and slope to vary between individuals. These individual differences are what is captured as the random effects. However, these models implicitly assume that the sample is drawn from a single population with a single set of population parameters around which individuals vary randomly. As described in previous sections this assumption is unlikely to hold in studies of cognitive ageing.

In a growth mixture model (GMM) individuals in the sample are allocated to classes of latent trajectory and vary randomly around the mean growth curve for their class. To do this one introduces a categorical latent trajectory variable $c_j$which represents unobserved group membership for individual *i.* There are K possible latent classes such that $c_j = 1, \dots, K$. Note that the time invariant variable now has effects on both likelihood of class membership and directly on the intercept and slope within class. Algebraically this can be represented using a superscript to denote that parameter estimates are contingent upon class membership.

7. $$y_{ti}^{c_i} = \eta_{0i}^{c_i}\lambda_{0t} + \eta_{1i}^{c_i}\lambda_{1t} + + \epsilon_{ti}^{c_i}$$

$$\eta_{0i}^{c_i} = \nu_{01}^{c_i} + \sum_i^k \nu_{02}^{c_i} x_{2i} + \zeta_{0i}^{c_i}$$

$$\eta_{1i}^{c_i} = \nu_{11}^{c_i} + \sum_i^k \nu_{12}^{c_i} x_{2i} + \zeta_{1i}^{c_i}$$

Here $y_{ti}^{c_i}$ denotes the outcome at occasion *t* for individual *i* who is a member of the latent class c_i. Otherwise all terms retain the same interpretation, except for where a c_i superscript indicates that the parameter is estimated within latent class and not the total sample. $\eta_{0i}^{c_i}$ is the class dependent latent intercept and $\eta_{1i}^{c_i}$ class dependent latent growth. The intercepts of these latent variables are $\nu_{01}^{c_i}$ and $\nu_{11}^{c_i}$. A vector of covariates which have a class dependent effect on the latent intercept and slope are given by $\sum_i^k \nu_{02}^{c_i} x_{2i}$ and $\sum_i^k \nu_{12}^{c_i} x_{2i}$. For simplicity all covariates in equation 7 have class dependent effects, however it is possible to have a mix of covariates with both class dependent and independent effects. The error terms for the intercept $\zeta_{0i}^{c_i}$ and slope $\zeta_{1i}^{c_i}$ are shown here also being free to vary by class. As with the covariates this can, and frequently is, be restricted to be constant across class. All residuals continue to be assumed to be normally distribute and independent and identically distributed, however they can be allowed to covary.

As described in the literature review the inclusion of latent classes of cognitive function is important to allow for population heterogeneity. The identification of population heterogeneity then allows for the effect of education on intercept and slope to vary by class. This explicitly models education as a source of brain resilience and not brain resistance. Furthermore, certain mechanisms of resilience may only function in health, or only need to function in disease, and

education's impact on these may not be equal. The ability to allow the effect of education to vary by class means that we are able to test this in an ante-mortem population sample.

3.1.3 Multiple Indicator Growth

As mentioned above one of the other advantages to the SEM approach to modelling change over time is that the outcome in question can be a latent variable itself. These models are variously known as multiple indicator growth curve models (MI-GCM, the preferred term here), second order growth curve models or curve-of-factors models.(Chan, 1998; Hancock, Kuo, & Lawrence, 2001)

These models have several advantages over single indicator growth curve models. These include greater statistical power, being the ability to separate change over time from measurement error and the ability to test measurement invariance over time.(Bishop, Geiser, & Cole, 2015; Chan, 1998; Ferrer, Balluerka, & Widaman, 2008; McArdle, Grimm, Hamagami, Bowles, & Meredith, 2009; von Oertzen, Hertzog, Lindenberger, & Ghisletta, 2010) The latter two are of particular importance for this research, with longitudinal invariance being particularly neglected in most cognitive ageing research.

Firstly, we return to the unconditional single class case in order to focus on the describing multiple indicator growth. The key difference is that on the left-hand side of the measurement model the outcome is no longer a single observed variable y_{ti}but a latent variable τ_{ti}. Let y_{1ti}, y_{2ti} and y_{3ti} be the 3 observed indicators of τ_{ti}:

8. $$\tau_{ti} = \eta_{0i}\lambda_{0t} + \eta_{1i}\lambda_{1t} + \zeta_{ti}$$

$$\eta_{0i} = \nu_0 + \zeta_{0i}$$

$$\eta_{1i} = \nu_1 + \zeta_{1i}$$

$$y_{1ti} = \alpha_{01} + \alpha_{11}\tau_{ti} + \epsilon_{1ti}$$

$$y_{2ti} = \alpha_{02} + \alpha_{12}\tau_{ti} + \epsilon_{2ti}$$

$$y_{3ti} = \alpha_{03} + \alpha_{13}\tau_{ti} + \epsilon_{3ti}$$

Here α_{01} to α_{03} represent the intercept of the observed factor indicators, also known as thresholds in the case of binary factor indicators. α_{11} to α_{13} are the factor loadings relating to the factor τ for each of y_{1ti} to y_{3ti}. The terms ϵ_{1ti} to ϵ_{3ti} represent the unique factor score for each individual at each time point. As mentioned above an important strength of this approach is that it is possible to separate measurement error from change over time and improvements in statistical power. The first of the individual research papers presents results comparing a multiple indicator growth curve model to those from a conventional regression analysis which uses of sum-score approach. This was an extension from initial descriptive work identifying to cognitive factors within the available dataset.

To extend equation 8 further, we can then add back into the equation multiple classes of change over time (note the factor structure is assumed to be invariant across classes):

9. $$\tau_{ti}^{ci} = \eta_{0i}^{ci}\lambda_{0t} + \eta_{1i}^{ci}\lambda_{1t} + \zeta_{ti}^{ci}$$

$$\eta_{0i}^{ci} = \nu_{0}^{ci} + \zeta_{0ti}^{ci}$$

$$\eta_{1i}^{ci} = \nu_{1}^{ci} + \zeta_{0ti}^{ci}$$

$$y_{1ti} = \alpha_{01} + \alpha_{11}\tau_{ti} + \epsilon_{1ti}$$

$$y_{2ti} = \alpha_{02} + \alpha_{12}\tau_{ti} + \epsilon_{2ti}$$

$$y_{3ti} = \alpha_{03} + \alpha_{13}\tau_{ti} + \epsilon_{3ti}$$

One of the core assumptions required by this model is that the factor structure (α_{01} to α_{03} and α_{11} to α_{13}) does not change over time. This is known as measurement invariance, which will be addressed in a subsequent section.

3.2 Modelling Missing Data Mechanisms in Bayesian Growth Mixture Models

Missing data is typically divided into data which is missing completely at random (MCAR; independent of observed and unobserved data, missing at random (MAR; independent of unobserved data conditional on observed data) and not missing at random (NMAR; not independent of unobserved data).(Rubin, 1976) Modern implementations of both frequentist full information maximum likelihood and Bayesian estimators are able to unbiasedly estimate growth curve models under MAR or MCAR but not NMAR.(Q. Chen & Ibrahim, 2014; X.-Y. Song, Lu, Hser, & Lee, 2011) The assumptions of MCAR or MAR are unlikely to be valid in the context of longitudinal analysis of cognitive function. The rate of decline is very likely to be related to the propensity to drop-out via death, loss to follow-up or other mechanisms. MCAR is therefore not a not a valid assumption. MAR is more probable, but it is still unlikely that, even with a wide range of covariates, dropout is random conditional on observed data.

This means that there is likely to be at least one unobserved, or latent, missing data mechanism. It is likely that this dropout mechanism is associated with both education and cognitive function. Individuals with higher levels of cognitive function are likely to be able to sustain a greater degree of loss before becoming functionally impaired. This would mean that those individuals are observed for longer than individuals with similar decline but lower education and poorer baseline functioning. If not adjusted for this would give the impression of greater education being associated with greater decline in cognition. If present, this would mean that the effect of education on cognitive reserve is mis-interpreted and so it is important to account for if possible.

A variety of approaches have been developed in the Bayesian and frequentist SEM literature for handling NMAR which focus on modelling that latent process. NMAR is also known as informative missingness because this latent process can be seen not simply as a nuisance in

the data but as an indicator of an important process in its own right. These models fall into two principle approaches, selection modelling (closely related to joint survival modelling) and pattern mixture modelling.(Muthen et al., 2011) These are able to account for at least one potential informative missingness model present in the data. The cost of this is that the missingness model must rely on untestable parametric modelling assumptions.(Enders, 2011)

The distinctive feature of pattern mixture modelling is that dropout is made a predictor of change over time.(Muthen, Asparouhov, Hunter, & Leuchter, 2011) One way of implementing this is that the latent variables for intercept and growth are regressed on missingness at each time point.(Little, 2008) This makes cognitive function at one time point conditional on survival to that time point. In the context of a GMM, an alternative formulation is to use missingness indicators to predict latent class of growth trajectory.(Roy, 2003) The latter model is not possible to implement in MPlus using Bayesian estimation at the current time and writing a software program to run these models was beyond the scope of this thesis.(Muthen et al., 2011) From a substantive perspective, these models are not a particularly good fit with cognitive function, as the model assumes that one's propensity to drop out influences cognitive function.

The other main family of NMAR models in SEM, selection models, do make good substantive sense. The distinctive feature being that cognitive function predicts dropout. The classical example of selection modelling is the Diggle and Kenward model, in which missingness at each time point is regressed on the observed outcome on that occasion and the previous occasion.(Diggle & Kenward, 1994) These models can alternatively be specified so that dropout is dependent not on occasion specific values, but on the latent intercept and slope. More recently, both of these models have been extended to the growth mixture modelling case in which missingness is dependent on latent class and the effect of slope and intercept may vary by latent class.(Beunckens et al., 2008; Lu, Zhang and Lubke, 2011; Muthen et al., 2011) If dementia pathology is the causal mechanism underlying latent class of cognitive trajectory, then it is highly plausible that both one's individual latent cognitive function and latent trajectory class would be likely to influence propensity to dropout. In this study, the analysis will be started using a model where missingness is dependent on latent class, intercept and slope. Applying this to the single indicator GMM with time-invariant covariates from equation 7, the latent class selection model is specified as:

10.
$$y_{ti}^{c_i} = \eta_{0i}^{c_i}\lambda_{0t} + \eta_{1i}^{c_i}\lambda_{1t} + + \epsilon_{ti}^{c_i}$$

$$\eta_{0i}^{c_i} = \nu_{01}^{c_i} + \sum_i^k \nu_{02}^{c_i} x_{2i} + \zeta_{0i}^{c_i}$$

$$\eta_{1i}^{c_i} = \nu_{11}^{c_i} + \sum_i^k \nu_{12}^{c_i} x_{2i} + \zeta_{1i}^{c_i}$$

$$logit(d_{ti}) = \beta_{0i}^{c_i} + \beta_{2i}^{c_i}\eta_{0i}^{c_i} + \beta_{3i}^{c_i}\eta_{1i}^{c_i} + \sum_1^n \beta_{(n+3)i}^{c_i} x_{ni} + \iota_{ti}$$

The first 3 lines of the equation remain unchanged to equation 7. A logistic regression is specified dropout at time t, denoted d_{ti}. $\beta_{0i}^{c_i}$ gives the class specific intercept for dropout at time t. The difference in the intercept between classes is interpreted as the effect of class on the

propensity to dropout. In this model there is a coefficient for both latent intercept ($\beta_{2i}^{c_i}\eta_{0i}^{c_i}$) and latent growth ($\beta_{3i}^{c_i}\eta_{1i}^{c_i}$) affecting propensity to dropout. If strictly implemented as per Beunckens (2008) then only a term for the latent intercept is included. A vector of covariates and coefficients for those covariates are included in $\sum_1^n \beta_{(n+3)i}^{c_i} x_{ni}$ as well as a residual variance for the dropout logistic regression ι_{ti}.

3.3 Measurement Invariance

3.3.1 Conventional (frequentist) Measurement Invariance

Tests of cognitive function are never pure measurements in the same sense as height or weight. Each test, whether aimed at a specific cognitive function or global cognition, measures that function and a range of other functions. For example, whilst a word recall task primarily tests short term memory, it also tests other cognitive functions such as attention. As well as other cognitive functions, physical functions such as hearing will also be tested by this task to a certain degree. This is the primary motivation for analysing cognitive function as a latent variable. Several tests of one primary function will place different demands on secondary cognitive and physical functions required to complete each of the tasks. By extracting the common variance between different working memory tasks, one is able to estimate a latent factor of working memory.(Horn & McArdle, 1992)

In the longitudinal setting, this creates a problem because these secondary functions may be affected at a different rate to the primary function by ageing or disease.(McAvinue et al., 2012; Wiegand et al., 2014). Even the way the data is collected, for example if the same task was performed using paper on one occasion and computerised later in the study, could affect the relative contributions of each function to the overall score. Practice effects can also play an important role. (Calamia, Markon, & Tranel, 2012) For example, more efficient retrieval strategies could be used in a working memory task which would increase the relative contribution of working memory to the overall test score.

Say one has a working memory task. At baseline this task measures 80% working memory, 15% attention and 5% hearing. After 8 years of follow-up this samples attention and hearing have deteriorated more rapidly than their working memory. The latent variables now measures 70% working memory, 20% attention and 10% hearing. You have no way to measure this directly. So, if your participants overall scores have reduced, you cannot be certain whether the change is due to deterioration in working memory, the primary cognitive function of interest, or one of the secondary functions.

In factor analysis, this is known as measurement invariance (MI).(van de Schoot et al., 2012). MI has been identified as a problem in longitudinal studies of cognitive function since at least the late 1980s and early 1990s. (Horn & McArdle, 1992; Schaie, Willis, Jay, & Chipuer, 1989).

Population research on cognitive function has frequently overlooked this issue.(Blankson & McArdle, 2013; McArdle, Fisher, & Kadlec, 2007; Wicherts, 2016). Most often a summed score, such as the MMSE, is used and the instrument's measurement properties are not examined.

If MI is ignored, it biases estimates of latent cognitive function towards the direction of the change in latent intercept.(Ferrer et al., 2008; Horn & McArdle, 1992; van de Schoot et al., 2013; Wicherts, 2016; Widaman, Ferrer, & Conger, 2011). An increase in latent intercept means an over-estimation of latent cognition. Practice effects would be expected to increase the latent intercept leading to an over-estimation of cognitive ability at follow-up visits. This would lead to an underestimation of decrease over time.(Wicherts & Dolan, 2010)

A change in factor loading is more complex and less predictable.(Wicherts, 2016) Take the example of the factor loading for a mean centred continuous variable where at a second time point the factor loading has decreased. In this case the latent mean will be under-estimated for those with an above mean score and over-estimated for those with a below mean score. For example, increased sensory impairment at later visits would weaken the association between measurable and latent cognitive function. This would decrease the factor loading and lead to overestimation of cognitive function for low scorers and underestimation for high scorers as time progresses. See Wicherts (2016) for a clear and concise discussion of this problem with additional illustrations. If either form of invariance is present it may lead to either quantitative or qualitatively incorrect inferences regarding the cognitive reserve provided by education.

We return to focus on the measurement part of our earlier growth model. Underlying the set of *l* (*n*=0,…,*l*) continuous observed variables (*y*) that have been measured, there is a latent variable τ (B. O. Muthén & Asparouhov, 2013; van de Schoot et al., 2013). As above, they are measured in individual *i* at time *t.* The measurement part is:

11. $$y_{lti} = \alpha_{0lt} + \alpha_{1lt}\tau_{ti} + \epsilon_{lti}$$

Here y_{lti} is the observed value of variable *l* at time *t* in individual *i*, α_{0lt} is the intercept for variable l at time *t*. α_{11t} is the loading for variable l at time *t.* τ_{ti} is the value of the latent variable at time *t* for the variable l and ε_{lti} is the error for individual *i* at time *t* for observed variable l. This model assumes independence amongst the observed variables conditional on the factor. The residuals are uncorrelated with the factors and normally distributed with a mean of 0. The factor metric is usually set by fixing $\alpha_{1lt} = 1$ for one observed variable across all measurement occasions. By convention $l = 1$ is set as this reference variable, but the variable order is arbitrary. In equation 9 it can be seen that the assumption of measurement invariance is shown by the fact that there is no t subscript for either the intercepts (α_{01} to α_{03}) or loadings (α_{11} to α_{13}). Linear growth of a factor over time can be specified as in equation 8 above, but with the addition of those subscripts:

12. $$\tau_{ti} = \eta_{0i}\lambda_{0t} + \eta_{1i}\lambda_{1t} + \zeta_{ti}$$

$$\eta_{0i} = \nu_0 + \zeta_{0i}$$

$$\eta_{1i} = \nu_1 + \zeta_{1i}$$

$$y_{1ti} = \alpha_{01t} + \alpha_{11t}\tau_{ti} + \epsilon_{1ti}$$

$$y_{2ti} = \alpha_{02t} + \alpha_{12t}\tau_{ti} + \epsilon_{2ti}$$

$$y_{3ti} = \alpha_{03t} + \alpha_{13t}\tau_{ti} + \epsilon_{3ti}$$

For continuous variables, the specification of MI consists of 4 steps.(van de Schoot et al., 2012; Widaman et al., 2011)

i) The same variables load onto the same factors at each time point (the same vector of y_{lti} for each τ_{ti}).

ii) The factor loadings are equal at each time point ($\alpha_{1l1} = \alpha_{1l2} = \cdots = \alpha_{1lt}$).

iii) The intercepts are equal at each time point ($\alpha_{0l1} = \alpha_{0l2} = \cdots = \alpha_{0lt}$).

iv) The residual variances fixed across time ($\epsilon_{l1i} = \epsilon_{l2i} = \cdots = \epsilon_{lti}$).

If only *i* holds, this is known as configural invariance, *i-ii* weak invariance, *i-iii* strong invariance and *i-iv* strict invariance. In the case of binary observed variables the second stage, weak factorial invariance is skipped because the item probability curve is influenced simultaneously by loading and intercept.(L. K. Muthén & Muthén, 2014) In many respects, the most important criteria are i to iii because strong measurement invariance needs to be established in order to compare latent means over time (Ferrer et al., 2008; Widaman et al., 2011). If the assumption of strong MI does not hold, then mean differences over time in a latent variable of cognitive function cannot be clearly attributed to change in true cognitive function. This is because the scale of the dependent variable has changed. This means that the latent mean at one time point is not directly comparable to the latent mean at another time point. It should be noted that this same problem applies to any cognitive score made from combining multiple cognitive tests whether they are combined using factor analysis, simple addition of results or any other method. One significant advantage of the latent variable method is that unlike, say simple addition, it is possible to test whether this assumption holds true.

However, testing for measurement invariance is not always straightforward. The simplest informal approach involves simply running models with and without MI. The results can then be compared and if there is conflict between these results then it is inferred that MI is not present.(Widaman et al., 2011). The results from these models will never match perfectly. So, with this approach an implicit decision is made about the degree of conflict in the results which is acceptable before MI is rejected. This decision is made using substantive prior subject knowledge and includes an assumption about the acceptable degree of invariance. Within a frequentist framework there is no way to formalise this judgement.

The standard approach to formally testing MI is sequentially testing global model fit for each level of measurement invariance. Initially configural invariance is specified and global model fit

checked to see if this adequately matches the data. Each increasing level of strictness is then specified (weak, strong then strict) and the global model fit statistics compared. The basic test is the chi-squared test of model fit, However, with large sample sizes this is a very strict test and strong factorial invariance over time may be rejected even in robust longitudinal studies of cognitive ageing (Blankson & McArdle, 2013; B. O. Muthén & Asparouhov, 2013). Therefore, especially with large sample sizes, alternative fit indices, in particular the comparative fit index (CFI), are frequently used instead (Cheung & Rensvold, 2002; Meade & Bauer, 2007).

Recommendations for the change in CFI which establishes MI differ between studies and these recommendations vary between 0.01 and 0.002.(F. F. Chen, 2007; Cheung & Rensvold, 2002; Meade & Bauer, 2007; Meade, Johnson, & Braddy, 2008; Short, 2014). These are clearly large relative differences. The most appropriate cut-off depends on both the number of occasions used and the number of indicators for the latent variable being used. As with any diagnostic test, in diagnosing measurement non-invariance, the pre-test probability of measurement invariance is also important in determining the post-test probability of MI. However, in a frequentist framework there is no formal way to include this in the calculation.

In addition to these uncertainties with tests of MI, they are not informative about which parameters are invariant. The global tests of MI rely on the choice of an invariant reference indicator.(Shi, Song, Liao, Terry, & Snyder, 2017) The reference variable sets the scale against which the other indicator variables are tested for invariance. If reference variable is non-invariant then that scale is changing over time. This can lead to other invariant factor indicators testing 'false positive' for non-invariance. Equally, if other observed variables show non-invariance of the same direction and magnitude then they can test 'false negative' for invariance.

Assuming a non-invariant reference variable has been chosen, there still remains the question of identifying which indicator variables are non-invariant. To identify which parameters are non-invariant one can either relax each equality constraint in turn or use modification indices (which give a measure of the improvement in model fit which would result from relaxing certain modelling assumptions). Relaxing each equality constraint sequentially means allowing each loading (α_{1lt}) or intercept (α_{0lt}) at each time point individually to be different to the same intercept or loading at all other time points. The change in model fit can then be assessed. If the model fit improves then this parameter is likely to be non-invariant. This rapidly escalates with increased numbers of measurement occasions or indicator variables. 3 indicator variables measured at 3 time points means a total of 12 parameters need to be tested. Simply increasing this to 4 observed variables at 4 time points means 24 parameters. This increases further if the reference variable is changed in order to assess for a non-invariant reference variable. In addition to being time-consuming, it significantly increases the chance that random variation will lead to different invariance solutions being identified depending upon the order in which the constraints are relaxed (R. C. MacCallum, Roznowski, & Necowitz, 1992; Bengt Muthén & Asparouhov, 2012). Modification indices avoid this problem by testing parameters simultaneously. However, they are limited in application because they are only validated for two

time points, less than is necessary for growth curve analysis.(B. O. Muthén & Asparouhov, 2013)

An additional consideration is the source of the invariance once this has been identified. In some cases, a small number of parameters may be highly non-invariant. It may be possible to omit one of these indicators and thus avoid the non-invariance. Whilst losing information in this manner is undesirable, it is usually preferable to lose information but have valid and interpretable results. If it is not possible for either substantive reasons or because there are 3 or less observed factor indicators partial measurement invariance can used. This is where the non-invariant loadings are freed and the others held constant. This may avoid making incorrect assumptions about measurement invariance but does not eliminate difficulties in interpretation of the results.

Non-invariance may also result from multiple loadings have minor amounts of non-invariance. This may result from random sampling variation if there are a large number of parameters or from slight changes across multiple cognitive domains influencing observed scores on cognitive tests. Commonly used criteria for the rejection of strong or strict measurement are overly sensitive to the rejection of models with trivial invariance of multiple factor loadings in large samples.(Meade & Bauer, 2007) Substantively insignificant amounts of non-invariance may then lead to the rejection of MI in situations where it would not result in meaningful differences to the conclusions drawn from the analysis. There is no particular frequentist solution to either quantifying or solving this problem. When I found uncertain results regarding measurement invariance using frequentist tests in my analysis of the ELSA cognitive function data, a Bayesian approach to measurement invariance offered a possible solution. This formed the basis of the second paper presented in this thesis.

3.3.2 Bayesian Measurement Invariance

Bayesian structural equation modelling (BSEM) has the potential to address some of the problems outlined above with conventional measurement invariance. Approximate MI has been developed to take account of multiple small or moderate non-invariances in loadings, intercepts or thresholds which are large enough to cause the model to be rejected by conventional fit statistics but small enough to be substantively unimportant. This makes it possible in some situations to avoid having to reject an essentially valid model or having a valid model where one cannot compare latent means over time. Additionally, it provides a one-step method of identifying which parameters are invariant.(B. O. Muthén & Asparouhov, 2013; van de Schoot et al., 2013; Verhagen & Fox, 2013) This provides similar information to that which is obtained using modification indices but is not limited by the number of measurement occasions.

The basic effect of approximate MI is that instead of requiring that all loadings be exactly equal, they are instead ‘tethered’. They do not have to be exactly equal but are allowed to differ only by a substantively unimportant amount. This means that it is possible to compare latent means over time despite small amounts of non-invariance. As described above the conventional

condition which must be met for strong factorial invariance that for each of the observed variables factor loadings must be equal ($\alpha_{1l1} = \alpha_{1l2} = \cdots = \alpha_{11t}$) and the intercepts must also be equal ($\alpha_{0l1} = \alpha_{0l2} = \cdots = \alpha_{0lt}$).

Let ђ be the difference between α's such that $\alpha_{1l1} - \alpha_{1l2} = ђ_{l12}$, $\alpha_{1l2} - \alpha_{1l3} = ђ_{l23}$ and $\alpha_{1l1} - \alpha_{1l3} = ђ_{l13}$. Also let be the difference between v's such that $\alpha_{0l1} - \alpha_{0l2} = и_{l12}$, $\alpha_{0l2} - \alpha_{0l3} = и_{l23}$ and $\alpha_{0l1} - \alpha_{0l3} = и_{l13}$. The conventional frequentist assumption of strong invariance can then be defined in Bayesian terms as the strongly informative priors of $ђ_{lXX} \sim N(0,0)$ and $и_{lXX} \sim N(0,0)$ (B. O. Muthén & Asparouhov, 2013). From a Bayesian perspective, the factor loadings and intercepts are themselves random variables. This makes the assumption of 0 variance between them defined in the priors above much less plausible. If there is past evidence or substantive logic which suggests there will be extremely limited differences between intercepts or loadings then this more plausible assumption can be included in the prior. Aside from the practical reasons to consider approximate measurement invariance, this provides a strong theoretical rational for preferring the Bayesian approach in this situation. Exact equality is a very strong assumption in longitudinal studies of cognitive function due to issues such as random variation across many time-points, attrition or practice effects.(Blankson & McArdle, 2013; Putnick & Bornstein, 2016)

Approximate measurement invariance is implemented as a strongly informative prior with 0 mean and small variance such as $ђ_{lXX} \sim N(0,0.01)$ and $и_{lXX} \sim N(0,0.01)$ for all loadings and intercepts except 1 factor loading at one time point which is fixed to set the scale. The researcher can decide a priori how long to make the tether by specifying an appropriate prior for the difference between loadings or intercepts over time. The size of the prior variance therefore sets the length of the tether and formalises the degree of invariance which is allowable. Difference statistics are then calculated for parameter by dividing the difference from the mean for the parameter by the standard deviation of the differences ($\frac{[и-\bar{и}]}{\sigma_и}, \frac{[ђ-\bar{ђ}]}{\sigma_ђ}$). The difference at each time-point is tested to see whether it is statistically significantly different from the mean of the loadings at all time-points. This tells you if any of the loadings have broken the tether and show a degree of non-invariance beyond that believed to be unimportant by the researcher. If desired one can then either reduce or elevate the variance in the prior and inspect the pattern by which factor loadings change. As both numerator and denominator are related to the prior variance then this relationship can be unpredictable as the prior variance in changed. However, there is as yet no established body of literature on model selection. So, whilst these additional results may be informative, to some extent they invalidate the theoretical value of the prior as a quantification of pre-test probability which was chosen with a specific rationale.

Once this model with all loadings and intercepts specified as approximate MI has been run then a further model is run. If there are no parameters which break their tether then it is appropriate to specify a model with full MI in the traditional sense.(van de Schoot et al., 2013) If there are a small number of highly non-invariant parameters then partial measurement invariance performs optimally. If it is the case that the source of non-invariance is many parameters small amounts

of non-invariance then continuing with approximate measurement invariance is most appropriate.

As an additional note approximate MI overcomes the problems in identifying the truly non-invariant parameters caused by fixing one indicator's loadings at 1 for all time-points. This is because, as mentioned above, when using the Bayesian approximate MI approach, one need only fix single loading for a single observed indicator at a single time-point to 1 (B. O. Muthén & Asparouhov, 2013; Xu & Green, 2015).

3.4 Marginal Structural Models and Inverse Probability of Treatment and Censoring Weighting

3.4.1 Statistical Motivation

As described in the previous section, the flexibility afforded by growth mixture models allows detailed exploration of which aspect of brain resilience and cognitive reserve is contributed to by education. Drawing on the theoretical models outlined in the introduction cognitive reserve is conceptualised as 'the accumulation of neural resources before the brain is affected by age-related processes'.(Cabeza et al., 2018) This framework fits well with education because the vast majority of education occurs during early life, a period of particular brain growth and development. This model does not fit as well with activities which older adults can undertake in later life which may improve their cognitive functioning. Once one's level of cognitive reserve has been achieved, concern then turns to how to prevent this being lost. This is known as cognitive maintenance as is defined as 'the preservation of neural resources, which entails ongoing repair and replenishment of the brain in response to damage'.(Cabeza et al., 2018)

This can include mechanisms involved in both brain resistance and brain resilience. As damage may be due to either disease or ageing, the study of brain maintenance at a population level orientates towards a pragmatic approach focussed on preventing clinical levels of impairment as a common end point. Cognitively stimulating activities (CSA) are one of the most important potential means by which individuals may be able to improve their brain maintenance. However, because it is a dynamic process, studying how CSA may contribute to brain maintenance presents methodological challenges distinct to those found in studying reserve.

The challenge which we will focus on is that good cognition is associated with greater exposure to CSA and greater exposure to CSA may be related to improved cognition. So, if CSA improves cognition then it also increases the chance of re-exposure and thus confounds itself. Does better brain maintenance cause increased CSA exposure, does increased CSA exposure improve brain maintenance or both? This problem is known as time dependent confounding where the confounder is affected by prior exposure.(Daniel et al., 2013) This effect is particularly problematic in the study of dementia due to the long prodromal phase of the illness prior to overt functional impairment. There is often a period of years in which developing ADP may cause an observable change individual's behaviour and social exposures before functional

decline. This may lead to a situation commonly referred to as reverse causation. Because the social change is observed before the dementia it appears as if the social change is influencing the dementia when in reality it may be the other way around. The situation may be further complicated if both cognitive function and CSA affect each other. One method which is able to account for this type of confounding, but not yet used to study cognitive maintenance, is marginal structural models (MSMs) to which we now turn.

The following section draw broadly from the works of Robins (Robins et al., 2000) for the initial development of MSMs, the exposition of the approach by VanderWeele (VanderWeele, Hawkley, Thisted, & Cacioppo, 2011) and Daniel (Daniel et al., 2013), the implementation in Stata by Fewell (Fewell et al., 2004) and Bodnar (Bodnar, Davidian, Siega-Riz, & Tsiatis, 2004) for the informative application of a MSM of treatment over time with a single final outcome, as well as all their respective co-authors. We also reference papers by Zou who has developed the use of a Poisson regression with robust error variance estimation to directly estimate relative risk in preference to the odds ratio more commonly obtained with logistic regression.(G. Zou, 2004; G. Y. Zou, 2009) This tends to provide a degree of additional efficiency, but the primary reason for its use is simply the more straightforward interpretation of the results.

The need to account for time varying confounding affected by prior treatment when making causal claims from observational data was one of the primary motivations for the development of MSMs and implementation using IPTW.(Robins et al., 2000) Unlike time-invariant confounding, time varying confounding affected by past exposure cannot be adjusted for using standard regression even if measured adequately. For example, volunteering and employment in later life have been associated with better maintenance of cognitive function.(Clouston & Denier, 2017a; Jenkinson et al., 2013; Kivipelto, Mangialasche, & Ngandu, 2018) Better maintenance of cognitive function is also associated with likelihood of remaining in employment and either continuing or starting volunteering.(Clouston & Denier, 2017b; Shen, 2017) This creates a hypothetical causal model where participation in CSA improves cognition, which in turn increases the probability of continuing CSA participation. If CSA improves cognition and better cognition makes future participation more likely, then the treatment effect of earlier CSA will be blocked, or collider stratification bias will be introduced using standard regression adjustment (see figure 3.1 for a directed acyclic graph of the model, which more specifically represents the model used in this thesis than the generalised model presented in figure 2.3).(Daniel et al., 2013)

Let a binary variable of CSA exposure be denoted $A_1, A_2, \ldots, A_t$. Let $L_1, L_2 \ldots, L_t$ represent a vector of observed confounders at each time-point (including cognition at that time) and $U_1, U_2 \ldots, U_t$ be a vector of unmeasured confounder's at each wave. Let $C_1, C_2 \ldots, C_t$ indicate whether an individual was censored at each measurement occasion.(Daniel et al., 2013) Let Y be observed cognitive status at the end of follow-up. For illustration, we will imagine that L contains only one variable, cognitive function measured longitudinally. We will also not directly address censoring and unmeasured confounding but will review the theory behind MSMs for readers new to the topic to illustrate why we have taken this approach to the analysis.

Starting from A_t in figure 3.1a let us assume CSA affects longitudinal cognitive function but cognitive function does not affect CSA exposure. From A_t we have the paths $A_t \rightarrow Y$, $A_t \rightarrow L_{t+1} \rightarrow Y$, $A_t \rightarrow A_{t+1} \rightarrow Y$ and several paths via the descendants of A_{t+1}. There is no path from any L_t to any A_{t+1} meaning L_t does not confound the association between any A and Y. This means that conditioning on L_t in the mistaken belief it is a confounder will block the path $A_t \rightarrow L_{t+1} \rightarrow Y$ and underestimate the total effect of A_t on Y. If the total effect is the quantity of interest, a naïve analysis not adjusting for L_t will provide an unbiased estimate under this condition.(Daniel et al., 2013)

In the example in figure 3.1b, the probability of CSA exposure is now affected by cognitive function but cognitive function is not affected by CSA. It can be seen that L_t is a confounder of the association between A_{t+1} and Y. Under these conditions those continuing CSA would become a progressively more cognitively elite group giving the appearance of CSA causing improvement in cognition. This should be possible to account for using standard regression analysis. However, difficulty may still arise if the direct effect of A_t not mediated by future treatment is of interest. This analysis would need to be conditioned on A_{t+1} which is a collider on the path $A_t \rightarrow A_{t+1} \leftarrow L_t \rightarrow Y$. So adjusting for A_{t+1} would inadvertently induce a conditional association between A_t and L_t and, therefore, between A_t and Y even if no true causal association exists.(Daniel et al., 2013) As such the effect of A_t may be estimated incorrectly.

In figure 3.1c, CSA affects cognition which in turn affects the probability of future CSA, all of which affect the risk of cognitive impairment or dementia Y. This is time varying confounding which is affected by past treatment. We now wish to condition on L_{t+1} because it is a confounder of A_{t+2} and Y. However, doing so also blocks the indirect effect of A_t mediated via L_{t+1} and its descendants, meaning the estimate of the effect of A_t is likely to biased. If both A_t and A_{t+1} have paths to L_{t+2} and Y then there is a backdoor path $A_{t+1} \rightarrow L_{t+2} \leftarrow A_t \rightarrow Y$. Conditioning on L_{t+2} may then introduce collider stratification bias for the association between A_{t+1} and Y. Lastly, if there is an unobserved confounder not of A_t and Y but L_{t+1} and Y then conditioning on L_{t+1} will also create collider stratification bias along the path $A_{t1} \rightarrow L_{t+1} \leftarrow U_t \rightarrow Y$ (Daniel et al., 2013). So if using a standard regression, one must therefore assume that confounders are not affected by prior treatment. In the case of CSA and cognitive function this seems a very strong assumption to make.

Inverse probability of treatment weighting is an alternative means of estimation which avoids having to make this assumption. Before describing this further we will briefly describe the conventional notation as applied to our specific MSM.(Robins et al., 2000) In ELSA wave 7 a telephone interview for cognitive status was used to diagnose probable dementia or cognitive impairment. Prior to this, 6 waves of data were collected on self-reported participation in a range of CSA and cognitive function was measured using episodic memory and verbal fluency. This allows the estimation of the effect of CSA from earlier waves on risk of dementia or cognitive impairment whilst being able to account for the effect of earlier cognitive function on exposure to CSA.

Let $A_2, \ldots., A_6$ now represent observed CSA exposure at of the corresponding waves of ELSA. A_2 to A_6 can take the values of 0 for not exposed or 1 for exposed. The confounding effect of L on A is lagged in our model. From this point forward we will be using V_1 to represent CSA exposure at wave 1. This is included as a baseline confounder rather than an exposure in order to account for unmeasured confounding at baseline mediated via baseline exposure by blocking the path $U_0 \rightarrow V_1 \rightarrow A_2 \rightarrow Y$. Let $\bar{A} = (A_2, \ldots., A_6)$ and $\bar{L} = (L_1, \ldots., L_5)$. Then let $\bar{a} = (a_2, \ldots., a_6)$, denote all the possible combinations of exposure which the participants could have been exposed to. Let $Y_{\bar{A}}$ be the observed outcome for exposure history $\bar{A}$. There will be one exposure history for each individual where $Y_{\bar{A}} = Y_{\bar{a}}$ and others where $Y_{\bar{A}} \neq Y_{\bar{a}}$. These $Y_{\bar{a}}$, the expected outcome given an exposure history of $\bar{a}$, represent a counterfactual quantity, the outcome that would have observed if a hypothetical intervention had set CSA exposure to any given $a_2, \ldots., a_6$. In my model, the association between CSA and covariates is always lagged though this need not always be the case.

Given that we are using Poisson regression where $E(Y_{\bar{a}}) = \lambda_{\bar{a}}$ and the use of a natural log link function the MSM takes the form:

13. $\log(\lambda_{\bar{a}}) = \beta_0 + \beta_2 a_2 + \beta_3 a_3 + \beta_4 a_4 + \beta_5 a_5 + \beta_6 a_6$

It is not possible to directly estimate this MSM precisely because all $\bar{a}$ are not observed. However, we are able to estimate

14. $\log(\lambda_{\bar{a}} \mid \bar{A} = \bar{a}) = \beta'_0 + \beta'_2 a_2 + \beta'_3 a_3 + \beta'_4 a_4 + \beta'_5 a_5 + \beta'_6 a_6$

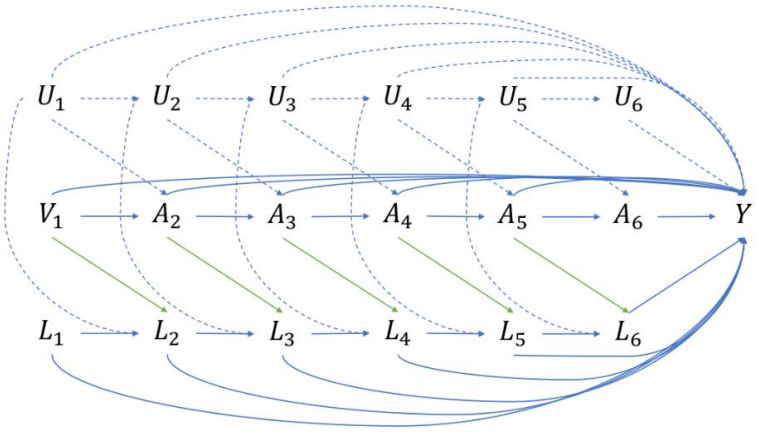
Whereas equation 13 describes the outcome under a hypothetical intervention to set the value of $\bar{a}$, equation 14 describes the relative risk of those with an observed history of $\bar{a}$. Assuming that all confounders are observed in L_1 to L_5 then we are able to unbiasedly estimate β_0 with β'_0, β_2 with β'_2 and so on.(Bodnar et al., 2004; Robins et al., 2000) L_1 contains both a subset of time-invariant confounders and the first measurement of time-varying confounders. To this model we add a term for baseline confounders $\beta_7 l_1$ and baseline CSA exposures $\beta_1 v_1$. These are required for the stabilised weights inverse probability of treatment weights used in the estimation of the model:

15. $\log(\lambda_{\bar{a}} \mid \bar{A} = \bar{a}, \bar{L} = L_1, \ \bar{V} = V_1) = \beta_0 + \beta_1 v_1 + \beta_2 a_2 + \beta_3 a_3 + \beta_4 a_4 + \beta_5 a_5 + \beta_6 a_6 + \beta_7 l_1$

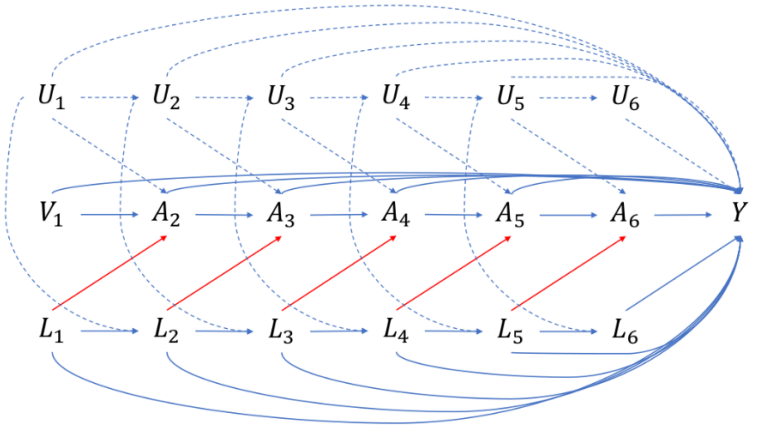
Interactions between exposure and either baseline exposure or baseline covariates are straightforwardly, but are omitted here for brevity. Interactions with time-varying covariates are not possible, but estimates of total effect are unbiased even if these interactions are present and not included in the model.

There are a small number of different techniques for estimating MSMs in the presence of time varying confounding. Probably the most commonly used of these is inverse probability of treatment and censoring weighting (IPTCW), which is the method I will employ. Instead of covariate adjustment in standard regression models, IPTCW accounts for confounding by weighting each individual by their probability of receiving their observed CSA exposure

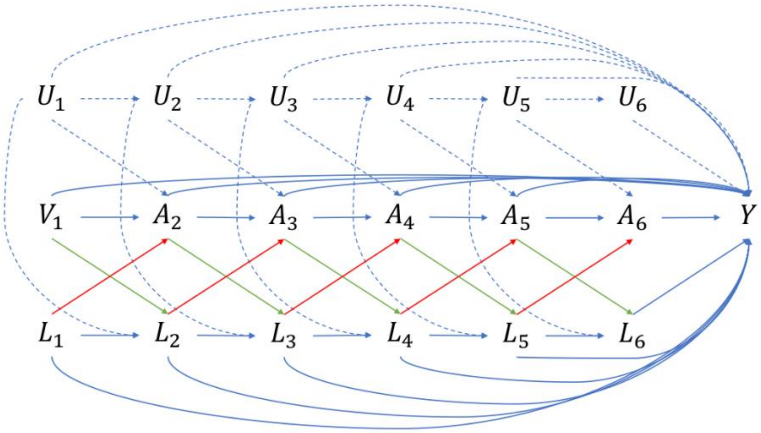
Figure 3.1. Causal diagrams for the effect of CSA on probable dementia/cognitive impairment showing: (a) indirect effects but no confounding.



(b) time dependent confounding.



(c) time dependent confounding affected by prior treatment.



Each line represents a hypothesised causal relationship, the colour has been added to highlight the difference between models only. Y represents probable cognitive impairment or dementia at wave 7. A_2, A_3…. A_t represents CSA exposure at each time point. V_1 represents CSA at baseline. L_1, L_2…. L_t represent all observed potential confounders. U_1, U_2…. U_t represent all unobserved potential confounders and the dashed lines unobserved potential causal relationships.

(hypothetical CSA treatment) estimated by their past CSA exposure and covariate history.(Daniel et al., 2013; Robins et al., 2000; VanderWeele et al., 2011)

The basic principle is that each individual is given a weight w inversely proportional to their probability of having received the exposure history they actually received conditional upon their measured covariate history and history of exposure prior to time t.(Hernán, Brumback, & Robins, 2002; Robins et al., 2000) This weight then effectively creates a 'pseudo-population' where there are a number of copies of individual i equal to the weight they are assigned and A_t is no longer confounded by L_{t-1}.(Robins et al., 2000) The overall weight is the product of an individual's weight at each time point. This is given by:

16. $w(t) = \prod_{t=2}^{T} 1/pr(A_t|A_2, \dots, A_{(t-1)}, V_1, L_1, \dots, L_{(t-1)})$

As these weights tend to have very high variance and may not be normally distributed due to a few individuals having very extreme weights, they are then usually stabilised. To produce the stabilised weight sw the numerator of 1 is exchanged for the probability of the observed exposure conditional on past exposure history and, as in our case, a vector of baseline covariates can also be included. If stabilised weights are used $E(Y_{\bar{a}})$ is now estimated within levels of the baseline covariates and additional terms must be added to the MSM as seen in equation 3 above. The stabilised weights are then given by:

17. $sw(t) = \prod_{t=2}^{T} pr(A_t|A_2, \dots, A_{(t-1)}, V_1, L_1) \,/pr(A_t|A_2, \dots, A_{(t-1)}, V_1, L_1, \dots, L_{(t-1)})$

This may alternatively be annotated in counterfactual form as:

18. $sw(t) = \prod_{t=2}^{T} pr(A_t = a_t|\bar{A}_{t-1} = \bar{a}_{t-1}, V_1 = v_1, L_1 = l_1) \,/pr(A_t = a_t|\bar{A}_{t-1} = \bar{a}_{t-1}, V_1 = v_1, \bar{L}_{t-1} = \bar{l}_{t-1})$

The weight at each time point is calculated for each measurement occasion t using a logistic regression model for the numerator and denominator.

The logistic regression models for the numerators were specified as:

19. $logit\ pr(A_t = 1|\bar{A}_{t-1} = \bar{a}_{t-1}, V_1 = v_1, \bar{L}_1 = \bar{l}_1) = \ \alpha_0 + \alpha_1 v_1 + \alpha_2 a_2 +, \dots, + \alpha_{t-1} a_{t-1} + \gamma_1 l_1$

The logistic regression models for the denominators were specified as:

20. $logit\ pr(A_t = 1|\bar{A}_{t-1} = \bar{a}_{t-1}, V_1 = v_1, \bar{L}_{t-1} = \bar{l}_{t-1}) = \ \alpha_0 + \alpha_1 v_1 + \alpha_2 a_2 +, \dots, + \alpha_{t-1} a_{t-1} + \gamma_1 l_1 +, \dots, + \gamma_{t-1} l_{t-1}$

These weights were then applied to estimate the MSMs above. Once the estimates have been obtained the results need interpreting. As mentioned above MSMs were developed as means of formally making causal inferences from observational data. If one is not making causal inferences, the assumptions underlying the MSM are the same as standard regression analysis with the exception of not having to assume that there are no confounders which are affected by past exposure (homoscedasticity, multivariate normality, no multi-collinearity).

Typical inference from observational data draws conclusions only regarding association between exposure and outcome, whilst theoretically remaining neutral to the issue of causality. In cases where prediction of the outcome is the primary concern this presents no real barriers. However, in many other contexts there is an implicit causal connection between exposure and outcome. Being able to make explicit causal claims using MSM is desirable, but rests on additional assumptions.(Cole & Hernán, 2008; Daniel et al., 2013) These are conditional exchangeability (no unmeasured confounders), positivity and correct specification of the structural (weighted regression) model and the exposure and censoring models. Whilst these assumptions are drawn out explicitly in the MSM literature, all of them would also be required to draw causal inferences from any analysis of observational data.

Conditional Exchangeability

Conditional exchangeability assumes that given the observed history of covariates $(\bar{L}_t)$ and past exposure $(\bar{A}_{t-1})$ you would observed the same propensity to dementia $(Y = 1)$ if those who were not exposed to CSA were exposed or those who were exposed were not exposed. (Cole & Hernán, 2008; Daniel et al., 2013) In other words you could switch who was exposed and who wasn't and it would make no difference to the outcome. More formally, conditional on $\bar{L}_t$ and $\bar{A}_{t-1}$, treatment received at time t is independent of the potential outcomes. This requires all predictors of dementia and exposure to CSA to have been measured appropriately and included in the exposure model used to calculate the IPTW. This is a very strong assumption in an observational study of complex social exposures and the influence on complex processes of brain maintenance.

One measure taken to address this in the MSM used is that wave 1 exposure and cognition are modelled as potential confounders and not exposures. A large portion of unmeasured confounding from pre-exposure covariates is likely to be mediated via baseline exposure and cognition. Pre-morbid IQ is an example of a confounder which is unmeasured in ELSA but whose effect is likely to be captured in large part by baseline cognitive function and baseline propensity to engage in CSA.

Using multiple CSA is another means by which there is some measure of testing for unobserved confounding. If one CSA exposure is associated with cognitive impairment but another CSA which is likely to be caused by similar social processes is not, then this increases the likelihood the effect is not due to unmeasured confounding. This is part of the rationale for why each CSA will be tested separately, not using a composite score of 'cognitive enrichment' as is often seen in the literature. Using a composite score also creates a path by which unmeasured confounders of every CSA included can bias the effect estimates and these vary between individuals with the same score. It is not possible to claim conditional exchangeability between the exposed and unexposed when exchangeability even between the exposed is not possible. Moreover, the counterfactual approach is based upon the theoretical construct of being able to set an individual's exposure to something other than what they actually experienced. This is much more clearly defined for an individual CSA than a composite score where the same level

of exposure can include very different activities. The cost of not using composite scores is that it is not possible to estimate potential multiplicate effects of exposure to several CSA.

Positivity

Positivity is the assumption that everyone *could* have had the exposure which they did not receive. More formally, that there are exposed and unexposed individuals within each strata of the confounding variables.(Cole & Hernán, 2008) It is not possible to directly estimate the effect of exposure within a certain stratum if everyone in that stratum was not exposed. In a randomised experiment this can be guaranteed because individuals are assigned to their exposure. For this reason, it is sometimes referred to as the experimental treatment assumption. This can occur either when there is a particular reason why a participant has a 0% probability of exposure or when empty cells occur because of small sample sizes or large numbers of covariates.

When considering social exposures with complex determinants over an extended follow-up period the problem of empty cells is particularly common. Parametric modelling assumptions can be made to estimate what would happen in these empty cells. Any inferences draw from the model then rely on those parametric assumptions. This is the same situation as with any observational data analysis. This generates a trade-off between the number of confounders which are included to reduce the chance of unmeasured confounding against the sparseness of the data and increasing reliance on parametric modelling assumptions.(Cole & Hernán, 2008)

3.4.2 Marginal Structural Models in cognitive epidemiology

Marginal structural models (MSMs) were developed by Robins and colleagues to formalise causal inference from observational data and account for confounding time dependent confounding where the confounder is affected by prior exposure.(Robins et al., 2000) The literature on their use in the field of social gerontology, or epidemiology, with dementia or cognitive function as their primary outcome was reviewed. This review was conducted in order to understand the scope of the current literature and to identify knowledge gaps. The main literature review on cognitively stimulating activities (CSA) drew upon recent systematic reviews conducted by other groups. It was not possible to know if studies implementing MSMs had been excluded from these reviews because of their methodology or other exclusion criteria. This meant it was important to identify any papers using MSMs in the study of cognitive function, especially those with any relevance to cognitively stimulating activities. With no existing systematic reviews to my knowledge regarding the use of MSMs in cognitive epidemiology, a rapid review of this literature was conducted which was updated in May 2019. The databases Scopus and PubMed were searched using the term *marginal structural model* with one of 3 additional keywords. These were *cognit**, *dementia* or *Alzheimer**. Documents citing the original article by Robins, Hernan and Brumback were searched with the same three keywords. The literature was restricted to studies of social exposures or social interventions where marginal structural models were used to estimate the effect of the exposure on cognitive function or risk

of dementia. This was a somewhat subjective inclusion criterion. Examples of exposures excluded include medications, exercise and cardiovascular diseases.

The searches by keyword and marginal structural model returned a total of 64 articles (including duplicates). After review of the title and abstracts, 2 appeared to be of relevance.(Barnes et al., 2010; Hikichi, Kondo, Takeda, & Kawachi, 2017) A third study regarding depressive symptoms and cognitive function was also included due to the complex interaction between depression and social factors as well as its methodological relevance.(Yao & Meng, 2015) Keyword search documents which cite the Robins et al article returned 407 articles (including duplicates). After review of titles and abstracts a further document was included.(Marden, Tchetgen Tchetgen, Kawachi, & Glymour, 2017)

Barnes and colleagues studied second hand smoke exposure in 970 adults aged 65 or older in the Cardiovascular Health Cognition Survey.(Barnes et al., 2010) Their outcome was incident physician diagnosed dementia over 6 years of follow-up. They used inverse probability of treatment weighted (IPTW) Cox proportional hazard model to estimated their MSM. Their focus in using the MSM was not on accounting for time-varying confounding affected by prior exposure but on causal mediation analysis. Specifically, to what extent the effect of second hand smoke on dementia risk was mediated or moderated by macrovascular disease. This study is an important contribution to understanding preventable risk factors for dementia but does not directly relate to either the substantive or methodological concerns of this thesis.

Yao and Meng studied the association between depressive symptoms and cognitive function in 3050 Mexican Americans aged 65 or older from the Hispanic established populations for epidemiologic studies of the elderly cohort.(Yao & Meng, 2015) The implemented their MSM using a IPTW linear regression testing the effect of 3 waves of depression scores on mini-mental state exam score on a 4^{th} measurement occasion. They contrasted a standard linear regression with covariate adjustment to the MSM and found that the effect of depression on the 3^{rd} occasion was larger in the MSM. This suggests that there may have been some time-varying confounding affected by prior exposure which was biasing the estimates towards the null. The MSM employed here is very similar in structure to that which will be used in this thesis and demonstrates the feasibility of the approach.

Hikichi and colleagues used an MSM to estimate the effect of visiting ‘salons’ on risk of cognitive functional impairment in Japanese adults aged over 65.(Hikichi et al., 2017) The used data from the Aichi Gerontological Evaluation Survey over 8 years. Over this time period public policy introduced into the area ‘salons’ which elders could attend a participate in a range of both intellectual (for example poetry or calligraphy) and physical (for example dance classes) activities. Estimating the effect of these salons raises similar difficulties to those for the cognitively stimulating activity participation of interest in this thesis. That is to say that cognitive function predicts participation and that participation may in turn improve cognition, creating complex bi-directional causality. They found that higher levels of salon attendance was associated with a modest reduction in the odds of cognitive disability.

These three studies have focussed on exposures in later life. An alternative life-course approach was taken by Marden and colleagues examining the effect of early life exposures.(Marden et al., 2017) Using data from the health and retirement survey on 10781 American adults aged 50 or more, they examined the effect of life-course socio-economic status (SES) on memory function and rate of decline. They took a critical periods approach with early life (parental), early adulthood (educational attainment) and later life (income, wealth and occupation immediately prior to the cognitive outcome) SES as the exposures of interest. Episodic memory was assessed using a marginal means regression model over the period 2002-2012. The IPTW were generated for SES at each point in the life-course account for prior exposure. The models were combined by including the measures of life-course SES as baseline exposures (plus interaction terms) in the regression model which was then estimated using IPTW. They found that higher SES in earlier life was more strongly associated with improved baseline performance and higher SES in late life with slower decline. This model estimated the total effects of SES and did not seek to account for any particular mediating pathways between SES and cognitive function. This is similar to the approach actually taken in the growth mixture modelling paper in this thesis.

These papers demonstrate that little use has yet been made of the potential for MSMs to advance our understanding of social factors related to cognitive maintenance and risk of cognitive impairment or dementia. One of the major limitations of the literature on cognitively stimulating activities as a whole is the potential for reserve causation to explain the findings.(Sajeev et al., 2016) MSMs have the potential to account some mechanisms by which reverse causation occurs. In doing so it will provide a greater degree of certainty that it is CSA which are improving cognitive maintenance and not that cognitive maintenance is causing greater exposure to CSA.

4. Data

This chapter will describe the English Longitudinal Study of Ageing, which is the source of the data used in this thesis, and key variables from it. Section 4.1 will provide a general overview of the study and participant recruitment. Section 4.2 will describe the measures of cognitive function used in this thesis. Section 4.3 will describe the main explanatory variables of interest, education and cognitively stimulating activities. Section 4.4 will describe the covariates and confounders considered.

4.1 The English Longitudinal Study of Ageing.

The English Longitudinal Study of Ageing (ELSA) is a nationally representative multidisciplinary cohort study of adults living in England aged 50 or more at recruitment in 2002/3.(Steptoe et al., 2013) ELSA has been described extensively in study reports and a journal published cohort profile, from which the information in this section is drawn.(Bridges, Hussey, & Blake, 2015; Littleford, Hussey, Begum, & Oskala, 2016; Steel et al., 2004; Steptoe et al., 2013)

The study sample was drawn from participants in Health Survey for England (HSE) years 1998, 1999 and 2001 who were born before 1st March 1952 and living in a private household. These are the core sample members.(Littleford et al., 2016) Those in their households who were new partners since HSE or ≤50 were also included as non-core members. This initial sample was nationally representative of the age specific English population. The initial HSE sample was drawn in a 2 step process.(Littleford et al., 2016) First, postcode sectors were randomly selected from the Postcode Address File. The postcodes were stratified by health authority and the proportion of households with non-manual occupation. Next, addresses were selected from within postcode sector. Within households up to 10 adults were eligible to be respondents. The ELSA sample was recruited only from those who responded to the HSE.

Additional recruitment was undertaken in waves 3, 4, 6 and 7 again from individuals who had previously participated in HSE. The wave 3, 6 and 7 refreshment samples were designed to maintain sample representativeness in the 50-55 age range, whilst the wave 4 refreshment sample recruited individuals aged 50-74 to replace losses from attrition.

Data was collected in biennial sweeps (wave 1 in 2002-2003, wave 2 in 2004–05, wave 3 in 2006–07, wave 4 in 2008–09, wave 5 in 2010–11, wave 6 in 2012–13 and wave 7 in 2014–15).(Littleford et al., 2016) Computer assisted interviews in the participant's homes were the primary source of data collection. Additional data was collected using self-report surveys which were left with participants following the interviews. In every other wave an additional visit was made by a research nurse who obtained measurements and blood samples.

The cross sectional response rates of eligible wave 1 participants invited to participate in each wave were 70% at wave 1, 82% at wave 2, 83% at wave 3, 77% at wave 4, 80% in wave 5, 86% in wave 6 and 83% in wave 7.(Bridges et al., 2015; Littleford et al., 2016) Longitudinal response rates, 'the proportion of remaining eligible wave 1 core members who gave an

interview in every wave up to and including the current wave' (Bridges et al., 2015), were 82% in wave 2, 71% at wave 3, 63% in wave 4, 59% in wave 5, 56% in wave 6 and 51.2% in wave 7.(Littleford et al., 2016) In wave 1 there were 12099 productive interviews carried out and of these 11391 were core members, 636 younger partners and 72 new partners. By the 7th wave 4894 of the original 11391 core participants responded. Of the missing core members in the 7th wave 3196 had died, 3132 did not respond to invitations to participate and 169 had moved out of Britain. Waves 1 to 5 are used for the first three papers presented in this thesis and waves 1 to 7 for the 4th paper. The specific sample sizes used for each analysis are presented with each paper. The sample size varied between papers depending on the exact waves used, whether or not the refreshment samples were included, completeness of data on specific variables, and whether outliers needed to be excluded for that analysis.

Ethical approval for ELSA was granted by the South Central Berkshire Research Ethics Committee (REC) through an application to the National Research Ethics Service (NRES).(Bridges et al., 2015) The current study was subject to the University of Manchester internal review process and no additional approval deemed necessary.

4.2 Measures of cognitive function.

ELSA contained a core battery of cognitive tests for the first 5 waves. This was modified in wave 6. This is why the first 3 papers presented used data from waves 1 to 5. A new set of tests was added in wave 7 which was based upon the telephone interview for cognitive status (TICS).(Crimmins, Kim, Langa, & Weir, 2011; Langa et al., 2017) This was used as the primary outcome for the 4th research question.

The cognitive tests were performed as part of the computer assisted interview.(Steel et al., 2004) The tests were orientation to time, immediate and delayed recall, prospective memory, verbal (semantic) fluency, and a letter cancelation task. Orientation to time was assessed by asking the participant to name the day, year, month and date. To assess immediate and delayed verbal recall a randomly assigned list of 10 common words was played from a standardised recording to participants. Delayed recall of the word list was tested after the other cognitive tests were undertaken to provide a distraction. The prospective memory task required participants to remember to write their initials in the top corner of a page they were handed. Participants were prompted if they did not complete the actions spontaneously. I treated this as binary with a correct response counting whether it was prompted or not. Semantic fluency was assessed by asking participants to name as many animals as they could in 1 minute. For the letter cancelation task participants were handed a clipboard with random letters in rows and columns. They were asked to cross out as many of the two target letters as possible in one minute. Participants were asked to complete the task by scanning from left to right as if reading. A memory index was calculated from the scores of the orientation, prospective memory and recall tasks.(Steel et al., 2004) The executive function index was calculated from score on

verbal fluency and letter cancellation. Both indices were scored out of 30. A global cognitive performance index combined these two into a total score ranging from 0 to 60.

The version of TICS which will be used in the 4th paper presented in this thesis is a 27-point scale. A higher score indicates better cognitive function. The score is calculated from a combined score on immediate and delayed 10-word free recall, backwards counting from 20 and serial 7 subtraction. This scoring system has previously been validated in the Ageing, Demographics and Memory sub-study of the Health and Retirement Survey.(Langa, Kabeto, & Weir, 2010) Scores of 0-6 are classified as probable dementia, 7-11 cognitive impairment no-dementia and 12-27 as normal.(G. Tampubolon et al., 2017)

4.3 Explanatory variables.

Educational attainment was recorded as no formal qualifications (this is the reference category in all analyses), high school completion (O-levels or equivalent), 6th form completion (A-levels of equivalent), non-degree level higher education and undergraduate degree or above. The first non-missing value was used. ELSA also contains data on the age of school leaving, rather than highest qualifications attained which was used in some analyses. Education was treated as time-invariant throughout all analyses.

I chose 6 CSA's from the range of activities reported by ELSA participants to represent variation in older adult's lifestyles and the type of cognitive challenge presented by the task. I also considered how clearly an intervention might be designed to alter an individual's exposure. This is important because of the counterfactual approach to causality used in marginal structural models. This is described in section 3.4. As an example, membership of a social club was chosen in preference to a measure of loneliness or social connectedness. It is possible to run a statistical model in which you estimate the effect of a hypothetical intervention setting a participant to 1 unit less of loneliness on a loneliness scale. However, setting someone to hypothetically be a member of a social club or not makes for comparatively unambiguous interpretation and clearer policy implications.

More general issues of measurement and interpretation were also considered in the selection of CSA. In particular, during earlier phases of analysis for the 4th paper gym class membership and having undertaken training in the past month were analysed as a potential CSA's. On closer inspection training was significantly under powered with less than 2% of participants above retirement age reporting participation. This was therefore dropped after initial inclusion. Gym membership was found to be associated with a reduced risk of cognitive impairment. However, when reflecting on the interpretation of the findings from this analysis the ability to interpret this finding was significantly limited. The crux of the problem was whether the observed effect was due to an additive effect of gym classes over and above self-reported exercise or whether gym membership simply correcting for measurement error in the self-report of exercise. I ultimately felt that this sufficiently undermined the findings to drop this exposure from the analysis as presented.

The CSA's chosen were working, volunteering, regular newspaper reading, attending arts/music/evening classes (hereafter 'evening classes'), internet or email use and attending a social club. All of these activities have been previously found to have an association with cognitive function. The past literature exploring the links between these activities and cognitive maintenance has been presented in section 2.2.3

For working and volunteering participants were asked in the main ELSA interview 'Did you do any of these activities in the last month?'. Participants attending evening classes and social clubs were asked in the self-completion questionnaire 'Are you a member of any of these organisations, clubs or societies?'. 'I read a daily newspaper' and 'I use the internet and/or email' were response options for the question 'Which of these statements apply to you?' which was also in the self-completion questionnaire.

4.4 Covariates.

A different set of covariates was used for each analysis and is specified in each paper. Age at baseline was centred for all analyses and wave of study was used as the metric of time. Whether or not the individual was at or above state retirement age was used as a separate time-varying variable. Gender and ethnicity (white and non-white) are treated as binary. Employment status was divided into retired (reference group), working, unemployed, long term illness and homemaker. For social class, the 5 category National Statistics Socio-Economic Classification (NSSEC-5) was used. Household wealth was grouped into quintiles. Household income was also grouped into quintiles. For some analyses income and wealth were divided into a binary variable with 60% lower and 40% higher categories. Marital status was classified as married or civil partnership (reference group), remarried, single, divorced/separated or widowed. For some analyses this was collapsed into married and non-married. Hypertension, angina, myocardial infarction, congestive cardiac failure (CCF), diabetes, stroke, COPD, asthma, osteoarthritis, osteoporosis, any psychiatric illness and Parkinson's disease diagnoses were all self-reported at each wave. Each condition was treated as irreversible. Self-reported usual cigarette consumption was divided into 0 per day (reference category), 1-9 per day, 10-19 per day and 20 or more per day. Frequency of alcohol consumption was given as less than monthly (reference group), monthly, weekly or daily or almost daily consumption. Depression was determined using the abbreviated 8 point version of the Centre for Epidemiologic Studies Depression scale (CES-D) which was dichotomised into less than or equal to 3 as non-depressed or 4 or more as (probable) depression.(Steffick, 2000) Parental smoking was comprised of two binary variables (smoker/non-smoker) for mother and father separately. Family structure in childhood was divided into being raised by two parents, being raised by a single mother or other family structures. Childhood social status was measured by father or main carers occupation divided into categories of unemployed, semi-skilled or manual, skilled manual, non-manual, managerial or professional and other (for example military service). Participation in other activities (including religious participation, charitable activities, daytrips, mobile phone use, holidays, voting, having a hobby and other class or society memberships)

were all based on self-report and binary. Self-rated health, hearing and eyesight were reported on a 5-point scale and treated as continuous when used. Participation in vigorous, moderate and light exercise was self-reported as a set of 3 binary variables.

Chapter 5. Differences Between Factor Analysis and Additive Scales in Population Studies of Ageing. Modelling Cognitive Function in The English Longitudinal Study of Ageing.

This paper presents the first paper which addresses the first research question of this thesis: whether using a sum score or factor score leads to different substantive conclusions about which variables are associated with cognitive maintenance. This question is divided into two main sections. The first is to establish the factor structure of the cognitive battery in the first 5 waves of data from the English Longitudinal Study of Ageing (ELSA). The second section is to compare the factor scores and sum scores (called index scores in the ELSA documentation). This was done through a qualitative comparison of the structure of the scores and through the results of a longitudinal multi-level model using either factor scores or index scores as the outcome.

This paper presents an example of the difference which can be made by using data driven factor scores rather than pre-defined sum scores in a real, rather than simulated, dataset. It was also an important step in the thesis as a whole. It was an important part of the development of the analysis, directly laying the foundations for the second and third papers. In presenting the results of my factor analysis it also makes this available for use by other researchers who may wish to use the cognitive tests in ELSA to answer their own substantive questions.

I conducted the analysis, drafted and revised the paper. My supervisors Prof. Chandola and Prof. Pendleton provided guidance on the analytic strategy and reviewed the drafts. Prof Gindo Tampubolon provided additional comments on a draft manuscript as part of an annual review.

This paper has not been published and is not under review at the time of submission.

Differences Between Factor Analysis And Additive Scales In Population Studies Of Ageing. Modelling Cognitive Function In The English Longitudinal Study Of Ageing.

Dr. Benjamin David Williams[1], BM,BS, MSc

Prof. Tarani Chandola[1], PhD

Prof. Neil Pendleton[2], MD

1. Cathie Marsh Institute for Social Research, University of Manchester

2. Institute of Brain, Behaviour and Mental Health, Stopford Building, University of Manchester

Abstract

Objectives

I firstly aimed to establish the factor structure of the cognitive function tasks in the English Longitudinal Study of Ageing (ELSA). I then compared the number and importance of differences in estimates of the association between cognitive function and common predictors when using factor scores or pre-specified sum scores (indices).

Method

I used data from ELSA, a large population-based ageing survey representative of the population of England aged ≥50, waves 1-5 (2002-2010). Exploratory factor analysis established the number of factors related to the cognitive function tests in different waves and confirmatory factor analysis was used to derive factors scores. Multilevel modelling was used to predict cognitive function factor scores or index scores (global, memory or executive function) using a number of common predictors of cognitive function and the differences in these associations were compared.

Results

A 3-factor solution was the best fit to the data, corresponding to attention (orientation to time), memory and visual scanning. The memory factor and index were closely correlated (0.95) and showed similar relationships with predictor variables, the global index demonstrated a lack of specificity, attention was moderately similar to memory and executive function indices whilst visual scanning did not correspond closely to any of the index scores.

Conclusion

In ELSA there was a loss of specificity resulting from the simple summation of different cognitive function tests into index score. Additive scales provide consistency across studies but I show that they can be misleading when compared to even simple factor models.

5.1 Introduction

Cognitive function is not directly measurable and must be estimated indirectly from single tests or their combination. The use of sum or mean scores combining multiple tests remains a commonplace approach in studies of cognitive maintenance. For example, in two systematic reviews on the association between education and cognition of the 26 research papers they included 21 used a summed or mean global cognitive function score as their primary outcome measure.(Beydoun et al., 2014; Plassman et al., 2010) This approach may be problematic for several reasons. It has previously been demonstrated to generate an unknown degree of error by, in effect, applying an arbitrary weight (or loading in factor analytic terms) to each individual cognitive test.(Jefferson et al., 2002; McGrory, Doherty, Austin, Starr, & Shenkin, 2014; Moafmashhadi & Koski, 2013) Moreover, the latent structure underlying test items may be either unexpected or overly general, leading to measures being combined in a manner which does not fit the observed data. For example, in their highly cited paper Miyake and colleagues demonstrated that cognitive tests commonly combined together as a measures of executive function in fact formed three separate factors.(Miyake et al., 2000)

Even in the mini-mental state exam (MMSE), probably the most commonly used measure for global cognition, studies in different populations have identified different factor structures underlying the test results.(Brugnolo et al., 2009; Shigemori, Ohgi, Okuyama, Shimura, & Schneider, 2010) Other studies have examined MMSE scores using item response theory methods and found substantial differences between the usual scoring system and the true weights responses should be given.(Ashford, Kolm, Colliver, Bekian, & Hsu, 1989; Gibbons et al., 2002; McGrory et al., 2014; Mungas & Reed, 2000) Whilst these studies have demonstrated some of the measurement problems with additive scales, they rarely explore what effect it has on the association of test score with either explanatory variables or change over time.

For example, by using global scores instead of using factors the specificity of patterns of cross-sectional and longitudinal association can be lost.(Salthouse, 2004) A global score could conflate decline in an age and pathology sensitive domain (such as working memory) and relative stability in another domain (such as verbal abilities).(Brugnolo et al., 2009; Shigemori et al., 2010) This would equate to a moderate decline in total score which would not accurately represent cognitive maintenance for either domain. Likewise, associations with exposures related to a specific cognitive factor could missed and the results interpreted as a moderate or non-association with global function. Population based or clinical studies analysed using sum or mean global scores form a considerable portion of the evidence on cognitive function.(Birks, 2006; Folstein, Folstein, & McHugh, 1975; Small, Rosnick, Fratiglioni, & Backman, 2004) It is therefore important to understand what effect, if any, these potential difficulties may have had on the estimates of the association between cognitive function and important sociodemographic, lifestyle and disease exposures.

5.1.1 Cognitive Tasks in ELSA

The English Longitudinal Study of Ageing (ELSA) which was chosen for this study has a set of pre-specified additive scores (termed index scores) divided into memory, executive function and global (which is the sum of the first two).(Huppert, Gardener, & McWilliams, 2006) The tasks in the memory index are self-reported memory, orientation in time, verbal recall and a prospective memory task. The executive function index includes category (semantic) fluency and a letter cancelation task.(Casey E. Krueger; Joel H. Kramer, 2010; Huppert et al., 2006) This formulation, whilst attractive in its simplicity, may not accurately reflect the underlying cognitive processes.

Considering the functional, anatomic and potential pathological processes which influence each of the individual cognitive tests undertaken one can see how this structure may struggle to represent the data. In ELSA, the first cognitive test is orientation to time. Orientation to time has been demonstrated to be the single most important indicator of current cognitive impairment and future prognosis out of those measures in the MMSE.(Folstein et al., 1975; Guerrero-Berroa et al., 2009; O'Keeffe, Mukhtar, & O'Keeffe, 2011; Tractenberg, Weiner, Aisen, Kaye, & Fuh, 2007) It has frequently been considered a measure of attention and it is also relatively strongly related to memory making it comparatively sensitive to changes related to Alzheimer's Disease (AD).(Chow, Hynan, & Lipton, 2006; Cossa, Sala, & Spinnler, 1995; Mioshi, Dawson, Mitchell, Arnold, & Hodges, 2006; Ryan, Glass, Bartels, Bergner, & Paolo, 2009)

Post mortem and in-vivo studies have found that disorientation in time is associated closely with pathology in the posterior cingulate gyrus and superior parietal lobe.(Giannakopoulos, Gold, Duc, & Patrick, 2000)(Hirono et al., 1998; Yamashita, Taniwaki, Utsunomiya, & Taniwaki, 2014) The superior parietal lobe is also part of the fronto-parietal network responsible for working memory tested in ELSA using a verbal immediate and delayed recall task.(Rottschy et al., 2012) Frontal regions used for working memory are also involved in the executive aspects of the categorical (semantic) fluency task.(Catheline et al., 2015; Jurado & Rosselli, 2007; Reverberi, Cherubini, Baldinelli, & Luzzi, 2014) However, semantic fluency also relies on lexical knowledge or the use of autobiographical memories, particularly in the later stages of the task which additionally utilise temporal lobe structures.(Henry & Crawford, 2004; Hirni, Kivisaari, Monsch, & Taylor, 2013; Sheldon & Moscovitch, 2012; Venneri et al., 2011) Thus semantic fluency may be strongly correlated with tests of both memory and executive function rather than executive functions alone.

The remaining tasks, prospective memory (PM) and letter cancelation (a test of visual scanning ability) also utilise multiple functional networks and discrete anatomical regions.(Cona, Scarpazza, Sartori, Moscovitch, & Bisiacchi, 2015; Leonards, Sunaert, Van Hecke, & Orban, 2000; McDaniel & Einstein, 2011; Squire, Noudoost, Schafer, & Moore, 2013; Todd & Marois, 2005; Uttl & Pilkenton-Taylor, 2001a) These examples show how complex the physiological and functional structures underlying maintenance of the different cognitive tests in ELSA are. As a well-known high quality multi-disciplinary ageing study it is important to explore the factor structure of the cognitive tests in ELSA because of the potential for error to be introduced by the inappropriate calculation of global function scores.(Steptoe et al., 2013) At the time of writing 21

studies using the cognitive function battery are listed on the ELSA publication database.(The Institute for Fiscal Studies, 2015) The majority of these have used the pre-specified cognitive function indices or their own sum/mean score. Moreover, the original ELSA investigators have proposed a unique set of summed scores using a limited battery of cognitive tests and as such this is relevant to many interdisciplinary studies not specifically examining cognitive function scores.(Steel et al., 2004).

To the author's knowledge, the factor structure of the data in ELSA has not been formally assessed in prior published work. Furthermore, the degree to which the different methods of combining the cognitive indices might bias associations has not been explored. Doing so will enable more informed appraisal of the reliability of existing work using similar scoring systems and inform future work on cognitive ageing in general and ELSA in particular.

Our research questions were:

1) When analysing variables associated with cognitive maintenance does analysing growth with a sum score or factor score lead to different substantive conclusions?
    a. What is the factor structure of the cognitive tests in the English Longitudinal Study of Ageing?
    b. How does this factor structure compare to the pre-specified cognitive index scores?

5.2 Methods

5.2.1 Participants

ELSA has been described in detail elsewhere.(Steptoe et al., 2013) In brief, the study sample was drawn from participants in Health Survey for England (HSE) years 1998, 1999 and 2001 who were born before 1st March 1952 and living in a private household or those in their households who were new partners or ≤50. This initial sample was nationally representative of the age specific English population. Additional recruitment was undertaken in waves 3 and 4, also from individuals who had previously participated in HSE. The wave 3 refreshment sample was designed to maintain sample representativeness in the 50-53 age range, whilst the wave 4 refreshment sample recruited individuals aged 50-74 to replace attritional losses. Data are collected in biennial sweeps by interview in the participant's homes. For this analysis data from waves 1 (2002) to 5 (2010) were utilised because the core cognitive battery was consistent through this time.

Response rates at each wave were 70% at wave 1, 82% at wave 2, 73% at wave 3, 74% at wave 4 and 80% in wave 5.(Steptoe et al., 2013) This resulted in final sample sizes of n=12009

in wave 1, n=9432 in wave 2, n=9971 in wave 3, n=11050 in wave 4 and n=10317 in wave 5 with the refreshment samples included.

5.2.2 Cognitive Measures

The cognitive tests were performed by computer assisted interview. Orientation to time was assessed by asking the participant to name the day, year, month and date. To assess immediate and delayed verbal recall a randomly assigned list of 10 common words was played from a standardised recording to participants. Delayed recall of the word list was tested after the other cognitive tests were undertaken to provide a distraction. The prospective memory task required participants to remember to write their initials in the top corner of a page they were handed. Participants were prompted if they did not complete the actions spontaneously which was included in the analysis as a correct response.

Semantic fluency was assessed by asking participants to name as many animals as they can in 1 minute. For the letter cancelation task participants were handed a clipboard with random letters in rows and columns. The aim was to cross out as many of the two target letters as possible in one minute. Participants were asked to complete the task by scanning from left to right as if reading. The number of the last letter reached was used as a measure of processing speed. All the non-binary variables were transformed to z-scores for the purpose of inclusion in the factor structure. The cognitive indices provided with the ELSA data release were used. The derivation of these indices is described in the wave 1 report.(Steel et al., 2004)

5.2.3 Covariates for Regression Analyses

Covariates were selected to encompass a range of potential predictors of cognitive function among older adults.(Plassman et al., 2010) Age was centred for the analysis and used as the metric of time in the multilevel model. Ethnicity was defined as white or non-white. Highest qualification was self-reported - no formal qualifications was used as the reference group. Age of completion of school education was included as a separate covariate as it may have an effect independent of qualification level.(Lenehan et al., 2015) Employment status was divided into retired (reference group), working, unemployed, long term illness and homemaker. For social class, the 5 category National Statistics Socio-Economic Classification (NSSEC-5) was used. Household wealth was grouped into quintiles. Marital status was classified as married or civil partnership (reference group), remarried, single, divorced/separated or widowed. Hypertension, angina, myocardial infarction, congestive cardiac failure (CCF), diabetes, stroke, COPD, asthma, osteoarthritis, osteoporosis and Parkinson's disease diagnoses were all self-reported at each wave. Self-reported usual cigarette consumption and divided into 0 per day (reference category), 1-9 per day, 10-19 per day and 20 or more per day. Frequency of alcohol consumption was given as less than monthly (reference group), monthly, weekly or daily or almost daily consumption. Depression was determined using the abbreviated 8 point version of

the Centre for Epidemiologic Studies Depression scale (CES-D) dichotomised into less than or equal to 3 as non-depressed or 4 or more as (probable) depression.(Steffick, 2000)

5.2.4 Statistical Analysis

The initial analysis focussed on establishing the factor structure observed in the data. The analysis was conducted in two stages. First, exploratory factor analysis (EFA) was conducted using each wave individually to identify the number of factors. Secondly, the structure of individual factors was established using confirmatory factor analysis (CFA).

The EFA was conducted using the weighted least squares means and variance adjusted estimator (WLSMV).(Barendse, Oort, & Timmerman, 2014; Beauducel & Herzberg, 2006) Geomin rotated factor loadings were used to permit correlation between factors and solutions with up to 4 latent factors were tested. The sampling weights provided with the ELSA dataset were used for the analysis. Comparative model fit was assessed using chi-squared test, root-mean square error of approximation (RMSEA), comparative fit index (CFI) and Tucker-Lewis fit index (TLI). Consideration was also given to substantive interpretability.

Following the EFA, the fit to the data of the most probable factor structure was confirmed using CFA in each wave, also using the sampling weights. The WLSMV estimator and the same fit indices were used. Due to the large sample size, greater importance was attached to the RMSEA, CFI and TLI than chi-squared test (Hu & Bentler, 1999). Factor scores were then generated using multilevel CFA in order to account for the clustering of scores within individuals. In order to provide comparability with the index scores invariance was assumed across time and socio-demographic groups and practice effects were not modelled separately to the effect of time.

To compare the indices or factor scores in relation to exposures the association of the variables listed above with the summed indices or factor scores was analysed. Multilevel models (level 1= waves of ELSA, level 2= individuals) were constructed for each of the factor scores and indices in turn. Computed factor scores were use in preference to multiple indicator multiple causes modelling as there is no comparable means of simultaneously performing the regression and accounting for measurement error when using index scores. This was because the aim of this paper is to focus principally on comparing the a priori factor structure and loadings implicit in using summed scores with data driven factor structure and loadings and no other aspects of SEM modelling. The factor scores and indices were standardized to permit comparability. Random effects at the individual level were included for age and gender. The non-response weights included in the dataset were used in order to weight results to be representative of the English population aged 50 or older.

The coefficients obtained from the regression analyses for the factors scores and indices were compared in 2 ways. Firstly, for each predictor variable the confidence interval for the regression coefficient was compared between each factor and each index. The total number of

differences between each factor and each index shows whether the magnitude of the associations is similar between the approaches. Secondly, the significance of the p-values for each coefficient was compared for each of the factors with each of the index scores and the total number of differences calculated. This gives an indication of what the differences in inferences about statistically significant predictors might be from the analysis. We compared the total number of differences in coverage and inferences between each factor and the index in both the extent they differed and how systematic the differences were.

The EFA and CFA were run using MPlus version 7.0 and the multilevel models run using MLwiN version 2.3 from Stata version 13.0 using the runmlwin command.(Leckie & Charlton, 2013; L. K. Muthén & Muthén, 2014; Rasbash, Charlton, Browne, Healy, & Cameron, 2016; StataCorp, 2013)

5.3 Results

5.3.1 Descriptive Statistics

The participant demographics are displayed in table 5.1. The gender, ethnic and marital composition of the sample was relatively stable at each wave with 55.4-56.2% female, 2.3-3.3% non-white and 54.9-56.3% married, 11-12.6% remarried, 5.2-6.1% widowed, 10.4-11.1% divorced or separated and 14.8-17.7% single. The proportion of participants with no formal qualifications decreased from 41.5% in wave 1 to 26.2% in wave 5. All other categories of qualification, except foreign qualifications, increased over the study period. In particular, the number of participants with a university degree rose from 11.5% in wave 1 to 19.6% in wave 5. Correspondingly, the number completing school at age 14 or failing to finish school decreased from 23.3% to 12.1% and the number completing full time education at age 19 or older increased from 12.4% to 19.6%. Membership of NSSEC class 5 (semi-routine and routine occupations) decreased from 35% to 30% whilst membership of class 1 (managerial, administrative and professional occupations) increased from 29.6% to 34.6%. The number of participants who smoked decreased from 17.0% to 11.7%.

Trends in the cognitive function measures can be seen in table 5.2. A general trend of modestly improving mean performance across waves, probably representing a combination of practice effects and selective dropout. From wave 1 to wave 5, the mean verbal fluency score increased from 19.3 to 20.8 animals named (standard deviation (S.D) increasing from 6.4 to 6.9). The mean immediate recall increased from 5.5 to 5.9 words recalled (S.D stable at 1.8) and the mean delayed recall increased from 4.0 to 4.6 words (S.D increasing from 2.1 to 2.2). The mean number of correct letters cancelled increased from 18.7 to 18.9 (S.D reducing from 6.0 to 5.5) and the number of letters missed decreased from 5.5 to 4.6 (S.D reducing from 5.0 to 4.0). The mean number of letters completed decreased from 307.2 to 301.7 (S.D decreasing from 96 to

86). Of the orientation tasks, naming the correct date was consistently the most difficult with 19.2% giving the incorrect date in wave 1, decreasing to 16.7% in wave 5. Similarly, the proportion of participants who gave the incorrect year, month or day of the week decreased from waves 1 to 5.

The correlation matrix of cognition function tests for all waves combined (appendix 5.1) shows that verbal fluency and prospective memory correlated moderately with all other variables except orientation to date and the letter cancelation tests. Immediate and delayed verbal recall correlated to a high degree (0.729) but also had weak to moderate correlation with orientation, prospective memory, verbal fluency and letter cancelation. The number of correct and missed letters correlated highly (-0.888) and correlated poorly with all of the other variables. The number of letters completed correlated weakly with all other variables- its strongest correlation was with the number of letters correct at -0.281. This suggests that separating letter reached from correctly cancelled and missed letters may have been relatively successful in disaggregating processing speed and executive control, which is a common difficulty in testing these specific cognitive domains.(Cepeda, Blackwell, & Munakata, 2013)

5.3.2 Factor Analysis

The model fits for the EFA showed that models with 1 or 2 factors fit poorly in all waves. Using the RMSEA, TLI and CFI criteria, the 3-factor solution was deemed to be an adequate fit to the data (RMSEA probability of a value <0.05 =1, TLI >0.95 and CFI >0.97 in all waves). The X^2 test of model miss-specification remained statistically significant for both the 3 and 4 factor solution in all waves. Given the fit to the data, interpretability and parsimony the 3-factor solution this was preferred to the 4-factor solution.

The 3-factor solution identified by EFA for each wave can be seen in appendix 5.3.2. The first factor was predominantly comprised of the orientation questions (Date 0.601 to 0.681; Month 0.883 to 0.994; Year 0.834 to 0.931; Day 0.698 to 0.826) with small loadings from prospective memory (0.267 to 0.403), verbal fluency (0.187 to 0.282) and processing speed (0.202-0.285). The second factor was comprised principally of immediate (0.831-0.891) and delayed (0.757-0.820) recall as well as verbal fluency (0.396-0.488) and prospective memory (0.204-0.341). The third factor was comprised of the letter cancelation task with the rate of correct (-0.971 to -1.026) and missed (0.829 to 0.897) letters as well as letter completed (0.365 to 0.398). From this point forward, these factors will be called attention (orientation to time), memory and visual scanning respectively.

This factor structure was then checked using CFA. This found good fit to the data in all waves using RMSEA (all <0.033), TLI (all >0.97) and CFI (all >0.98). The factor structure identified across waves using multilevel analysis can be seen in figure 5.1. The loadings on the orientation factor were fixed at 1 for year and estimated as 0.334 (S.E 0.019) for date, 0.712 (S.E 0.043) for month and 0.548 (S.E 0.032) for day of the week. Loadings were <0.3 in the EFA were not included in the CFA model. The loadings on the memory factor were fixed at 1 for

Table 5.1 ELSA Participant Demographics by Wave

Variable		Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
	n	12095	9425	9760	11035	10260
		Mean				
Age		64.2 (s.d 11.1)	65.8 (s.d 10.7)	64.6 (s.d 11.4)	65.3 (s.d 10.5)	65.1 (s.d 13.3)
		Proportion				
Gender	Male	5334 (44.1%)	4124 (43.8%)	4292 (44.0%)	4922 (44.6%)	4566 (44.5%)
	Female	6761 (55.9%)	5301 (56.2%)	5468 (56.0%)	6113 (55.4%)	5694 (55.5%)
Ethnicity	White	11658 (96.3%)	8988 (97.7%)	9203 (97.0%)	10302 (96.8%)	9513 (96.7%)
	Non-white	360 (3.0%)	208 (2.3%)	285 (3.0%)	342 (3.2%)	325 (3.3%)
Marital Status	1st Marriage	6811 (56.3%)	5231 (55.5%)	5355 (54.9%)	6107 (55.4%)	5730 (55.9%)
	Remarried	1349 (11.2%)	1036 (11.0%)	1182 (12.1%)	1395 (12.6%)	1190 (11.6%)
	Widowed	669 (5.5%)	491 (5.2%)	568 (5.8%)	670 (6.1%)	592 (5.8%)
	Divorced or separated	1260 (10.4%)	997 (10.6%)	1082 (11.1%)	1229 (11.1%)	1156 (11.3%)
	Single	2003 (16.6%)	1669 (17.7%)	1572 (16.1%)	1632 (14.8%)	1586 (15.5%)
Qualifications Obtained	No Qualifications	5007 (41.5%)	3561 (37.9%)	2815 (29.0%)	2979 (27.9%)	2664 (26.2%)
	High School	2555 (21.2%)	2064 (21.9%)	2232 (23.00%)	2465 (23.1%)	2364 (23.2%)
	6th Form	763 (6.3%)	649 (6.9%)	780 (8.0%)	888 (8.3%)	868 (8.5%)
	Higher Education	1332 (11.0%)	1131 (12.0%)	1448 (14.9%)	1595 (15.0%)	1568 (15.4%)
	Degree	1388 (11.5%)	1188 (12.6%)	1699 (17.5%)	1975 (18.5%)	1990 (19.6%)

	Foreign Qualifications	1015 (8.4%)	815 (8.7%)	744 (7.7%)	764 (7.2%)	727 (7.1%)
Age of leaving FT education	14 or never	2754 (23.3%)	1898 (21.0%)	1592 (17.0%)	1424 (13.4%)	1181 (12.1%)
	15 or 16	6121 (51.9%)	4735 (52.3%)	4962 (52.9%)	5743 (54.2%)	5314 (54.5%)
	17 or 18	1464 (12.4%)	1192 (13.2%)	1341 (14.3%)	1603 (15.1%)	1494 (15.3%)
	19 or over	1464 (12.4%)	1223 (13.5%)	1478 (15.8%)	1834 (17.3%)	1761 (18.1%)
NSSEC class	1	3526 (29.6%)	2869 (31.2%)	3103 (32.6%)	3567 (34.1%)	3423 (34.6%)
	2	1607 (13.6%)	1303 (14.2%)	1339 (14.1%)	1437 (13.7%)	1377 (13.9%)
	3	1236 (10.4%)	981 (10.7%)	1078 (11.3%)	1230 (11.8%)	1164 (11.8%)
	4	1329 (11.2%)	995 (10.8%)	1012 (10.6%)	1025 (9.8%)	960 (9.7%)
	5	4151 (35.0%)	3061 (33.2%)	2988 (31.4%)	3212 (30.7%)	2972 (30.0%)
Cigarette Consumption	0	10042 (83.0%)	8045 (85.4%)	8344 (85.5%)	9596 (87.0%)	9058 (88.3%)
	1-9	446 (3.7%)	331 (3.5%)	360 (3.7%)	349 (3.2%)	307 (3.0%)
	10-19	800 (6.6%)	527 (5.6%)	531 (5.4%)	571 (5.2%)	482 (4.7%)
	20+	807 (6.7%)	518 (5.5%)	525 (5.4%)	519 (4.7%)	413 (4.0%)

immediate recall, and estimated as 1.023 (S.E 0.007) for delayed recall, 0.693 (S.E 0.008) for verbal fluency and1.165 (S.E 0.023) for prospective memory. The loadings on the visual search factor were fixed at -1 for missed letters and was estimated as 0.978 (S.E 0.024) for correct letters and -0.259 (0.018) for letters completed. Despite the loading being <0.3 the latter was retained in order to ensure there were 3 indicators of the visual scanning factor. All loadings were statistically significant with p values <0.001.

The correlations between factors were 0.826 between attention and memory, 0.369 with attention and visual scanning and 0.354 between memory and visual scanning (appendix 5.3.3). The correlations between the attention factor and the three index scores were 0.798 for the global index, 0.824 for the memory index and 0.531 for executive function. The memory factor had correlations of 0.913 with global function, 0.949 for the memory index and 0.599 for the executive function index. In contrast, the visual scanning factor was only moderately correlated with the index scores, with correlations of 0.426 with the global index, 0.294 with the memory index and 0.464 with the executive function index.

Table 5.2 Cognitive Function Test Results by Wave

	Wave				
	1	2	3	4	5
n	12095	9272	9483	10559	9667
	mean[1]				
Verbal Fluency	19.3 (6.4)	19.8 (6.7)	20.2 (6.9)	20.7 (6.9)	20.8 (6.9)
Immediate	5.5 (1.8)	5.7 (1.8)	5.8 (1.8)	5.8 (1.8)	5.9 (1.8)
Delayed	4.0 (2.1)	4.3 (2.1)	4.5 (2.1)	4.6 (2.1)	4.6 (2.2)
Letters Correct	18.7 (6.0)	18.5 (5.9)	19.0 (5.8)	19.1 (5.5)	18.9 (5.5)
Letters Missed	5.5 (5.0)	5.0 (4.5)	4.7 (4.2)	4.5 (4.1)	4.6 (4.1)
Letters Completed	307.2 (96.0)	297.6 (93.5)	301.4 (90.2)	301.1 (87.0)	301.7 (86.1)
	Proportion incorrect[2]				
Date	2262 (19.2%)	1708 (18.5%)	1719 (18.1%)	1871 (17.8%)	1615 (16.7%)
Month	301 (2.6%)	221 (2.4%)	249 (2.6%)	235 (2.2%)	202 (2.1%)
Year	324 (2.8%)	194 (2.1%)	220 (2.3%)	229 (2.2%)	215 (2.2%)
Day	242 (2.1%)	200 (2.2%)	205 (2.2%)	200 (1.9%)	198 (2.1%)
Prospective	2364 (20.3%)	1724 (18.6%)	1550 (16.6%)	1551 (15.0%)	1297 (13.7%)

[1] Data displayed as mean (standard deviation)

[2] Data Displayed as number giving incorrect answer (percentage giving incorrect answer)

Figure 5.1 CFA Factor Structure for Cognitive Function in the English Longitudinal Study of Ageing.

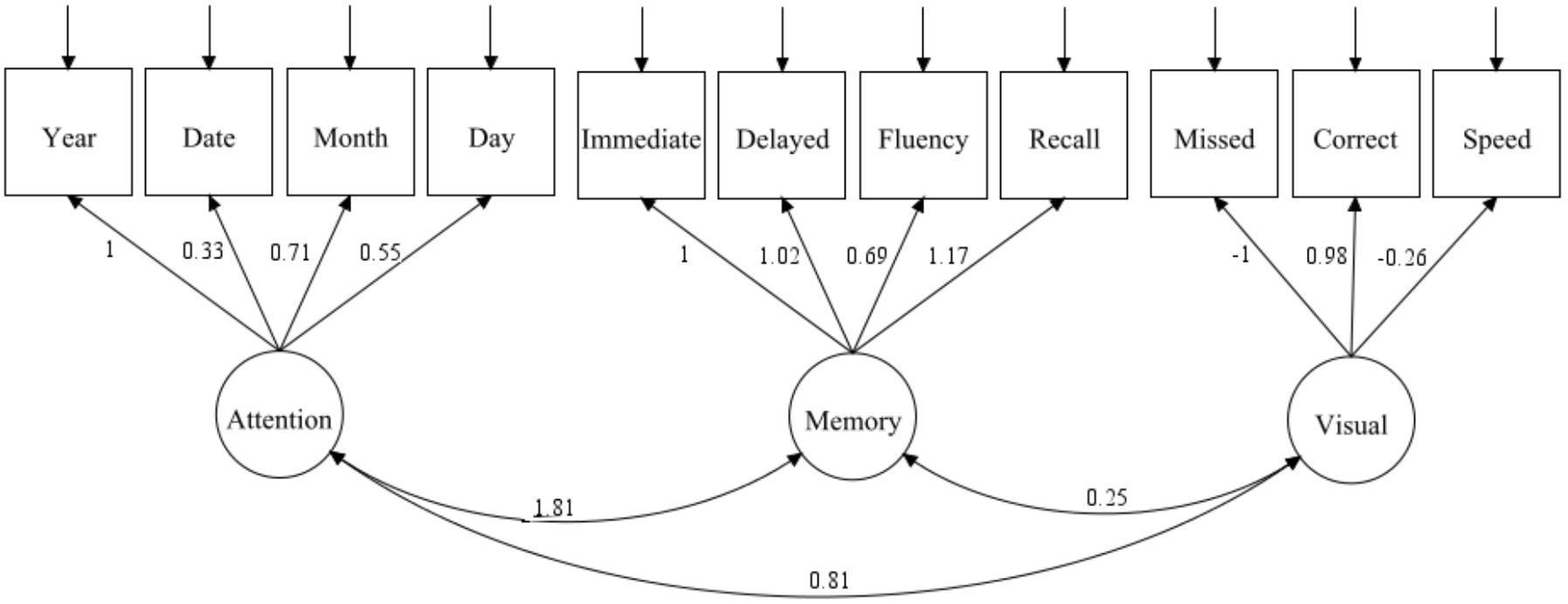


Table 5.3 Results of the multilevel models (time invariant covariates): cognitive measures regressed on covariates.

		Attention Factor	Memory Factor	Visual Search Factor	Global Index	Memory Index	Executive Index
Intercept		**-0.391 (-0.442 to -0.339)**	**-0.496 (-0.546 to -0.447)**	**-0.253 (-0.309 to -0.198)**	**-0.744 (-0.793 to -0.695)**	**-0.577 (-0.627 to -0.527)**	**-0.676 (-0.728 to -0.625)**
Gender (female)		**0.228 (0.203 to 0.252)**	**0.246 (0.220 to 0.272)**	**0.048 (0.021 to 0.075)**	**0.249 (0.223 to 0.275)**	**0.246 (0.221 to 0.271)**	**0.171 (0.143 to 0.199)**
Ethicity (non-white)		**-0.486 (-0.561 to -0.411)**	**-0.622 (-0.704 to -0.540)**	**-0.616 (-0.712 to -0.519)**	**-0.765 (-0.848 to -0.681)**	**-0.574 (-0.655 to -0.492)**	**-0.770 (-0.857 to -0.684)**
Education	Highschool	**0.271 (0.239 to 0.304)**	**0.323 (0.289 to 0.357)**	**0.179 (0.143 to 0.215)**	**0.351 (0.316 to 0.385)**	**0.279 (0.245 to 0.312)**	**0.306 (0.268 to 0.343)**
	Sixth Form	**0.328 (0.279 to 0.377)**	**0.422 (0.370 to 0.474)**	**0.191 (0.138 to 0.245)**	**0.472 (0.420 to 0.523)**	**0.362 (0.312 to 0.412)**	**0.428 (0.372 to 0.485)**
	Non-degree	**0.278 (0.237 to 0.320)**	**0.379 (0.335 to 0.422)**	**0.108 (0.063 to 0.153)**	**0.408 (0.364 to 0.452)**	**0.312 (0.270 to 0.355)**	**0.373 (0.325 to 0.421)**
	Degree	**0.400 (0.353 to 0.447)**	**0.545 (0.493 to 0.597)**	**0.177 (0.126 to 0.229)**	**0.589 (0.538 to 0.640)**	**0.453 (0.404 to 0.503)**	**0.533 (0.478 to 0.589)**
	Foreign	**0.189 (0.146 to 0.232)**	**0.210 (0.163 to 0.256)**	**0.149 (0.100 to 0.197)**	**0.215 (0.169 to 0.262)**	**0.204 (0.158 to 0.249)**	**0.155 (0.105 to 0.205)**
School Leaving Age	15	**0.147 (0.108 to 0.185)**	**0.215 (0.177 to 0.253)**	**0.199 (0.157 to 0.241)**	**0.253 (0.215 to 0.291)**	**0.142 (0.104 to 0.180)**	**0.285 (0.246 to 0.325)**
	17	**0.245 (0.197 to 0.293)**	**0.314 (0.264 to 0.364)**	**0.232 (0.179 to 0.286)**	**0.350 (0.300 to 0.401)**	**0.225 (0.176 to 0.274)**	**0.371 (0.317 to 0.425)**
	19	**0.327 (0.275 to 0.380)**	**0.426 (0.370 to 0.482)**	**0.260 (0.203 to 0.318)**	**0.475 (0.419 to 0.530)**	**0.323 (0.269 to 0.377)**	**0.487 (0.427 to 0.546)**
Social Class	1 (highest)	**0.176 (0.145 to 0.207)**	**0.196 (0.164 to 0.228)**	**0.093 (0.059 to 0.127)**	**0.196 (0.164 to 0.227)**	**0.211 (0.179 to 0.243)**	**0.133 (0.098 to 0.169)**
	2	**0.187 (0.153 to 0.220)**	**0.195 (0.159 to 0.231)**	**0.142 (0.104 to 0.180)**	**0.186 (0.150 to 0.223)**	**0.198 (0.162 to 0.233)**	**0.129 (0.089 to 0.169)**
	3	**0.090 (0.047 to 0.133)**	**0.151 (0.108 to 0.194)**	0.039 (-0.007 to 0.086)	**0.154 (0.111 to 0.196)**	**0.151 (0.109 to 0.194)**	**0.114 (0.069 to 0.160)**
	4	**0.070 (0.031 to 0.108)**	**0.088 (0.049 to 0.126)**	0.037 (-0.006 to 0.079)	**0.061 (0.022 to 0.099)**	**0.081 (0.043 to 0.119)**	0.033 (-0.009 to 0.076)
Wealth Quintile	2	0.018 (-0.010 to 0.046)	0.016 (-0.009 to 0.041)	0.013 (-0.017 to 0.043)	0.013 (-0.011 to 0.036)	0.018 (-0.008 to 0.044)	0.006 (-0.021 to 0.032)

3	0.020 (-0.009 to 0.048)	0.025 (-0.001 to 0.051)	0.021 (-0.010 to 0.051)	0.017 (-0.007 to 0.041)	**0.032 (0.006 to 0.059)**	0.003 (-0.024 to 0.031)
4	**0.058 (0.029 to 0.087)**	**0.044 (0.017 to 0.071)**	0.025 (-0.006 to 0.056)	**0.047 (0.022 to 0.072)**	**0.056 (0.028 to 0.083)**	**0.040 (0.011 to 0.069)**
5 (highest)	**0.062 (0.031 to 0.093)**	**0.053 (0.024 to 0.083)**	**0.034 (0.001 to 0.066)**	**0.054 (0.027 to 0.082)**	**0.067 (0.037 to 0.096)**	**0.036 (0.004 to 0.067)**
VPC	0.349	0.458	0.388	0.551	0.700	0.500

Table 5.4 Results of the multilevel models (time varying covariates): cognitive measures regressed on covariates.

		Attention Factor	Memory Factor	Visual Search Factor	Global Index	Memory Index	Executive Index
Age (centred)		**-0.021 (-0.022 to -0.019)**	**-0.023 (-0.024 to -0.021)**	**-0.011 (-0.013 to -0.009)**	**-0.021 (-0.022 to -0.019)**	**-0.026 (-0.027 to -0.024)**	**-0.010 (-0.012 to -0.009)**
Age^2 (centred)		**-0.001 (-0.001 to -0.001)**	**-0.001 (-0.001 to -0.001)**	**-0.000 (-0.000 to -0.000)**	**-0.001 (-0.001 to -0.001)**	**-0.001 (-0.001 to -0.001)**	**-0.001 (-0.001 to -0.001)**
Employment Status	Working	**0.070 (0.044 to 0.097)**	0.022 (-0.004 to 0.049)	**0.029 (0.000 to 0.057)**	**0.053 (0.029 to 0.077)**	0.014 (-0.011 to 0.040)	**0.074 (0.047 to 0.102)**
	Unemployed	**0.056 (0.011 to 0.101)**	0.043 (-0.002 to 0.088)	0.023 (-0.024 to 0.071)	**0.048 (0.006 to 0.089)**	0.031 (-0.013 to 0.075)	**0.052 (0.006 to 0.099)**
	Long term sick	0.044 (-0.038 to 0.125)	0.051 (-0.027 to 0.129)	0.057 (-0.031 to 0.145)	0.039 (0.000 to 0.109)	0.049 (0.000 to 0.123)	0.016 (0.000 to 0.100)
	Homemaker	**-0.142 (-0.191 to -0.092)**	**-0.102 (-0.145 to -0.058)**	-0.030 (-0.081 to 0.022)	**-0.073 (-0.115 to -0.031)**	**-0.111 (-0.156 to -0.066)**	**-0.050 (-0.096 to -0.003)**
Marital Status	Remarried	0.029 (-0.002 to 0.060)	0.028 (-0.005 to 0.060)	0.003 (-0.031 to 0.038)	**0.036 (0.005 to 0.067)**	0.023 (-0.009 to 0.055)	**0.039 (0.003 to 0.075)**
	Single	-0.026 (-0.073 to 0.020)	**-0.075 (-0.125 to -0.024)**	**-0.108 (-0.164 to 0.053)**	**-0.103 (-0.153 to -0.054)**	**-0.075 (-0.125 to -0.026)**	**-0.118 (-0.171 to -0.065)**
	Divorced	**0.038 (0.005 to 0.070)**	0.013 (-0.022 to 0.048)	-0.033 (-0.072 to 0.006)	0.027 (-0.007 to 0.061)	0.017 (-0.017 to 0.051)	**0.041 (0.003 to 0.078)**
	Widowed	0.024 (-0.010 to 0.057)	0.015 (-0.018 to 0.049)	-0.036 (-0.073 to 0.001)	-0.016 (-0.048 to 0.017)	0.013 (-0.021 to 0.046)	-0.034 (-0.068 to 0.001)
Chronic Disease	Hypertension	-0.014 (-0.034 to 0.005)	**-0.023 (-0.043 to -0.004)**	-0.011 (-0.033 to 0.011)	**-0.022 (-0.040 to -0.004)**	0.000 (-0.020 to 0.019)	**-0.048 (-0.068 to -0.028)**
	Angina	**-0.049 (-0.091 to -0.008)**	**-0.059 (-0.096 to -0.022)**	**-0.045 (-0.089 to -0.001)**	**-0.069 (-0.105 to -0.033)**	-0.027 (-0.065 to 0.012)	**-0.104 (-0.144 to -0.064)**
	MI	-0.026 (-0.076 to 0.024)	-0.032 (-0.078 to 0.013)	0.017 (-0.036 to 0.070)	**-0.048 (-0.093 to -0.004)**	-0.039 (-0.085 to 0.008)	-0.021 (-0.068 to 0.026)
	CCF	-0.099 (-0.246 to 0.049)	-0.033 (-0.162 to 0.095)	0.004 (-0.135 to 0.143)	-0.047 (-0.163 to 0.070)	-0.017 (-0.142 to 0.108)	-0.073 (-0.191 to -0.045)
	Type 2 Diabetes M.	**-0.043 (-0.081 to -0.006)**	**-0.052 (-0.088 to -0.015)**	-0.013 (-0.052 to 0.027)	**-0.041 (-0.076 to -0.006)**	**-0.080 (-0.117 to -0.044)**	0.010 (-0.027 to 0.047)

	Stroke	**-0.221 (-0.288 to -0.154)**	**-0.197 (-0.254 to -0.139)**	-0.033 (-0.096 to 0.029)	**-0.226 (-0.285 to -0.167)**	**-0.209 (-0.269 to -0.150)**	**-0.179 (-0.238 to -0.120)**
	COPD	-0.032 (-0.077 to 0.013)	-0.003 (-0.045 to 0.040)	-0.035 (-0.083 to 0.013)	-0.017 (-0.058 to 0.024)	0.013 (-0.029 to 0.055)	**-0.058 (-0.101 to -0.014)**
	Asthma	-0.015 (-0.047 to 0.018)	-0.002 (-0.034 to 0.030)	-0.018 (-0.053 to 0.017)	-0.021 (-0.052 to 0.010)	-0.005 (-0.037 to 0.026)	-0.035 (-0.070 to 0.001)
	Osteoarthritis	**0.036 (0.015 to 0.057)**	**0.029 (0.008 to 0.050)**	**0.043 (0.020 to 0.066)**	**0.033 (0.013 to 0.053)**	0.020 (-0.002 to 0.041)	**0.044 (0.022 to 0.067)**
	Osteoporosis	0.004 (-0.038 to 0.046)	0.003 (-0.038 to 0.044)	-0.008 (-0.053 to 0.038)	-0.007 (-0.047 to 0.032)	-0.015 (-0.058 to 0.027)	0.006 (-0.037 to 0.048)
	Parkinson's Disease	**-0.273 (-0.437 to -0.109)**	**-0.256 (-0.390 to -0.122)**	0.051 (-0.089 to 0.191)	**-0.246 (-0.372 to -0.121)**	**-0.240 (-0.387 to -0.093)**	**-0.304 (-0.417 to -0.192)**
Cigarette Consumption	1-9	**-0.059 (-0.110 to -0.007)**	-0.039 (-0.086 to 0.008)	-0.001 (-0.055 to 0.053)	**-0.073 (-0.118 to -0.028)**	-0.043 (-0.091 to 0.005)	**-0.106 (-0.155 to -0.057)**
	10-19	**-0.061 (-0.103 to -0.020)**	**-0.044 (-0.085 to -0.003)**	-0.031 (-0.076 to 0.015)	**-0.063 (-0.100 to -0.025)**	-0.030 (-0.070 to 0.010)	**-0.099 (-0.141 to -0.057)**
	>20	-0.033 (-0.076 -to 0.009)	-0.041 (-0.083 to 0.001)	-0.021 (-0.067 to 0.024)	**-0.054 (-0.094 to -0.014)**	-0.031 (-0.074 to 0.012)	**-0.082 (-0.128 to -0.037)**
Alcohol Consumption	Monthly	**0.056 (0.027 to 0.085)**	**0.050 (0.022 to 0.077)**	**0.037 (0.005 to 0.068)**	**0.040 (0.015 to 0.066)**	**0.036 (0.009 to 0.064)**	**0.045 (0.016 to 0.074)**
	Weekly	**0.033 (0.007 to 0.060)**	**0.047 (0.022 to 0.073)**	**0.034 (0.006 to 0.062)**	**0.038 (0.014 to 0.062)**	**0.047 (0.022 to 0.073)**	0.021 (-0.006 to 0.048)
	Daily	**0.087 (0.060 to 0.114)**	**0.111 (0.084 to 0.137)**	**0.052 (0.022 to 0.081)**	**0.111 (0.086 to 0.137)**	**0.098 (0.072 to 0.124)**	**0.102 (0.073 to 0.131)**
Depression		**-0.104 (-0.130 to -0.078)**	**-0.096 (-0.120 to -0.72)**	-0.025 (-0.052 to 0.002)	**-0.075 (-0.097 to -0.053)**	**-0.087 (-0.111 to -0.063)**	**-0.069 (-0.094 to -0.045)**

5.3.4 Regression Analyses

Substantial differences between the predictors of the factor and index scores were found in the multilevel models (table 5.3). The variance partition coefficients (VPC) for the random slope models assuming mean age and male gender were 0.349 for the attention factor, 0.458 for the memory factor, 0.388 for the letter cancelation factor, 0.551 for the global index, 0.700 for the memory index and 0.500 for the executive function index. In general, this finds that the proportion of variance at the individual level is estimated to be substantially higher when using indices. Conversely, there is much greater within individual change in cognitive factor scores. The index scores may represent an underestimation of the degree of intra-individual variability over time.

The full regression coefficients for the explanatory variables in the multilevel models are presented in table 5.3 for time invariant covariates or table 5.4 for time variant covariates and the summary of differences in coverage and p-value based inferences for the 47 regression coefficients can be seen in table 5.5. These show that inferences about the memory factor are very similar to those made using the index for memory, with only 1 parameter estimate having non-overlapping confidence intervals and only 6 parameters with different p values. The differences between the memory factor and the global and executive function indices were 1 and 3 differences in coverage and 6 and 10 differences in inferences respectively. The attention factor differs substantially more in terms of both coverage and inferences in relation to global function, memory and executive function indices (9, 2 and 9 differences in coverage and 6, 10 and 8 in inferences respectively). The visual scanning factor was very unlike any of the scores for the global, memory and executive indexes (15, 12 and 14 for coverage; 16, 13 and 16 for inferences).

This means that the difference between memory and all index scores is only what one would expect by chance in terms of the coverage of parameter estimates. The parameter estimates for attention are also most similar to the memory index with moderate differences to the global index and executive function index. Both the memory and attention factors show a moderate number of differences in the number of significant p-values when compared with all of the index scores. The visual scanning factor shows far more differences both in terms of the parameter estimates themselves and the number of those estimates found to be statistically significant. This suggests that, in an applied research setting, using index scores would fail to reflect the inclusion of the visual scanning aspect of the cognitive tasks in ELSA and its unique pattern of association with predictor variables.

Some of the key coefficients which illustrate the ways the coverage and inferences differ include that the association of the visual scanning factor with gender (being female) is substantially smaller in magnitude (0.021 to 0.075) than for the index scores (global index 0.223 to 0.275; memory index 0.221 to 0.271; executive index 0.143 to 0.199). Likewise, the association of the visual scanning factor with depression is substantially smaller and non-significant (-0.052 to 0.002) when compared to the index scores (global index -0.097 to -0.053; memory index -0.087 to -0.063; executive index -0.094 to -0.045). On the other hand, the associations with the

attention (-0.130 to -0.078) and memory factors (-0.120 to -0.072) are in line with those of the index scores. This demonstrates a risk of using the index scores that include visual scanning tasks which are not associated with depression or gender- including visual scanning in an overall index score reduces the association of the latter with gender and depression.

Table 5.5 Summary of Differences in Coverage and Inferences between factor and index scores

	Coverage Differences				
		Index Scores			
		Global	Memory	Executive	Total
Factor Scores	Attention	9	2	9	20
	Memory	1	1	3	5
	Visual	15	12	14	41
	Total	25	15	26	66
	Inference Differences				
		Global	Memory	Executive	Total
Factor Scores	Attention	6	10	8	24
	Memory	6	6	10	22
	Visual	16	13	16	45
	Total	28	29	34	91

5.4 Discussion and Conclusions

These results demonstrate the grouping of test scores into memory and executive function indices are not an accurate representation of the data structure in ELSA. Indeed, no 2-factor model is found to adequately fit the observed data. A 3-factor model of attention, working memory and visual scanning factors fits the observed data well. However, even within this 3-factor structure the variables which are grouped together do not match those of the indices.

In particular, the verbal fluency task is summed with the letter cancelation task to create the executive function index. It is in fact closer to verbal and prospective memory tasks than the letter cancelation task. The letter cancelation task is further complicated by the fact that speed on this task is associated with reduced accuracy but better performance in other areas. This is noted in the first report on the ELSA cognitive function tests, but is not accounted for in the summed index scores where speed and accuracy are summed together.(Steel et al., 2004) Indeed, this complex association with the other measures of cognitive function makes the letter cancelation task quite problematic when it comes to interpretation of the cognitive data.

Contrastingly, for inferences about which variables are associated with memory there were minimal differences between the factor and index. The memory factor and index are highly correlated and have quite similar associations with explanatory variables. The number of inferences which would differ was found to be 6 out of 47 when compared with each other. This

is likely to be due to the factor loadings being nearly 1 for immediate and delayed recall. This nearly equal weighting for those two items is similar to their simple addition in the memory index. The global index is associated with the most covariates overall which seems to represent a lack of specificity. Using the global index is likely to result in conclusions which are over generalised, so claims for an association with global functioning may be made which are actually associations with only one of the specific cognitive functions included in the global index. The loss of specificity is also present in the executive function index. No overall executive function factor of the type envisaged in the index scores is found. Rather visual scanning stands alone as a single task, separate from the other variables and significantly associated with fewer predictor variables than the executive function index, which is its closest parallel out of the 3 indices.

As well as different inferences based on conventional tests of statistical significance, there were several important differences in coverage and the substantive interpretation of the regression results. In particular, in comparison with the visual scanning factor several substantive differences are missed by the executive function index. Visual scanning shows substantially less female advantage when using factor scores, which corresponds to reported associations between male gender and comparatively high performance in visual tasks in cognitive assessment.(Marja J Aartsen, Martin, & Zimprich, 2004; de Azeredo Passos et al., 2015; Ferreira, Ferreira Santos-Galduróz, Ferri, & Fernandes Galduróz, 2014; van Hooren et al., 2007) Moreover, the visual scanning factor was less strongly associated with depression, a replication of previous findings in studies of depression and cognition which specifically use a letter cancelation task as part of the their cognitive battery.(Channon, Baker, & Robertson, 1993; Reppermund, Ising, Lucae, & Zihl, 2009; Tarbuck & Paykel, 1995) An additional difference for the visual scanning factor based on the letter cancelation task is that the decline was closer to linear, also consistent with previous research on cancellation tasks specifically.(Byrd, Touradji, Tang, & Manly, 2004; Uttl & Pilkenton-Taylor, 2001b)

Although the reduction in bias associated with the different measures was not estimated as would be done in a simulation study, this study provides a useful comparison of two common methods. In order to do so, I have provided a qualitative description of differences which may occur between the CFA method and the summed indices. In order to keep the factor analysis and summed scores as comparable as possible, I have assumed longitudinal invariance, which is one weakness of the analysis. Furthermore, it must be acknowledged that factor analysis is not the only data reduction technique which could have been used to meet the aims of this paper. However, it has the distinct advantage of being readily incorporated under the broader structural equation modelling framework which is both widely used and highly versatile.

This study illustrates how methods for combining cognitive scores determined a priori can influence the conclusions of a research paper compared with methods such as factor analysis, which are data driven. The summed index scores did not accurately reflect the structure of the data and thus the factors derived differed substantially from the indices. However, this is primarily with regards to visual scanning and executive function, as the memory index appeared

quite comparable to the results obtained using CFA. Researchers using summed scores for simplicity or comparability should carefully weigh up whether those assumptions are reasonable to make and consider checking whether the summed score has at least the same underlying structure as that revealed using data reduction methods such as factor analysis.

Chapter 6. An application of Bayesian measurement invariance to modelling cognition over time in the English Longitudinal Study of Ageing.

This paper presents the second paper which addresses the second research question of this thesis: when conventional tests for longitudinal measurement invariance based on the comparative fit index provide inconclusive results can Bayesian approximate measurement invariance be used as a suitable alternative. The secondary part to this question is whether the cognitive function latent factors in ELSA show longitudinal measurement invariance.

In a similar fashion to the first paper, this paper presents an empirical example of testing assumptions about how cognitive function is being measured in the ELSA dataset. In this case whether or not two of the factors identified in the first paper have the property of longitudinal measurement invariance. A novel use of Bayesian measurement invariance is presented in complimenting existing tests of measurement invariance when the results of conventional tests are ambiguous. In the context of the thesis as a whole, it builds upon the analysis present in the first paper in laying the foundations for the analysis in the third paper.

I conducted the analysis, drafted and revised the paper. My supervisors Prof. Chandola and Prof. Pendleton are co-authors and provided guidance on the analytic strategy and reviewed the drafts. Prof Gindo Tampubolon provided additional comments on a draft manuscript as part of an annual review for which he has my thanks.

This paper has been published. The citation is:

Williams, B. D., Chandola, T., Pendleton, N. An application of Bayesian measurement invariance to modelling cognition over time in the English Longitudinal Study of Ageing. *International Journal of Methods in Psychiatric Research.* 27(e1749), 1-8. https://doi.org/10.1002/mpr.1749

An application of Bayesian measurement invariance to modelling cognition over time in the English Longitudinal Study of Ageing.

Dr. Benjamin David Williams[1], BM,BS, MSc

Prof. Tarani Chandola[1], PhD

Prof. Neil Pendleton[2], MD

1. Cathie Marsh Institute for Social Research, Humanities Bridgeford Street Building, University of Manchester, M13 9PL

2. Institute of Brain, Behaviour and Mental Health, Stopford Building, University of Manchester

Abstract

Objectives

Recommended cut-off criteria for testing measurement invariance (MI) using the comparative fit index (CFI) vary between -0.002 and -0.01. We compared CFI results with those obtained using Bayesian approximate measurement invariance (MI) for cognitive function.

Methods

We used cognitive function data from waves 1-5 of the English Longitudinal Study of Ageing (ELSA; wave 1 n=11951), a nationally representative sample of English adults aged ≥50. We tested for longitudinal invariance using CFI and approximate MI (prior for a difference between intercepts/loadings ~N(0,0.01)) in an attention factor (orientation to date, day, week and month) and a memory factor (immediate and delayed recall, verbal fluency and a prospective memory task).

Results

Conventional CFI criteria found strong invariance for the attention factor (CFI + 0.002) but either weak or strong invariance for the memory factor (CFI -0.004). The approximate MI results also supported strong measurement invariance for attention but found 9/20 intercepts or thresholds were non-invariant for the memory factor. This supports weak rather than strong invariance.

Conclusions

Within ELSA, the attention factor is suitable for longitudinal analysis but not the memory factor. More generally, in situations where the appropriate CFI criteria for invariance are unclear, Bayesian approximate MI could alternatively be used.

Key words/phrases: approximate measurement invariance, statistics, old age, cognitive function, ELSA.

6.1 Introduction

Measurement invariance (MI) is an often-underappreciated problem in psychiatric research. Whilst some outcomes in psychiatry are discrete and directly observable, many are impossible to observe directly. Cognitive function is an example of this and is the focus of this study. Latent variable analysis is one common method used to combine multiple measures into a single measure of an underlying concept of interest. However, a frequently occurring problem when using latent variables longitudinally is that the association between the observed variables and the unobserved latent variable changes over time.

In tests of cognition, performance on the tests will be determined by the individual's ability in the target function (say working memory) but also their ability in a range of other cognitive and physical functions (such as attention and hearing). The demands on other functions will differ between tests. Each additional function utilized in performing each individual task may be differentially affected by ageing, disease or setting (McAvinue et al., 2012; Wiegand et al., 2014). As well as different rates of change secondary to cognitive or physical processes, the size of practice effects may also vary between different tests of the same cognitive function (Calamia et al., 2012). Any of these may change the strength of the association between the individual cognitive tests and the latent cognitive function over time. In factor analysis, this manifests as a change in factor loading or intercept and is known as measurement invariance (van de Schoot et al., 2012).

MI has been discussed extensively elsewhere and has been identified as a problem in longitudinal studies of cognitive function since at least the late 1980s and early 1990s (Horn & McArdle, 1992; Schaie et al., 1989). With some notable exceptions, population and clinical research on cognitive function has had a tendency to overlook this issue with a preference for using summed scores, the measurement properties of which are often not examined (Blankson & McArdle, 2013; McArdle et al., 2007; Wicherts, 2016). If this issue is ignored it biases estimates of change in cognitive function over time towards the direction of the change in latent intercept or varying effects for a change in factor loading (Ferrer et al., 2008; Horn & McArdle, 1992; van de Schoot et al., 2013; Wicherts, 2016; Widaman et al., 2011). For example, practice effects would be expected to increase the intercept leading to an over-estimation of cognitive ability at follow-up visits and thus underestimation of decrease over time (Wicherts & Dolan, 2010). Alternatively, a decrease in factor loading due to increased sensory impairment over time weakening the association between measurable and latent cognitive function could lead to overestimation of cognitive function for low scorers and underestimation for high scorers as time progresses (Wicherts, 2016).

6.1.1 Conventional Measurement Invariance

Underlying a set of k (n=0,…,k) continuous observed variables c that have been measured, there is a latent variable η (B. O. Muthén & Asparouhov, 2013; van de Schoot et al., 2013). If they are measured in individual i at time t the measurement part is:

1) $c_{ikt} = v_{kt} + \lambda_{kt}\eta_{it} + \varepsilon_{ikt}$

Here c_{ikt} is the observed value of variable *k* at time *t* in individual *i*, v_{kt} is the intercept for variable *k* at time *t*, λ_{kt} is the loading for variable *k* at time *t,* η_{it} is the value of the latent variable at time *t* for the variable *k* and ε_{ikt} is the error for individual *i* at time *t* for observed variable *k*. This model assumes independence amongst the c's conditional on the factor, that the residuals are uncorrelated with the factors and the errors are normally distributed with a mean of 0. The factor metric is usually set by fixing $\lambda = 1$ for one observed variable.

A linear growth curve for factor scores (the structural model) is:

2) $\eta_{it} = \eta_{0i} + x_t\eta_{1i} + \zeta_{it}$

Here η_{0i} is the intercept of the latent variable, η_{1i} is the slope growth factor and ζ_{it} the time and individual specific residual. The binary case is a straightforward extension of equation 1 and if a probit link function is assumed then the latent variable is assumed to follow a continuous distribution and the structural model is unchanged. Otherwise, it should be noted that the intercept v_{kt} is replaced with the threshold $-\tau_{tl}$ (Muthen, 2004).

For continuous variables, the specification of MI consists of: *i*) the same variables load onto the same factors at each time point (the same vector of c_{ikt} for each η_{it}), *ii*) the factor loadings are equal at each time point ($\lambda_{k1} = \lambda_{k2} = \cdots = \lambda_{kt}$) , *iii*) intercepts are equal at each time point ($v_{k1} = v_{k2} = \cdots = v_{kt}$) and *iv*) residual variances fixed across time ($\varepsilon_{ik1} = \varepsilon_{ik2} = \cdots = \varepsilon_{ikt}$)(van de Schoot et al., 2012; Widaman et al., 2011). If only *i* holds, this is known as configural invariance, *i-ii* weak invariance, *i-iii* strong invariance and *i-iv* strict invariance. In the case of binary observed variables the second stage, weak factorial invariance is skipped because the item probability curve is influenced simultaneously by loading and intercept (L. K. Muthén & Muthén, 2014).

Strong invariance needs to be established in order to compare latent means over time (Ferrer et al., 2008; Widaman et al., 2011). If this assumption does not hold, then mean differences over time in a latent variable of cognitive function cannot be clearly attributed to change in true cognitive function because the scale of the dependent variable has changed. Additionally, tests of MI are sensitive to the choice of indicator variable (Shi et al., 2017). This is fixed at 1 for every time point and is used to establish the scale of the latent variable, so at least one factor loading must be assumed to be invariant. Thus, the choice of a non-invariant reference variable can make decisions regarding MI significantly more difficult.

The standard approach to testing MI is to sequentially compare each level of increasing invariance using the chi-squared test of model fit. With large sample sizes this is a strict test and strong factorial invariance over time may be rejected even in robust longitudinal studies of cognitive ageing (Blankson & McArdle, 2013; B. O. Muthén & Asparouhov, 2013). Therefore with large sample sizes alternative fit indices, in particular the comparative fit index (CFI), are frequently used instead (Cheung & Rensvold, 2002; Meade & Bauer, 2007). However, recommendations for the change in CFI which establishes MI differ between studies and these

recommendations vary between 0.01 and 0.002 depending on author and the number of factors, indicators of those factors and groups/occasions used.(F. F. Chen, 2007; Cheung & Rensvold, 2002; Meade & Bauer, 2007; Meade et al., 2008; Short, 2014).

Whilst these methods can be used to identify invariance, they are not informative about which parameters are invariant. To do this, one can either relax each equality constraint in turn or use modification indices, which give a measure of the improvement in model fit which would result from relaxing certain modelling assumptions. Relaxing each equality constraint sequentially means allowing each loading or intercept at each time point individually to be different to the same intercept or loading at all other time points. The change in model fit can then be assessed. This is laborious and random variation can lead to different invariance solutions being identified depending upon the order in which the constraints are relaxed (R. C. MacCallum et al., 1992; Bengt Muthén & Asparouhov, 2012). Modification indices are limited in application because they are only validated for two samples or time points (B. O. Muthén & Asparouhov, 2013).

6.1.2 Bayesian Measurement Invariance

Bayesian structural equation modelling (BSEM) has introduced the concept of approximate MI to take account of multiple small or moderate non-invariances in loadings, intercepts or thresholds. Additionally, it provides a one-step method of identifying which parameters are invariant (B. O. Muthén & Asparouhov, 2013; van de Schoot et al., 2013; Verhagen & Fox, 2013). The basic effect of approximate MI is that instead of requiring that all loadings be exactly equal, they are instead 'tethered' so that they do not have to be exactly equal but are allowed to differ only by a substantively unimportant amount.

As described above the conventional condition which must be met for strong factorial invariance (and therefore the ability to measure change in latent means over time) is that for each of the observed variables $\lambda_{k1} = \lambda_{k2} = \cdots = \lambda_{kt}$ and $v_{k1} = v_{k2} = \cdots = v_{kt}$. Let ђ be the difference between λ's such that $\lambda_{k1} - \lambda_{k2} = ђ_{k12}$, $\lambda_{k2} - \lambda_{k3} = ђ_{k23}$ and $\lambda_{k1} - \lambda_{k3} = ђ_{k13}$. Also let be the difference between v's such that $\nu_{k1} - \nu_{k2} = и_{k12}$, $\nu_{k2} - \nu_{k3} = и_{k23}$ and $\nu_{k1} - \nu_{k3} = и_{k13}$. The conventional frequentist assumption of strong invariance can then be defined in Bayesian terms as the strongly informative priors of $ђ_{kXX} \sim N(0,0)$ and $и_{kXX} \sim N(0,0)$ (B. O. Muthén & Asparouhov, 2013).

Given that, from a Bayesian perspective, the factor loadings and intercepts are random variables, the assumption of 0 variance is difficult to envisage in this framework. With approximate MI this is instead relaxed slightly to a still strong but more plausible informative prior with 0 mean and small variance such as $ђ_{kXX} \sim N(0,0.01)$ and $и_{kXX} \sim N(0,0.01)$. One reason for preferring the Bayesian approach in this situation is that this assumption of exact equality is relatively unrealistic in a number of situations due to issues such as random variation across many time-points, attrition or practice effects (Blankson & McArdle, 2013; Putnick & Bornstein, 2016). The researcher can decide a priori how long to make the tether by specifying an

appropriate prior for the difference between loadings or intercepts over time. The size of the prior variance therefore sets the length of the tether and formalises the degree of invariance which is allowable.

The difference at each time-point is tested to see whether it is statistically significantly different from the mean of the loadings at all time-points. This tells you if any of the loadings have broken the tether and show a degree of non-invariance beyond that believed to be unimportant by the researcher. Additionally, this overcomes the problems in identifying the truly non-invariant parameters caused by fixing one indicator's loadings at 1 for all time-points. Using the Bayesian approximate MI approach one need only fix single loading for a single observed indicator at a single time-point to 1 (B. O. Muthén & Asparouhov, 2013; Xu & Green, 2015).

An alternative frequentist approach to testing for MI is running models with and without MI to see if the results are conflicting (Widaman et al., 2011). With this approach an, often informal, decision is made about the degree of conflict in the results which is acceptable before MI is rejected. This decision is made using substantive prior subject knowledge and implicitly includes an assumption about the acceptable degree of invariance. The Bayesian approach formalises the same substantive knowledge into the prior which can therefore be specifically tested.

When assessing for longitudinal invariance in the English Longitudinal Study of Ageing we encountered several of the aforementioned problems with conventional MI testing. The sample size is large, therefore the chi-squared test likely to be overly conservative (F. F. Chen, 2007; Cheung & Rensvold, 2002; Steptoe et al., 2013). Additionally, as we will show, different cut-offs for the CFI produced different conclusions. Moreover, the kind of invariance we were expecting was of multiple small deviations rather than few large deviations from invariance. Given the number of variables and time points in use, relaxing each constraint in turn would be both laborious and highly prone to the risk of error due to chance. For these reasons, we applied Bayesian approximate MI to test whether the conclusions about the level of MI drawn from this method differed to those drawn from the chi-squared test and CFI rules.

Our primary research questions were; firstly, in ELSA's cognitive function battery is there longitudinal measurement invariance for an attention and a memory factor? Secondly, can Bayesian approximate measurement invariance be used to identify measurement invariance (or the lack thereof) in situations where CFI has an uncertain result?

6.2 Methods

6.2.1 Participants and Procedure

ELSA has been described in detail elsewhere (Steptoe et al., 2013). The study sample was drawn from participants in Health Survey for England (HSE) years 1998, 1999 and 2001 who were born before 1st March 1952 and living in a private household or those in their households

who were new partners or ≤50. We used the ELSA core sample which was nationally representative of the age specific English population at the time of recruitment. Data is collected in biennial sweeps by interview in the participants homes. For this analysis data from waves 1 (2002) to 5 (2010) were utilised because the core cognitive battery was consistent through this time.

Response rates at each wave were 70% at wave 1, 82% at wave 2, 73% at wave 3, 74% at wave 4 and 80% in wave 5 (Steptoe et al., 2013). After the exclusion of extreme values (see below) final sample sizes at each wave were n=11951 in wave 1, n=9313 in wave 2, n=7850 in wave 3, n=6911 in wave 4 and n=6535 in wave 5.

6.2.2 Cognitive Measures

The cognitive tests were performed by computer assisted interview. Orientation to time was assessed by asking the participant to name the day, year, month and date. To assess immediate and delayed verbal recall 10 common words were played to participants (Steel et al., 2004). Immediate recall is assessed straight away and delayed recall of the word list was tested after the other cognitive tests were undertaken (this also serve as a distraction technique). The word lists used were randomly assigned and a standardised recording was used for all participants.

The prospective memory task required participants to remember to write their initials in the top corner of a page they were handed. Participants were prompted if they did not complete the actions spontaneously. A binary variable was used for remembering the correct action (either prompted or spontaneous). Semantic (category) fluency was assessed by asking participants to name as many animals as they can in 1 minute. All the non-binary variables were transformed to z-scores for the purpose of inclusion in the factor structure.

6.2.3 Statistical Analysis

Initially, extreme values with regards to the relationship between cognitive variables were identified by regressing each cognitive function variables in turn on all the others at each wave. The standardised residuals and leverage statistics were then compared and regression re-run with the exclusion of influential cases to see if the results were substantively different (Institute for Digital Research and Education, 2009). Only for month and year did the exclusion of high residual cases appear to make a substantive difference.

Due to the rarity of giving the incorrect response to year and month, almost all incorrect answers were considered extreme values by conventional recommendations. However, analysis of those cases with particularly high residual values identified a subset of cases who were incorrect on either year or month but achieved average or better results on all other tests. These cases were felt likely to represent either errors in recording or single item inattention. In

total 97 and 85 measurement occasions were excluded for year and month respectively, meaning approximately 35 in total per wave of data collection. Other missing data were considered missing at random, which is as a property of the Bayesian estimation (Q. Chen & Ibrahim, 2014). Research on how missingness affects longitudinal invariance has only been implemented in a single study using full information maximum likelihood and, whilst a topic warranting further investigation, is beyond the scope of this analysis (Sterba, 2017).

Initial EFA and CFA assuming invariance were performed as part of an earlier study currently in submission. Two of the factors from this, attention (loaded onto by orientation questions) and memory (loaded onto by immediate and delayed recall, prospective memory and verbal fluency) were used. The model was specified using CFA with configural invariance and modification indices checked to see if there was any need to make additional modifications beyond the basic factor structure (L. K. Muthén & Muthén, 2014). This identified that allowing residual covariance over time in verbal fluency and within factor covariance's for immediate and delayed recall resulted in substantially improved model fit. This improved model was then tested using the chi-squared test and CFI for MI.

Next the Bayesian approximate MI model was specified. A prior variance of N~(0,0.01) for all differences between loadings, intercepts and thresholds at each wave with the mean across all waves was used. The MPlus default non-informative priors were used for all other model parameters (Bengt Muthén & Asparouhov, 2011). The conclusions about the level of MI in the data was then compared between frequentist chi-squared test and CFI and Bayesian approximate MI.

The primary analysis was run for all ages in the ELSA data. Sensitivity analyses were run using age bands to check for one possible source of longitudinal non-invariance. Though there was slightly less non-invariance for older participants, and slightly more for younger participants, the overall pattern of results was very similar for all ages. Due to this, they are not presented here.

The data was edited using Stata version 12 and the structural equation modelling performed using MPlus version 7.0 (L. K. Muthén & Muthén, 2014; StataCorp, 2011). MCMC estimation was utilised with the MPlus default Gibbs sampler and convergence criterion, 105000 iterations (of which the first 55250 are burn-in) and no thinning (Bengt Muthén & Asparouhov, 2011).

6.3 Results

The participants at wave 1 were 55.7% female, had a mean age of 64.2 and 2.8% of the sample were of non-white ethnicity (table 6.1). The large minority of participants were retired (47.7%) and the majority of the rest of the sample worked as either employed (28.1%) or self-employed (5.7%). Most participants were married (56.2% first marriage; 11.1% remarried). The modal educational attainment was no-qualifications (41.7%) with 11.5% having attained a degree. There was a bimodal distribution of social class with the largest groups being class 5 (manual

Table 6.1 English Longitudinal Study of Ageing participant demographics at wave 1.

Variable	Total	Percentage
Age	64.2	S.D 11.1
Female	6676	55.7%
Non-White	328	2.8%
Employment Status		
Retired	5715	47.7%
Employed	3370	28.1%
Self Employed	687	5.7%
Unemployed	123	1.0%
Permanent Sick	783	6.5%
Homemaker	1173	9.8%
Other	131	1.1%
Marital Status		
1st Marriage	6741	56.2%
Remarried	1331	11.1%
Single	658	5.5%
Divorced/Separated	1256	10.5%
Widowed	2003	16.7%
NS-SEC Social Class		
1 Professional	3487	29.7%
2	1596	13.6%
3	1223	10.4%
4	1320	11.3%
5 Manual	4112	35.0%
Highest Qualification		
No Qualifications	4986	41.7%
High School	2522	21.1%
6th Form	748	6.3%
Non-degree higher Ed.	1317	11.0%
Degree	1370	11.5%
Foreign Qualification	1014	8.5%

and routine occupations; 35.0%) and the second largest class 1 (managerial and professional roles; 29.7%).

Cognitive function data was available for 11630 of 11951 participants at wave 1, 9066 of 9313 at wave 2, 7659 of 7850 at wave 3, 6656 of 6911 at wave 4 and 6216 of 6535 at wave 5. The results showed a slight improvement in the memory factor tasks over time (table 6.2). Mean immediate word recall was 5.4 (s.d 1.8) in wave 1 and 5.7 (s.d 1.9) in wave 5. Mean delayed recall was 4.0 (s.d 1.8) in wave 1 and 4.4 (s.d 2.2) in wave 5. The number of participants

correctly remembering the prospective memory task was 79.3% in wave 1 and 85.8% in wave 5. The orientation to time tasks were stable over time. The proportion of participants in wave 1 correctly identifying the year was 97.4%, date 80.6%, month 97.6% and day 97.9%. This was not dissimilar to wave 5 where the proportion of participants correctly identifying the year 97.3%, date 81.7%, month 97.8% and day 97.5%.

The factors structure was modelled based on previous EFA and CFA. The attention factor was comprised of orientation to year, date, month and day. The memory factor was comprised of immediate and delayed recall, verbal fluency and prospective memory. In the memory factor the residual variances of verbal fluency were allowed to correlate over time and the residual variances between immediate and delayed recall were allowed to correlate at each time point, reflecting the more similar nature of these tasks.

Table 6.2 Mean or proportion of correct responses for each cognitive task in waves 1 to 5 for ELSA core participants.

	Wave				
	1	2	3	4	5
n	11630	9066	7659	6656	6535
	mean†				
Immediate	5.4 (1.8)	5.7 (1.8)	5.7 (1.8)	5.7 (1.8)	5.7 (1.9)
Delayed	4.0 (1.8)	4.3 (2.1)	4.4 (2.2)	4.4 (2.2)	4.4 (2.2)
Verbal Fluency	19.3 (6.4)	19.8 (6.6)	19.8 (6.8)	20.2 (7.0)	20.2 (7.0)
	Proportion correct				
Year	97.4%	98.1%	97.5%	97.4%	97.3%
Date	80.6%	81.4%	80.8%	80.8%	81.7%
Month	97.6%	97.7%	97.2%	97.7%	97.8%
Day	97.9%	97.8%	97.6%	97.7%	97.5%
Prospective	79.6%	81.3%	82.9%	84.3%	85.8%

† Results displayed as mean (standard deviation).

When testing for longitudinal invariance using the χ^2 test, all levels of MI were rejected for both the attention and memory factor with a p values of <0.001 (table 6.3). We then compared the CFI between the models. The model with configural invariance both attention and memory had a CFI of 0.976. Setting strong invariance for the attention factor (not weak invariance as all 4 indicator variables were binary) actually improved the CFI to 0.978. Weak invariance for the memory factor also increased the CFI to 0.978 whereas strong invariance reduced it to 0.972. Strong invariance for both factors resulted in a CFI of 0.973 showing that the misfit induced by strong invariance in the memory factor was not compensated for by the improvement in fit from strong invariance in the attention factor.

Table 6.3 Model fit tests for conventional frequentist CFA.

	χ^2 test versus baseline model	χ^2 test versus less restrictive model	CFI	CFI Difference versus less restrictive model
All configural	<0.001	-	0.976	-
Attention Strong	<0.001	0.002	0.978	0.002
Memory Weak	<0.001	<0.001	0.978	0.002
Memory Strong	<0.001	<0.001	0.972	-0.006
Both Strong	<0.001	<0.001	0.973	0.001 or -0.005

Therefore, using the CFI criteria, longitudinal MI was not rejected for the orientation factor by any criteria. On the other hand, the decrease in CFI of 0.006 in the change between weak and strong invariance for the memory factor falls between different recommendations from different studies.

The approximate MI results found that the there was one parameter in the attention factor which showed a minor degree of non-invariance (table 6.4); the 1st wave loading for recall of the day (0.326) which is 0.036 less than the mean loading across all waves (0.362); this was a statistically significant difference based on the 95% credible interval. This is not likely to have a substantively important impact on the results of longitudinal analysis.

For the memory factor, there is only one non-invariant loading; the wave 4 verbal fluency loading (0.927) which is 0.029 greater than the mean across all waves (0.898). However, 9 of the 20 intercepts and thresholds are non-invariant. For immediate recall the 2nd (0.052 above the mean), 3rd (0.032 above the mean) and 5th (0.057 below the mean) loadings show significant non-invariance. For delayed recall the 2nd (0.036 above the mean) and 3rd (0.053 above the mean) occasions are non-invariant. In verbal fluency the 2nd measurement occasion is estimated as being 0.009 above the mean. For prospective memory task the threshold on the

1st occasion is 0.069 above the mean and the 5th occasion is 0.054 below the mean for all measurement occasions.

This means that across multiple time points there are different expected values of the indicator variables for memory when the mean of the factor is zero. Whilst the individual differences are small, the number of non-invariant parameters suggest that the latent mean at one time point is not directly comparable with another. It may be better not to use the memory factor for longitudinal analysis but to analyse the individual memory tasks separately. These results support the use of the stricter CFI criteria for MI in this case.

6.4 Discussion

When analysing cognitive function data from ELSA we encountered a situation where different recommendations for using the CFI to establish MI led to different conclusions. We sought to use approximate MI to provide an alternative method of deciding which level of MI to accept or reject. In this case, the approximate MI approach identified small but significant non-invariance in the loadings of the memory and attention factors which was not identified by the use of CFI (which did not reject weak invariance). However, the degree of invariance in loadings which was identified using approximate MI but missed by CFI was relatively trivial. This suggests that the assumptions of strong longitudinal MI in the attention factor and weak invariance in the memory factor are plausible.

The main source of longitudinal non-invariance was not in the factor loadings, but the intercepts of the memory factor. This led to strong invariance to being rejected by both the stricter CFI criteria and approximate MI. This is particularly important because strong invariance is required to compare latent means over time and therefore necessary for longitudinal analysis. However, using alternative CFI cut-off rules for MI would have led the authors to a different conclusion about the presence or absence of strong invariance for the memory factor. Using a cut-off of -0.01 such as that recommended by Chen (2007) or Cheung and Rensvold (2002) would have suggested not rejecting strong MI. By the more stringent recommendations of Meade, Johnson and Brady (2008) of -0.002, strong but not weak invariance would have been rejected. Moreover, as discussed by Short (2014) the truly suitable cut-off for CFI may be different again when using the specific number of time-points and observed variables available. Using approximate MI revealed that there was a high proportion of non-invariant intercepts and thresholds for the memory factor caused by multiple small deviations from non-invariance. This would have been difficult to accurately identify in a step-wise fashion using a frequentist estimator.

If using factor analysis or another data reduction method, including sum scores, then ignoring this MI would have resulted in bias in the estimation of the memory factor latent mean (B. O. Muthén & Asparouhov, 2013; van de Schoot et al., 2013). In our results, the wave 2 and 3 memory factor latent means would have been over-estimated due to increases in the immediate and delayed recall intercepts. Wave 5 would have been underestimated because of decreases

Table 6.4. Factor Loadings using Bayesian Approximate Measurement Invariance for both factors at each time point.

Item		Approximate Invariance Factor Loadings (0.01 prior variance)					
		Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Mean
Orientation Factor	Year	1	1.021	1.034	1.029	1.045	1.026
	Date	0.278	0.295	0.298	0.264	0.264	0.28
	Month	0.51	0.54	0.555	0.513	0.516	0.527
	Day	**0.326***	0.387	0.35	0.369	0.377	0.362
Memory Factor	Immediate Recall	1	1.013	1.025	1.021	0.985	1.009
	Delayed Recall	1.08	1.102	1.101	1.082	1.064	1.086
	Verbal Fluency	0.856	0.897	0.896	**0.927***	0.914	0.898
	Prospective Mem.	0.88	0.934	0.875	0.911	0.855	0.891
		Approximate Invariance Intercepts† and Thresholds‡ (0.01 prior variance)					
		Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Mean
Orientation Factor	Year‡	-5.887	-5.9	-5.892	-5.898	-5.89	-5.893
	Date‡	-1.095	-1.099	-1.088	-1.040	-1.062	-1.077
	Month‡	-3.483	-3.446	-3.463	-3.457	-3.476	-3.465
	Day‡	-2.796	-2.853	-2.754	-2.847	-2.811	-2.812
Memory Factor	Immediate Recall†	-0.013	**0.053***	**0.033***	-0.013	**-0.055***	0.001
	Delayed Recall†	**-0.014***	**0.069***	**0.086***	0.037	-0.013	0.033
	Verbal Fluency†	-0.011	**0.009***	-0.013	-0.013	-0.063	-0.018
	Prospective Memory‡	**-0.963***	-1.013	-1.034	-1.064	**-1.086***	-1.032

*Statistically significant using 95% credible interval.

in the immediate recall intercept and prospective memory threshold. These effects would result in bias in both the estimation of both the rate and shape of the latent growth curve.

The non-invariance in the memory factor seems to be a combination of several isolated deviations and a linked increase in immediate and delayed recall in waves 2 and 3. It is possible that the non-invariance seen at wave 2 and 3 for the intercepts of immediate and delayed recall represents unequal practice effects in the indicators of this factor. The reduction in waves 4 and 5 may represent fatiguing practice effects, an initial practice effect followed by more rapid decline in performance on those tasks or practice effects for the other indicators catching up relative to the recall tasks (Calamia et al., 2012). Whether Bayesian MI could be used to detect non-uniform practice effects may be an avenue for further research.

The present study has the strength of using data from a high-quality multidisciplinary survey with a large sample size. This study is relevant to researchers with a wide variety of longitudinal research questions relating to phenomena which cannot be directly observed. It is especially pertinent for those researching common mental health disorders who wish to utilise the richness of multidisciplinary surveys but lack a validated measure (previously demonstrated to be invariance over time in the population of interest) of the construct of interest, as with cognition in the first 5 waves of ELSA. Here, difficulties due to a large number of small non-invariances are particularly likely to occur. Furthermore, the specific number of groups or time-points may not to have been covered in previous simulation studies, thus the most appropriate cut-off for the CFI or other fit indices not known.

The large sample size to some extent does cover one of the potential weaknesses of BSEM in that it can be highly sensitive to prior specifications (Depaoli, Yang, & Felt, 2017; van Erp, Mulder, & Oberski, 2018). Informative priors for one parameter have the property of inducing implicit priors for other covariant parameters in a fashion which is difficult to predict and manage (Robert C. MacCallum, Edwards, & Cai, 2012). It should be noted that if there is insufficient data to generate informative priors, or they are not desired for substantive reasons, then BSEM estimates with non-informative priors tend to converge with maximum likelihood estimates (Helm, Castro-Schilo, & Oravecz, 2017; S. Y. Lee, Song, & Tang, 2007). Whilst the single step identification of non-invariant parameters offers significant theoretical advantages over methods such as modification indices, in terms of the reduction of the capitalization of chance in inferences, there are few simulation studies to confirm this finding (Robert C. MacCallum et al., 2012).

BSEM retains the common practical problems of many types of Bayesians analysis in terms of computational intensity, challenges with assessing convergence and unfamiliarity to many users. This is particularly the case in comparison to approaches to identifying MI such as straightforwardly comparing parameters between models which assume or don't assume MI. Whilst this approach may provide rapid answers in some clear-cut situations, in many cases even if an acceptable difference between estimates is pre-specified (for example 5% or 10%) the results are borderline (Flora & Curran, 2004). This approach will also be model specific if the

target of interest is a predictor of growth or a distal outcome and the additional information about invariant parameters will not be obtained, unlike with approximate MI.

Approximate measurement invariance, whilst not a panacea, is designed to handle multiple small invariances and its power to detect non-invariance is not known to be affected by changing the number of groups or occasions being compared, which provides substantial flexibility. As such it may be useful for future researchers to consider when testing the measurement properties of their instruments in longitudinal research. With regards to ELSA specifically we find an attention factor which essentially shows strong measurement invariance over time but only weak invariance for a memory factor. Whilst the degree of non-invariance was relatively small, it was on a large number of parameters and therefore researchers may wish to either avoid using the memory factor for longitudinal research or accommodate the non-invariance using approximate or partial measurement invariance.

Chapter 7. Does the association between cognition and education differ between cognitively stable and cognitively declining older adults?

This paper presents the third paper which addresses the third research question of this thesis: is education associated with cognitive maintenance and does this association vary by latent class of decline?

This paper presents an analysis using growth mixture modelling to allow the effect of education on cognitive function to vary dependent on underlying trajectory. This makes it possible to test the hypothesis that different mechanisms of cognitive reserve are active in different states. Simultaneously included in the structural equation model is a class dependent informative missingness model which is implemented in a way not used in studies of cognitive ageing previously.

In the context of the thesis as a whole this analysis builds upon the previous two papers efforts to understanding the structure underlying the cognitive tests in the English Longitudinal Study of Ageing. It is because of the results of the previous two papers that individual tests, rather than simple combinations or factor scores, were used in this analysis. It leads into the fourth paper by examining the effects of social exposures on cognitive maintenance, whilst making use of newer analytical methods which can overcome certain limitations in the existing literature.

I conducted the analysis, drafted and revised the paper. My supervisors Prof. Chandola and Prof. Pendleton provided guidance on the analytic strategy and reviewed the drafts. Prof Gindo Tampubolon provided additional comments on a draft manuscript as part of an annual review.

This paper has not been published and is not under review at the time of submission.

Does the association between cognition and education differ between cognitively stable and cognitively declining older adults?

Dr. Benjamin David Williams[1], BM,BS, MSc

Prof. Neil Pendleton[2], MD

Prof. Tarani Chandola[1], PhD

1. Cathie Marsh Institute for Social Research, University of Manchester

2. Institute of Brain, Behaviour and Mental Health, Stopford Building, University of Manchester

Abstract

Objectives

Education is associated with baseline cognitive performance in older adults, but it's association with maintenance of cognitive function is less clear. I hypothesised that education is associated with different types of active cognitive reserve in later life depending on whether an individual is cognitively stable or declining, which may explain conflicting previous results.

Methods

I used data on n=5642 adults aged >60 from the English Longitudinal Study of Ageing (ELSA) over 5 waves (8 years). Verbal fluency and immediate recall were used as the observed measures of cognitive function. I used a Bayesian growth mixture model for each outcome which including a model for informative missingness. The effect of education on rate of change was allowed to vary by latent class in order to test whether latent class moderated the association between education and cognitive maintenance.

Results

For recall, 91.5% (n=5164) of ELSA participants were in a stable class and 8.5% (n=478) in a declining class. For fluency 90.0% (n=4907) were in a stable class and 10.0% (n=561) were in a declining class. Educational attainment was associated with improved baseline performance for both verbal fluency and recall. In the declining classes, educational attainment was not associated with rate of change for either verbal fluency or immediate recall. In the stable classes, the only significant association with rate of change was for verbal fluency amongst those with higher (an extra 0.05 to 0.38 words per 2 years) or degree level education (an extra 0.04 to 0.42 words per 2 years) compared to those with no formal qualifications.

Conclusions

Educational attainment had a strong effect on baseline performance, but little effect on cognitive maintenance overall. There was some evidence of active cognitive reserve for verbal fluency in the most highly educated individuals.

7.1 Introduction

Education in childhood and early adulthood (for simplicity referred to throughout as ‘education’) is thought to be one of the most important sources of cognitive reserve, defined here as the degree of disease or age related change which can be tolerated by the brain before impairment becomes apparent.(Barulli & Stern, 2013) This is demonstrated by the consistent finding that educational attainment is associated with a reduced risk of a clinical diagnosis of dementia.(Beydoun et al., 2014; X. Meng & D’Arcy, 2012) However, the exact nature of this relationship, and more specifically the relationship between education as a source of cognitive reserve and cognitive maintenance over time, has been more contested. Several theoretical concepts of reserve have been developed, each leading to contrasting hypotheses about the relationship which would be observed between education and cognitive maintenance.

Stern proposes that, broadly speaking theories, about reserve can be divided into brain reserve and cognitive reserve.(Stern, 2012) Cognitive reserve has also been divided into active and passive reserve depending on whether moderation of cognitive decline is seen or not.(Stern, 2002) The brain reserve hypothesis, which is largely synonymous with passive reserve, sees reserve principally as the degree of pathology that can accumulate before clinical expression of impairment.(Barulli & Stern, 2013; Stern, 2002) The amount of brain reserve is quantified principally in structural and anatomical terms such as brain volume, synaptic density or white matter tract integrity.(Arnold et al., 2013; Negash et al., 2013; Teipel et al., 2009) Under this hypothesis higher levels of education would lead to improved baseline performance but no change in cognitive maintenance.(Stern, 2002)

On the other hand, cognitive reserve is hypothesised to moderate decline over time. It can be sub-divided into the concepts of neural cognitive reserve and neural compensation reserve. These generate opposing predictions about the effect of reserve of cognitive maintenance. The former broadly overlaps with the theory of active reserve. It predicts that, as individuals with higher levels of education have greater network capacity and efficiency, cognitive decline in those individuals will be slower.(Boller, Mellah, Ducharme-Laliberté, & Belleville, 2017) Contrastingly, in the compensation hypothesis secondary networks are recruited as primary networks fail, but then as the secondary networks also fail decline rapidly accelerates.(Barulli & Stern, 2013) This hypothesis predicts that individuals with high levels of education should experience more rapid cognitive decline once a certain threshold in the severity and extent of brain pathology is reached.

Earlier studies and the systematic reviews based on those studies largely found evidence that education improved cognitive maintenance which supports the neural cognitive reserve hypothesis.(Valenzuela & Sachdev, 2006) More recently, Lenehan and colleagues questioned the findings of these studies on the basis of methodological limitations present in those studies.(Lenehan et al., 2015) These included the frequent use of only two time-points, the practice of regressing on baseline function and the use of simpler statistical methods less robust to missing data or uneven spacing of measurement occasions such as repeated measures ANOVA and simple linear regression.(Lenehan et al., 2015) They found that, in later analyses of

cohort studies with three or more measurement occasions, there was typically no association between education and rate of decline and therefore little evidence to support the idea that education contributes to cognitive maintenance.

Other authors have found that, especially in individuals affected by dementia rather than those with age-related cognitive decline, education is associated with more rapid cognitive decline.(Barulli & Stern, 2013; X. Meng & D'Arcy, 2012) This may be due to a common end point where pathology overwhelms compensatory reserve. This difference between healthy old age and dementia found in epidemiological or clinical studies is supported by evidence from functional magnetic resonance imaging studies. These have shown that different mechanisms of compensation appear to be utilised depending on disease status.(Colangeli et al., 2016) It is not known whether this effect would also be seen in those with pre-clinical dementia pathology.

This leads to one limitation of many of the major analyses of education's association with cognition in the general population of older adults. This is that they have implicitly made the assumption that all the individuals in that sample are from the same population and therefore share the same underlying trajectory (or random effects around this). This would include examples such as the Victoria Longitudinal Study, the ARIC Neurocognitive study or the ACTIVE study.(Gottesman et al., 2014; Tucker-Drob, Johnson, & Jones, 2009; Zahodne et al., 2011a) However, it is likely that, even in these ostensibly health samples, individuals were actually drawn from at least two populations. Those with a pre-clinical dementia pathology (a high burden of tau, amyloid and/or vascular pathology in the absence of functional impairment) and those without.(Braak & Del Tredici, 2015; Riley et al., 2011) Of course there are a great number of diseases which lead to cognitive impairment, but most are too rare to identify as separate subgroups using statistical means. They will be considered as a single group here.

The presence of this sub-population creates both problems and opportunities for understanding the relationship between cognitive reserve and education. In analyses which assume a single homogenous population when there are two or more sub-populations, the estimated longitudinal change will be biased away from both true trajectories. In particular, it is possible that in these samples an association between longitudinal change and education has been obscured for those with a declining trajectory suggestive of pre-clinical dementia pathology, as suggested by imaging studies of people with dementia. However, if one can identify a latent sub-population with more rapid decline in cognitive function from a population sample then this gives the opportunity to examine the influence of social and environmental factors on cognitive maintenance in earlier stages of disease than clinical samples. In this case, to see if the more rapid rate of decline in cognitive function seen in highly education persons with diagnosed dementia is also seen in those with probable pre-clinical dementia.

Several studies to date have used growth mixture modelling (GMM) to address the issue of rates of change in cognitive function in population samples with latent sub-populations.(Hayden et al., 2011; Marioni et al., 2014; Muniz-Terrera et al., 2010; Olaya et al., 2017; Pietrzak et al., 2014; Royall et al., 2014; Small & Bäckman, 2007) Hayden et al and Pietrzak et al. combined genotypic and clinicopathological data with latent class analysis and found that membership of a

rapidly declining latent class was strongly associated with higher relative risk of amyloid beta pathology and apolipoprotein ε4 carrier status.(Hayden et al., 2011; Pietrzak et al., 2014) This is important because it suggests growth mixture modelling is able to accurately identify those in a pre-clinical disease state who have more rapid cognitive decline.(Riley et al., 2011)

Of the studies which have used GMM specifically to analyse cognitive trajectories and education, most have used education as a predictor of class membership.(Hayden et al., 2011; Marioni et al., 2014; Pietrzak et al., 2014; Royall et al., 2014; Small & Bäckman, 2007) The results of these studies have been conflicting with some finding a strong association between class of cognitive trajectory and education, some a weak association and others finding none.

It is important to note that the, often not explicitly stated, assumption underlying these models is that education has a direct influence on the process underlying the latent classes. Whilst there is likely to be several sub-classes of cognitive function within healthy ageing, the single most important determinant of class in this case must surely be disease. This will include a range of pathologies, but in a population study by far the most substantial are tau/amyloid pathology and vascular disease which may occur individually or comorbidly.(Santos et al., 2017) If we assume that the latent class structure is driven principally by disease status, then these studies presuppose that education affects not only observable cognitive and functional status but also unobserved (in most population studies) disease status. However, clinicopathological studies have in general found that education is not associated with the degree of pathological change observed post-mortem.(Brayne et al., 2010; Koepsell et al., 2008; Roe et al., 2007; Serrano-pozo et al., 2013) In the theoretical model used in this analysis the mechanism underlying the latent classes in cognitive function is driven by the presence or absence pathology, the effect of which is moderated by education. This approach is more consistent with the findings of the clinicopathological studies and the long-acknowledged need to adjust cognitive scores for education when diagnosing dementia (see figure 7.1 for a generalized version of my model).(Kittner et al., 1986; Uhlmann & Larson, 1991)

Terrera et al. have previously utilised a similar theoretical model and examined the association between education and decline within class using the mini-mental state exam (MMSE).(Folstein et al., 1975; Muniz-Terrera et al., 2010) They found 2 sharply declining classes and 1 high performance group with very slight decline over time. A lower level of education was associated with more rapid decline in the high-performance class but not in either of the two sharp decline classes. However, the MMSE is known to have a strong ceiling effect which can conceal change in high performance groups.

I sought to test whether the association between cognitive decline and education was moderated by latent class to test the hypothesis that different mechanisms of cognitive reserve are utilised in different states. To do this I utilised data from the English Longitudinal Study of Ageing (ELSA), a large multi-disciplinary study of ageing, to permit the identification of relatively small subgroups, and which includes cognitive measures with relatively minimal ceiling or floor effects.

7.2 Methods

7.2.1 Participants and Procedure

ELSA has been described in detail previously.(Steptoe et al., 2013) The study sample was drawn from participants in Health Survey for England (HSE) years 1998, 1999 and 2001 who were born before 1st March 1952 and living in a private household or those in their households who were new partners or ≤50 years old. This initial sample was nationally representative of the age specific English population. Additional recruitment was undertaken in waves 3 and 4 but these participants were not included in the current analysis due to the informative dropout modelling used. Data was collected in biennial sweeps by interview in the participants homes. Data from waves 1 (2002) to 5 (2010) were utilised because the core cognitive battery was kept consistent through this time.

I limited the sample to individuals born before 1940, who were therefore aged 61 or above at the first wave. This eliminates at least one source of cohort effect (prenatal exposure to World War 2 rationing) and restricts the analysis to those more likely to show a greater degree of cognitive decline. I also excluded those individuals with foreign qualifications because this is a highly heterogenous category from which no meaningful substantive conclusions could be drawn about my research question.

The size of the full sample eligible for analysis was 5643 at wave 1, 103 were dropped due to missing data on gender, ethnicity or education. 1256 participants dropped out or died between waves 1 and 2, 765 between waves 2 and 3, 634 between waves 3 and 4 and 533 between waves 4 and 5.

For verbal fluency the latent trajectory class structure was initially driven by extreme outliers. Therefore, for this analysis I identified outliers by regressing each measurement occasion on the previous one. Results with standardised residuals >2.9 or <-2.9 were checked individually. They were coded as missing from the analysis if the results were inconsistent with the other results for those individuals (for example a 0 despite normal performance on other tests or a result far higher or lower than the results both before and after that occasion). This removed 81 observations at wave 2, 111 observations from wave 3, 99 observations from wave 4 and 91 observations from wave 5.

7.2.2 Cognitive Measures

The cognitive tests were performed by computer assisted interview. Of the cognitive measures in ELSA orientation to time, delayed recall and prospective memory task were not utilised due to strong ceiling or floor effects.(Marmot et al., 2014) Immediate recall and verbal (semantic) fluency were utilised because the floor effects were much weaker (appendices 7.1 and 7.2). To

assess immediate recall 10 common words were played to participants which they were asked to repeat immediately after presentation. The word lists used were randomly assigned and a standardised recording was used for all participants. Semantic (category) fluency was assessed by asking participants to name as many animals as they could in 1 minute.

7.2.3 Education and Covariates

Educational attainment was recorded as no formal qualifications (reference category in all analyses), high school completion (O-levels or equivalent), 6th form completion (A-levels of equivalent), non-degree level higher education and undergraduate degree or above. The first non-missing value was used. Age at baseline was centred for the analysis and wave of study was used as the metric of time for all analyses. Gender and ethnicity (white and non-white) are treated as binary.

7.2.4 Statistical Analysis

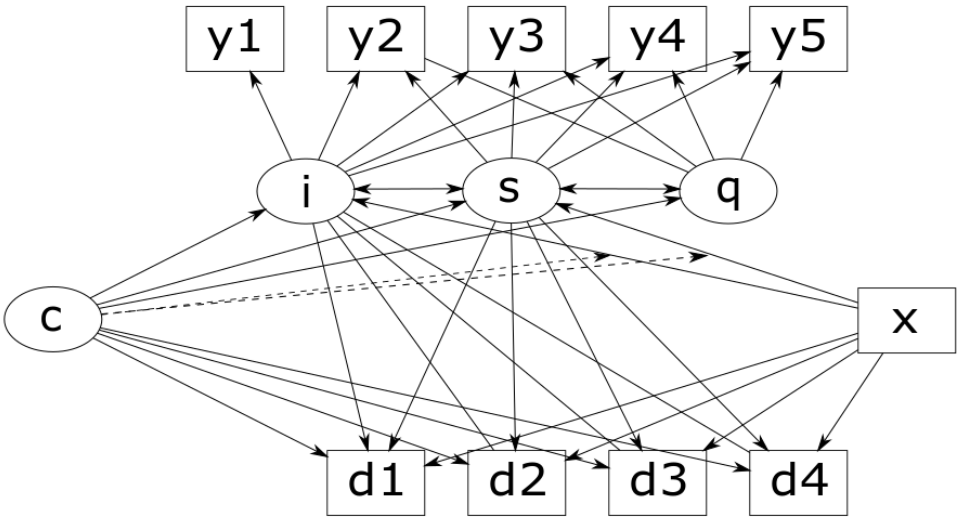
Growth mixture models were estimated for immediate recall and verbal fluency separately using Bayesians MCMC estimation. Each model was tested to see if quadratic or cubic curves improved model fit. Intercept and slope were allowed to vary by class. The latent intercept and slope were regressed on all covariates and the effect of each covariate was allowed to vary by class. This tests the principle research question that the association between education and cognitive maintenance may vary by latent class.

Missing data were handled using a not missing at random (NMAR) Beunckens model in which missingness at each wave is dependent on observed covariates, the latent variables for the outcome (intercept and growth factors) and latent class.(Beunckens, Molenberghs, Verbeke, & Mallinckrodt, 2008) Separately modelling dropout and death lead to model under-identification and problems with convergence. They were therefore modelled jointly. After assessing model fit the immediate recall model utilised only the intercept to predict missingness whereas in the verbal fluency model both intercept and slope independently predicted missingness. The effect of each variable on missingness was fixed to be equal across all waves. In neither the fluency nor the recall model did allowing the regression of missingness on covariates to vary by class improve model fit. See figure 7.1 for the generalised representation of the structural equation model used.

The number of latent classes was assessed using Rousseau and Mengersen's over-fitting method.(Nasserinejad, Rosmalen, De Kort, & Lesaffre, 2017; Rousseau & Mengersen, 2011) For this 6 classes were initially specified with a pre-specified cut-off for the posterior mode of the number of classes larger than 0.05 and the Dirichlet prior for class proportion of (5,3) for fluency and (4,3) for recall (half the number of free parameters between classes). In both cases this method identified 2 classes meeting the pre-specified cut-off. Model fit for the 2 class models was then assessed using the Bayesian posterior predictive p value (PPPV), entropy and

whether the classes were substantively coherent. Weakly informative priors were used for all regression coefficients, missingness thresholds and class specific latent intercepts and means. An analysis adjusting for potential pre-education confounders (parental smoking, family structure in childhood and parental occupation) there was little difference in the substantive conclusions drawn from the analysis, therefore these results are not presented. The data were edited using Stata version 12 and the structural equation modelling performed using MPlus version 7.0.(L. K. Muthén & Muthén, 2014; StataCorp, 2011) MCMC estimation was utilised with the MPlus default Gibbs sampler and convergence criterion, 120000 and 200000 iterations were used for fluency and recall respectively of which the first 50% are treated as burn-in with no thinning.(Bengt Muthén & Asparouhov, 2011)

Figure 7.1 Generalised structural equation model for the verbal fluency and immediate recall growth mixture models including informative missingness modelling and the effect of covariates moderated by latent class membership.



C = latent class of change over time; X = all time invariant covariates; Y1-y5 = the outcome at waves 1 through 5; D1-d4 = whether participants died or dropped out at each wave 2-5; I = latent intercept; S = latent linear rate of change; Q = latent quadratic rate of change

7.3 Results

The participant demographics can be seen in table 7.1 separated by when those individuals left the study. Across all waves the mean age at baseline was 73.2, 53.7% (n=3028) of participants were female and 2.3% (129) were a non-white ethnicity. 58.3% (n=3291) had no formal educational qualifications, 19.0% (n=1073) had high school certificates, 4.5% (n=253) had sixth form equivalent qualifications, 10.0% (n=566) had non-degree higher educational qualifications

and 8.2% (460) had an undergraduate degree or higher. As the study progressed the remaining participants were younger, more likely to be female, more likely to be white, and less likely to have no formal educational qualifications.

For a 1 class fluency model the PPPV was 0.078 (-12.8 to 78.5 credible interval for a difference between the observed and replicated chi-squared values), for the 2 class model the PPPV was 0.148 (-28.6 to 93.6) and for a 3-class model 0.041 (-9.3 to 152.5). The entropy for the 2-class fluency model was 0.904 and for the 3-class model it was 0.617. For the 1 class immediate recall model the PPPV was 0.78 (-12.3 to 78.2), for the 2-class model 0.081 (-18.5 to 108.3) and for the 3-class model 0.011 (12.1 to 177.1). Entropy for the 2 class recall model was 0.872 and 0.826 for the 3-class model.

Table 7.1 Demographics by wave and dropout numbers for ELSA core participants utilised in the growth mixture model analysis.

	Wave 1 n=5643	Wave 2 n=4387	Wave 3 n=3622	Wave 4 n=2988	Wave 5 n=2455
Mean Age at W1 (s.d)	73.2 (7.2)	72.3 (6.8)	71.8 (6.5)	71.2 (6.1)	70.6 (5.7)
Verbal fluency (s.d)	17.5 (6.0)	18.0 (6.3)	17.9 (6.4)	18.0 (6.7)	17.9 (6.8)
Recall (s.d)	4.9 (1.8)	5.1 (1.8)	5.0 (1.8)	5.0 (1.8)	5.0 (1.9)
Female	3028 (53.7%)	2379 (54.2%)	1977 (54.6%)	1647 (55.1%)	1358 (55.3%)
Non-white ethnicity	129 (2.3%)	83 (1.9%)	65 (1.8%)	45 (1.5%)	36 (1.5%)
No formal qualification	3291 (58.3%)	2419 (55.1%)	1932 (53.3%)	1507 (50.4%)	1184 (48.2%)
Highschool Education	1073 (19.0%)	889 (20.3%)	751 (20.7%)	656 (22.0%)	562 (22.9%)
Sixth Form Education	253 (4.5%)	210 (4.8%)	179 (4.9%)	156 (5.2%)	129 (5.3%)
Non-degree Higher Ed.	566 (10.0%)	474 (10.8%)	412 (11.4%)	364 (12.2%)	318 (13.0%)
Degree Level Education	460 (8.2%)	395 (9.0%)	348 (9.6%)	305 (10.2%)	262 (10.7%)
Dropout	-	1256	765	634	533

The final class proportions based on the estimated posterior probabilities for the fluency model were 90.0% (n=4907) of the sample in a stable class and 10.0% (n=561) in a rapid decline class. The final class proportions based on the estimated posterior probabilities for the recall model were 91.5% (n=5164) of the sample in a stable class and 8.5% (n=478) in a declining class.

The coefficients from the GMMs are shown in table 7.2 for verbal fluency and 7.3 for recall. In the analysis of verbal fluency in the first (stable or probable healthy cognitive ageing) latent class the latent intercept in number of animals named per minute was 16.59 (95% Credible Interval 16.17 to 16.93). The linear rate of change was -0.21 (95% CI -0.39 - -0.05) with a quadratic rate of change of -0.12 (95% CI -0.16 to -0.09), indicating gradual decline in fluency over time which accelerates slightly. In the second (rapid decline or probable disease) latent class the latent intercept in number of animals named per minute was 14.75 (95% CI 13.01 to 18.03). This shows a wider confidence interval due to the smaller number of participants in this class, but with the lower point estimate suggesting some decline prior to initial data collection as one would expect. The linear rate of change was -2.20 (95% CI -3.36 to -1.21) with a quadratic rate of change of -0.59 (95% CI -0.87 to -0.31). This latent class is estimated to decline in fluency at an initial rate around 10 times as fast as the healthy ageing class, and this decline also accelerates more rapidly.

Increasing age at baseline was associated with decreased baseline fluency score in both classes (Class 1 -0.24 - -0.1; class 2 -0.53 - -0.19), but with linear decline only in class 1 (class 1 -0.06 - -0.05; class 2 -0.01 – 0.02). In neither class was female gender associated with either intercept (class 1 -0.64 – 0.08; class 2 -1.91 – 2.42) or rate of change (class 1 -0.04 – 0.18; class 2 -1.44 – 0.69). Non-white ethnicity was associated with a decreased baseline score in class 1 but not class 2 (class 1 -7.38 - -5.41; class 2 -4.70 – 1.31), however the number of individuals with non-white ethnicity in class 2 was only 2.3% (n=13; table 7.4). In neither class was non-white ethnicity associated with rate of decline (class 1 -0.62 – 0.13; class 2 -2.74 – 2.16).

The association of education with latent intercept in the stable fluency class showed essentially a dose response relationship with greater education associated with higher baseline fluency scores (high school 1.75 – 2.62; Sixth Form 2.17 – 3.62; Higher Education 3.56 – 4.09; Degree 3.93 – 5.12). In the stable fluency class lower levels of educational attainment were associated with no difference in change over time (high school -0.10 – 0.13; Sixth form -0.06 – 0.41) but higher levels of education attainment were associated with a modest decrease in rate of decline (higher education 0.05 – 0.38; degree 0.04-0.42). The difference in rate of decline was approximately equivalent to being 4 years younger for both higher education and degree education.

In the rapid decline fluency class level of education was significantly associated with intercept only with high school education but not sixth form, non-degree higher or degree level educational attainment (high school 0.28 – 4.95; Sixth Form -1.62 – 3.64; Higher Education -0.50 – 3.90; Degree -1.48 – 5.59). Although mostly non-significant, the point estimates showed

a similar dose-response pattern to that seen in the stable function class. In the rapid decline fluency class no level of educational attainment was associated with rate of decline (high school -1.71 – 0.96; Sixth Form -1.07 – 1.98; Higher Education -1.65 – 0.63; Degree -0.19 – 0.53).

Table 7.2. Estimates of verbal fluency latent class specific parameters for baseline score and rates of change, with the effect of covariates on these.

	Class 1 - Stable			Class 2 - Decline		
		95% Credible Interval			95% Credible Interval	
	Coeff.	Lower	Upper	Coeff.	Lower	Upper
Baseline Score	**16.619**	**16.244**	**16.992**	**14.99**	**12.986**	**18.312**
Linear Rate of Change	**-0.189**	**-0.347**	**-0.023**	**-1.539**	**-2.499**	**-0.654**
Quadratic Rate of Change	**-0.098**	**-0.134**	**-0.066**	**-0.717**	**-0.956**	**-0.431**
Date of birth						
Baseline Score	**0.209**	**0.181**	**0.233**	**0.334**	**0.167**	**0.452**
Rate of Decline	**0.045**	**0.036**	**0.053**	-0.08	-0.216	0.005
Female Gender						
Baseline Score	-0.255	-0.589	0.09	0.878	-1.48	2.887
Rate of Decline	0.062	-0.047	0.162	-0.456	-1.822	0.237
Non-white ethnicity						
Baseline Score	**-6.371**	**-7.328**	**-5.427**	-1.987	-4.995	1.272
Rate of Decline	-0.093	-0.446	0.263	-0.67	-3.103	1.952
Highschool Education†						
Baseline Score	**2.137**	**1.72**	**2.545**	**3.182**	**0.956**	**5.426**
Rate of Decline	0.05	-0.075	0.177	-0.105	-1.102	0.949
Sixth Form Education†						
Baseline Score	**2.946**	**2.252**	**3.636**	0.695	-2.223	3.271
Rate of Decline	0.118	-0.101	0.344	0.496	-0.77	1.674
Higher Non-degree Education†						
Baseline Score	**3.404**	**2.873**	**3.929**	2.077	-0.379	4.55
Rate of Decline	**0.221**	**0.062**	**0.382**	0.011	-1.215	0.82
Degree Level Education†						
Baseline Score	**4.564**	**3.979**	**5.117**	1.793	-1.428	5.099
Rate of Decline	**0.172**	**0.002**	**0.344**	-0.104	-1.776	1.09

†Baseline no formal educational qualifications

The model for immediate recall in the healthy cognitive ageing latent class estimated a latent intercept in number of words correctly recalled of 4.27 (95% CI 4.19 – 4.36). The linear rate of change was 0.16 (95% CI 0.03 – 0.29), the quadratic rate of change was -0.19 (95% CI -0.28 - -0.11) and the cubic rate of change was 0.03 (95% CI 0.02 – 0.05). As can be seen in figure 7.2 this results in very slight mean decline of around 0.3 words over the 8 years of follow-up. In the rapid decline or probable disease latent class the latent intercept was 4.09 (95% CI 3.67 – 4.50) words recalled correctly. The linear rate of change was 0.35 (95% CI 0.08 – 0.60), the quadratic rate of change was -0.13 (95% CI -0.05 – 0.33) and the cubic rate of change was -0.10 (95% CI -0.14 - -0.06). As can be seen in figure 7.2 this results in a decline which is only minimally evident during the first couple of waves but then declines very sharply in later waves. Compared to the more gradual decline throughout follow-up for verbal fluency this may be evidence either of reduced sensitivity to change or a greater practice effect masking initial declines.

Increasing age at baseline was associated with decreased immediate recall intercept in class 1 but not class 2 (class 1 -0.09 - -0.08; class 2 -0.04 – 0.03), although the association in real terms is extremely small even in class 1. Age at baseline was associated with linear decline in recall in both classes (class 1 -0.01 - -0.00; class 2 -0.03 – 0.00), though again the magnitude of the association is very slight. Female gender was associated with intercept in class 1 only (class 1 0.33 – 0.51; class 2 -0.02 – 0.90) though the confidence intervals overlap for both estimates. Female gender was associated with rate of change in class 2 only (class 1 -0.02 – 0.04; class 2 -0.27 – -0.13). As with verbal fluency, non-white ethnicity was associated with a decreased baseline recall score in class 1 but not class 2 (class 1 -1.17 - -0.50; class 2 -3.33 – 0.76). In neither class was non-white ethnicity associated with rate of decline (class 1 -0.24 – 0.02; class 2 -0.42 – 0.13).

The association with education for the latent intercept of recall in the stable class showed a dose response relationship, similar to fluency, with greater education associated with higher baseline fluency scores (high school 0.56 – 0.80; Sixth Form 0.62 – 1.03; Higher Education 0.86 – 1.16; Degree 1.29 – 1.61). In the stable class no level of educational attainment was associated with change over time (high school -0.02 – 0.06; Sixth form -0.05 – 0.09; higher education -0.07 – 0.02; degree -0.09 -0.01). In the rapid decline class level of education was associated with intercept for sixth form and degree level attainment but not high school or non-degree higher education (High school -0.21 – 0.95; Sixth Form 0.26 – 2.26; Higher Education -0.76 – 0.85; Degree 0.73 – 2.27). In the rapid decline class, no level of educational attainment was associated with rate of decline (High school -0.10 – 0.25; Sixth Form -0.34 – 0.43; Higher Education -0.31 – 0.22; Degree -0.50 – 0.12).

Table 7.3. Estimates of immediate recall latent class specific parameters for baseline score and rates of change, with the effect of covariates on these.

	Class 1 - Stable			Class 2 - Decline		
		95% Credible Interval			95% Credible Interval	
	Coeff.	Lower	Upper	Coeff.	Lower	Upper
Baseline Score	**4.305**	**4.235**	**4.398**	**4.094**	**3.674**	**4.502**
Linear Rate of Change	**0.168**	**0.083**	**0.256**	**0.346**	**0.077**	**0.601**
Quadratic Rate of Change	**-0.189**	**-0.249**	**-0.136**	0.134	-0.049	0.334
Cubic Rate of Change	**0.033**	**0.023**	**0.043**	**-0.101**	**-0.143**	**-0.063**
Date of birth						
Baseline Score	**0.083**	**0.077**	**0.089**	-0.006	-0.038	0.027
Rate of Decline	**0.006**	**0.003**	**0.008**	**0.02**	**0.009**	**0.03**
Female Gender						
Baseline Score	**0.421**	**0.335**	**0.509**	**0.517**	**0.063**	**0.979**
Rate of Decline	0.013	-0.017	0.043	**-0.288**	**-0.443**	**-0.14**
Non-white ethnicity						
Baseline Score	**-0.838**	**-1.155**	**-0.53**	-1.001	-3.056	0.899
Rate of Decline	-0.087	-0.2	0.031	0.103	-0.679	1.049
Highschool Education†						
Baseline Score	**0.666**	**0.555**	**0.781**	0.417	-0.18	0.989
Rate of Decline	0.016	-0.022	0.054	0.076	-0.104	0.26
Sixth Form Education†						
Baseline Score	**0.831**	**0.635**	**1.02**	**1.09**	**0.182**	**1.996**
Rate of Decline	0.013	-0.052	0.08	-0.1	-0.433	0.192
Higher Non-degree Education†						
Baseline Score	**1.017**	**0.87**	**1.158**	-0.062	-0.836	0.777
Rate of Decline	-0.027	-0.073	0.021	-0.013	-0.279	0.225
Degree Level Education†						
Baseline Score	**1.43**	**1.283**	**1.579**	**1.585**	**0.842**	**2.313**
Rate of Decline	-0.04	-0.089	0.01	-0.26	-0.556	0.032

†Baseline no formal educational qualifications

Table 7.4 Descriptive statistics by latent class for verbal fluency and recall.

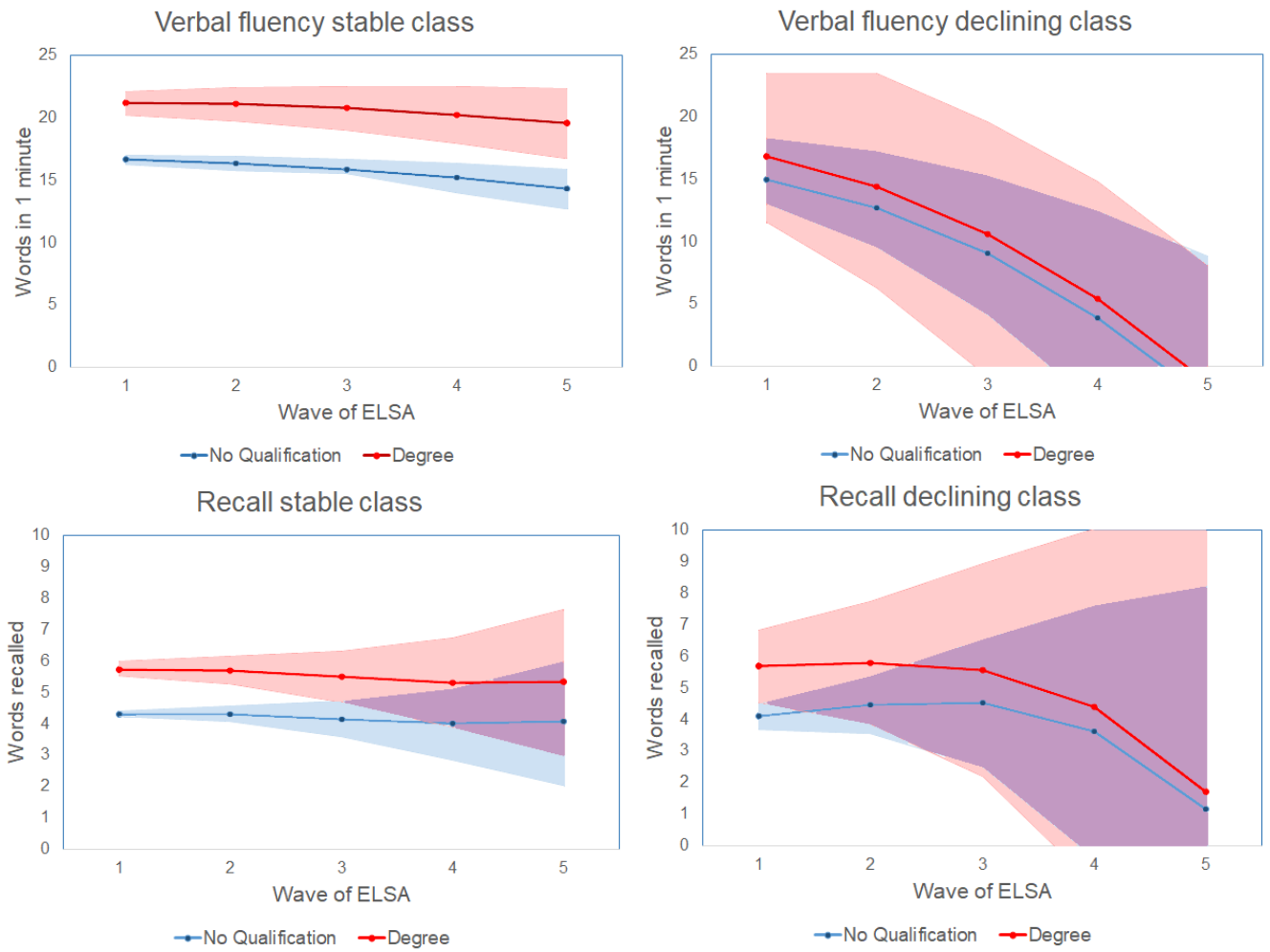
		Verbal Fluency		Recall	
		1. Stable	2. Decline	1. Stable	2. Decline
Education	Gender	53.6%	54.0%	53.5%	54.9%
	Non-white	2.4%	2.3%	2.3%	2.3%
	No formal qual.	58.4%	60.4%	58.1%	60.0%
	High school	19.2%	17.4%	19.1%	19.3%
	Sixth Form	4.5%	4.6%	4.5%	4.2%
	Higher non-degree	9.8%	10.2%	10.1%	9.2%
	Degree	8.1%	7.4%	8.2%	7.3%
Mean Cognitive Score (animals named or words recalled)	Age	0.07	-0.74	0.027	-0.3
	Wave 1	17.6	15.7	4.9	4.6
	Wave 2	18.2	14.2	5.1	5
	Wave 3	18.2	11.3	5	5.2
	Wave 4	18.3	7.4	5	4.6
	Wave 5	18.2	4.3	5.2	2.2

7.4 Discussion

For both cognitive measures, educational attainment was not associated with rate of decline in cognitive function amongst the decline/disease class. The somewhat higher number of individuals classified as decliners/disease in the fluency model is likely to indicate a greater degree of sensitivity to change for this measure. Nonetheless the numbers classified as decliners is close enough between the fluency and recall models to support the idea that they are identifying the same class. Previous studies using GMM have demonstrated that this is likely to represent a class with pre-clinical dementia.(Hayden et al., 2011; Pietrzak et al., 2014)

In the stable function class there was a suggestion that those with the highest levels of education did have a slightly slower rate of decline for verbal fluency only. However, this association was not seen for immediate recall and the magnitude of the association was small. It is possible that the finer level of differentiation possible with verbal fluency makes it more sensitive to small effects such as this or alternatively that education has an effect one of the cognitive processes involved in verbal fluency but not recall. Therefore my results in general provide support for the brain or passive reserve hypothesis in both health and early disease, with suggestion of a degree of neural compensation reserve for cognitively health older adults with the highest levels of educational attainment.(Lenehan et al., 2015) Incidentally, the non-linear decline observed in the decline/disease class supports the explanation that observation of faster decline in more highly educated individuals with dementia is likely to be due to being at a more advanced disease state by the time they reach the threshold for diagnosis.(X. Meng & D'Arcy, 2012)

Figure 7.2 Estimated mean curves of the two latent classes for verbal fluency and immediate recall comparing the effect of no formal qualifications (baseline) and degree level education.



My findings have both agreement and contrast with those of Muniz-Terrera *et al.* who used a very similar statistical methodology.(Muniz-Terrera et al., 2010) They found that lowers levels of education predicted faster decline in their high performance class. Due to the strong ceiling effect of the MMSE, it was possible that it may not to detected decline in those with higher levels of education. This would artificially make it appear that those with lower levels of education declined more rapidly. However, in my analysis using verbal fluency, which does not have a ceiling effect, I observed a broadly similar pattern. Those with higher or degree level education showed slightly better cognitive maintenance than those with no formal qualifications or secondary school level educational attainment. This is consistent with education providing a degree of neural reserve in older adults with stable, likely healthy, cognition. It makes it unlikely that the previous finding was due to the ceiling effect alone. However, the effect size is very small and likely to be of limited practical importance. The contrast with the analysis of Muniz-Terrera *et al.* is that I did not observe the same effect for immediate recall of a 10 word list. This difference stems from the choice of using specific cognitive tests rather than a measure of global cognitive function. My analysis suggests that the association between education and cognitive maintenance is domain specific. It is quite plausible that educational attainment would be more closely associated with verbal skills than memory alone.(McDaniel & Einstein, 2011)

The finding of 2 latent classes contrasts with several previous studies of latent classes of cognitive function not using the MMSE, including previous analysis of ELSA data, which have found 3 or 4 classes.(Hayden et al., 2011; Olaya et al., 2017) However, not all previous studies have found more classes. Of those studies with 3 or 4 classes the pattern is frequently of 2-3 essentially ordinal classes and 1 qualitatively different class (for example the 3 stable classes with differing baseline performance and 1 declining class as seen in Olaya et al.).(Olaya et al., 2017; Royall et al., 2014) Allowing the effects of education and age to vary within class rather than predict class in my model is likely to have resulted in the loss of the ordinal classes (whose differences in baseline performance are instead modelled as a function of education within class) and the preservation of the qualitatively different trajectories.

Though a limited range of measures was available, the relative lack of ceiling and floor effects in the measures used is an important strength of this analysis. Another important strength of this study is the fact that education is used to predict change and not class which, for the reasons described in the introduction, I feel more accurately translates the results of post-mortem studies into a population setting. Additionally, there are many strengths of the ELSA dataset in general including, but far from limited to, the large sample size, a representative general population sample and good duration of follow-up.(Steptoe et al., 2013) The large number of individuals with lower levels of educational attainment is of special relevance to this study. This is important not only for statistical power but also reduces the chances of results being related to sampling bias. The inclusion of an informative missingness model is also an important strength of the analysis as it relaxes the missing at random assumption for at least one missingness process.

That said, one weakness in this analysis is that the classes identified are both classes of cognitive decline and missingness pattern.(Muthen et al., 2011) Whilst these processes are closely linked it would be preferable to model them separately. Unfortunately, Bayesian estimation using multiple-membership latent classes are not yet implementable within available software. The use of Bayesian estimation could been seen as a weakness as Bayesian mixture modelling in general can be sensitive to prior specification, although they tend to converge with frequentist estimation with less informative priors.(Depaoli et al., 2017; Helm et al., 2017)

It seems relatively unlikely that my results have been unduly influenced by unmeasured confounding. What early life measures preceding education I had did not alter the principle finding of no association between educational attainment and rate of decline for most participants. Other unmeasured confounders would be anticipated to bias results away from, rather than towards, the null hypothesis. However, my method does not account for the various post-education pathways to cognitive decline. This being the case my results cannot say how much of the observed association is caused by mediating pathways rather than being the direct effect of education itself. I considered the inclusion of a range of post-education variables such as adult social status or occupational complexity. However, their inclusion would introduce a large number of additional modelling assumptions which are not necessarily sustainable. Ultimately, my research question was about estimating whether there is a total effect, not the many possible pathways this might take. It is also worth noting that it was not possible to elucidate cohort effects because of using both time and age in the model.(Bell & Jones, 2013)

In conclusion I identified two latent classes of cognition in a representative sample of the English older adult population, one of very stable function with minimal decline and another with rapid decline likely to represent a population with mostly pre-clinical dementia. In my analysis of change, after relaxing the assumptions of population heterogeneity and MAR, there was no evidence to support the hypothesis the cognitive reserve from early life education moderates longitudinal cognitive decline in those in the declining group. In the cognitively health older adult group there was evidence of a small degree of active or neural compensation reserve in verbal fluency for those with the highest levels of education, but no association was seen for recall.

Chapter 8. Cognitively stimulating activities and risk of probable dementia or cognitive impairment in the English Longitudinal Study of Ageing.

This paper presents the fourth paper which addresses the fourth research question of this thesis: does exposure to cognitively stimulating activities in later life reduce risk of dementia or cognitive impairment once time-varying confounding affected by past exposure is accounted for? As a secondary outcome cognitive function as a continuous score, rather than a set-cut off point will be used.

This paper presents the first use of marginal structural models to estimate the association between cognitively stimulating activities in later life and cognitive maintenance. This builds on previous research by using a method which can account for time-varying confounding affected by past exposure. The inability to exclude this as a cause of reverse causation is an important limitation in the current literature.

In the context of the thesis as a whole this analysis moves from cognitive stimulation in earlier life to later life. In doing so, it builds upon the approach which aims to faithfully represent an aspect of the underlying causal pathways in a way not previously done.

I conducted the analysis, drafted and revised the paper. My supervisors Prof. Chandola and Prof. Pendleton provided guidance on the analytic strategy and reviewed the drafts. Prof Gindo Tampubolon provided additional comments on a draft manuscript as part of an annual review.

This paper has not been published and is not under review at the time of submission.

Cognitively stimulating activities and risk of probable dementia or cognitive impairment in the English Longitudinal Study of Ageing.

Dr. Benjamin David Williams[1], BM,BS, MSc

Prof. Tarani Chandola[1], PhD

Prof. Neil Pendleton[2], MD

1. Cathie Marsh Institute for Social Research, Humanities Bridgeford Street Building, University of Manchester, M13 9PL

2. Institute of Brain, Behaviour and Mental Health, Stopford Building, University of Manchester

Abstract

Objectives

Cognitive stimulating activities (CSA) are an important potential preventative factor for cognitive impairment and dementia, but this association may be due to reverse causation. We aimed to estimate CSA's effect on risk of cognitive impairment whilst accounting for time-varying confounding affected by past exposure, one source of reverse causation.

Methods

We analysed data from n=11992 participants of the English Longitudinal Study of Ageing waves 1 (2002) to 7 (2014), a nationally representative prospective cohort of adults in England aged ≥50 (UK Data Service SN5050 https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=5050). Self-reported participation in internet use, employment, volunteering, evening classes, social clubs and newspaper reading was measured from waves 2 (2004) to 6 (2012). The primary outcome was wave 7 probable cognitive impairment or dementia (≤11/27 on the Telephone Interview for Cognitive Status). Inverse probability of treatment and censoring weighted repeated measures Poisson regressions were used.

Results

Volunteering was associated with reduced risk of cognitive impairment at wave 3 RR=0.633 (0.407 to 0.984), wave 4 RR=0.626 (0.401 to 0.977), wave 5 RR=0.516 (0.302 to 0.881) and wave 6 RR=0.564 (0.340 to 0.935). Internet use was associated with reduced risk of cognitive impairment at wave 2 RR=0.659 (0.518 to 0.839), wave 3 RR=0.652 (0.523 to 0.814), wave 4 RR=0.620 (0.502 to 0.767), wave 5 RR=0.666 (0.540 to 0.821) and wave 6 RR=0.691 (0.562 to 0.848). There was no significant association with employment, social clubs, evening classes or newspaper reading.

Conclusion

We found that volunteering and internet use was associated with reduced risk of cognitive impairment or dementia.

8.1 Introduction

Dementia and cognitive impairment are now established global health problems. Cognitively stimulating activities (CSA) are an important potential modifiable factor affecting cognitive decline and dementia.(Sajeev et al., 2016) CSA have been defined as 'mentally engaging activities or exercises that challenge a person's ability to think'.(Global Council On Brain Health, 2017) An effect of targeted interventions, particularly computerised cognitive training, has been found in healthy old age, mild cognitive impairment and dementia.(García-Casal et al., 2017; Hill et al., 2017) However, a major systematic review found 'moderate-strength evidence shows cognitive training in adults with presumed normal cognition improves performance in the cognitive domain trained…, but not transfer of benefits to other cognitive areas and little evidence for benefit beyond 2 years'.(Kane et al., 2017) This limitation stresses the importance of understanding the effect of cognitive activities which older adults already engage in. Mental activity is the dementia prevention strategy most commonly identified by adults in Western countries.(Friedman et al., 2015) Accordingly, it is common for older adults to engage in CSA with the intention of reducing their risk of dementia.(Hosking, Sargent-Cox, & Anstey, 2015) It is therefore important to understand whether these activities are effective to inform both public health interventions and individual choice.

In general, the popular view that CSA are beneficial agrees with recent systematic reviews and recommendations.(Global Council On Brain Health, 2017; Sajeev et al., 2016) However, it is possible that this association may be due to reverse causation, that is to say, it is better cognitive performance which predicts CSA.(Gow et al., 2012; Sajeev et al., 2016) If present, reverse causation could be time invariant, as in the case of childhood intelligence or time-varying but not affected by prior exposure.(Gow et al., 2012; Robins et al., 2000) However, there may also be time-varying confounding present which has been influenced by exposure at a prior time point.(Robins et al., 2000) Those with declining cognition are less likely to continue engagement in CSA, meaning cognition at one time may confound the relationship between future CSA and cognition.(M. J. Aartsen et al., 2002) This type of confounding cannot be accounted for using standard regression methods, but can be adjusted for using marginal structural models (MSMs).(Robins et al., 2000)

Marginal structural models, estimated with inverse probability of treatment weights (IPTW), are a means of, making causal inferences from observational data under strict assumptions.(Robins et al., 2000) They are especially valuable when time-varying confounding affect by prior exposure is highly probable, but conducting high quality randomised controlled trials (RCTs) is extremely difficult. Even in the absence of complete adherence to the assumptions required for causality, MSMs still relax the assumption that there is no time varying confounding affected by prior exposure implicitly present in other longitudinal models. Whilst gaining broad use across epidemiology in general, they have seen relatively limited application in cognitive

epidemiology.(Marden et al., 2017) We hypothesise that the association between CSA and cognitive function or dementia is affected by time-varying confounding influenced by prior exposure. Therefore, that exposure to CSA will be associated with risk of dementia or cognitive impairment using standard regression but that this association will be attenuated or not observed in inverse probability of treatment weighted marginal structural models.

8.2 Methods

8.2.1 Participants and Procedure

ELSA has been described in detail previously.(James Banks et al., 2016; Steptoe et al., 2013) Participants were drawn from a nationally representative multistage probability sample of adults aged 50 or more living England in 2002. Data from core sample participants was collected in biennial sweeps by interview in the participants homes. For this analysis data from waves 1 (2002) was used as baseline. Waves 2 to 6 (2004-2012) were used for exposure to CSA, the outcome from wave 7 (2014) and analysis conducted in 2018.

The size of the initial sample at wave 1 was 11992, falling to 4062 by wave 7. Data was used from all participants available in each wave for the creation of the treatment and non-response weights. Only data from participants present in all waves with full exposure data were utilised in the final analysis resulting in final sample sizes between n=3937 and n=2530 being included in the final analysis for each CSA.

Ethical approval for ELSA was granted by the South Central Berkshire Research Ethics Committee and the current study was subject to the University of Manchester internal review process.(James Banks et al., 2016) Written informed consent was obtained from all participants.

8.2.2 Outcome measures

The primary outcome was probable dementia or cognitive impairment at wave 7. This was diagnosed using questions from a modified telephone interview for cognitive status (TICS), is a 27 point scale using immediate and delayed 10 word free recall, backwards counting from 20 and serial 7 subtraction.(Langa et al., 2017) We used a binary outcome of non-impaired (12-27) and probable dementia or cognitive impairment (0-11; of which 0-6 is dementia and 7-11 cognitive impairment).(P. J. Clarke et al., 2012) Our secondary outcome was the continuous TICS-27 score (a higher score indicates better cognition).

8.2.3 Cognitive Stimulating Activities

We chose 6 CSA's from the range of activities in ELSA to represent variety in older adult's lifestyles and the type of cognitive challenge presented. We chose CSA where it is relatively clear how an intervention might be designed to alter an individual's exposure. The CSA's chosen were working, volunteering, regular newspaper reading, attending arts/music/evening classes (hereafter 'evening classes'), internet or email use and social club membership. All of these activities have been previously found to have an association with cognitive function, including some in other analyses of ELSA data using standard regression methods.(d'Orsi et al., 2017; Guiney & Machado, 2018; Hikichi et al., 2017; Liapis & Harding, 2017; A. Meng et al., 2017; Sajeev et al., 2016)

For working and volunteering participants were asked in the main ELSA interview 'Did you do any of these activities in the last month?'. Participants attending evening classes and social clubs were asked in a separate self-completion questionnaire 'Are you a member of any of these organisations, clubs or societies?'. 'I read a daily newspaper' and 'I use the internet and/or email' were response options for the question 'Which of these statements apply to you?', also in the self-completion questionnaire. Exposure at wave 1 was treated as a baseline variable and waves 2-6 were used to measure time-varying exposure.

8.2.4 Covariates

Due to the large number of exposures, a wide range of potential covariates were considered that might plausibly confound the association between exposure and outcome. Across all exposures the time invariant covariates used were gender, age at recruitment, highest educational qualification, income quartile, ethnicity (white or non-white), and parental smoking. The time varying covariates were being above retirement age, other activities (including religious participation, charitable activities, daytrips, mobile phone use, holidays, voting, having a hobby and other class or society memberships), caring, homemaking, self-rated health, self-rated hearing, self-rated eyesight, marital status, psychiatric illness, depression score, number of cigarettes smoked per day (0, 1-10, 11-19 or 20), vigorous, moderate and light exercise and cognitive function. Verbal fluency (number of animals named in one minute) and episodic memory (sum score of immediate and delay recall of a 10-word list) were used as longitudinal measures of cognitive function.

8.2.5 Statistical Analysis

A marginal structural model was constructed for the effect of exposure to each individual CSA from waves 2 to 6 on risk of dementia or cognitive impairment at wave 7. Further detail on the rationale for and calculation of marginal structural models are presented in appendix 1. In brief our core MSM is specified as:

1. $\log(\lambda_{\bar{a}} \mid \bar{A} = \bar{a}, \bar{L} = L_1, \bar{V} = V_1) = \beta_0 + \beta_1 v_1 + \beta_2 a_2 + \beta_3 a_3 + \beta_4 a_4 + \beta_5 a_5 + \beta_6 a_6 + \beta_7 l_1$

This estimates the expected risk of cognitive impairment or dementia had it hypothetically been possible to intervene and set each individuals exposure to CSA at each wave to a pre-specified

value. $\bar{A}$ represents the observed history of exposure to CSA $(A_2, \ldots, A_6)$ and $\bar{a}$ all possible exposure histories which could have been observed. $\lambda_{\bar{a}} = E(Y_{\bar{a}})$ where $Y_{\bar{a}}$ is the potential outcome given an exposure history of $\bar{a}$. Each of the terms $\beta_2 a_2 \ldots \beta_6 a_6$ are the effect of intervening to set the exposure to the CSA for each of waves 2 to 6. Also included are baseline confounders $\beta_7 l_1$ and baseline CSA exposure $\beta_1 v_1$which are included because the weights are stabilised. Interactions between exposure in waves 2-6 and baseline covariates were tested but not significant for any of the CSA and therefore not included above. Interactions with time-varying covariates are not possible but estimates of total effect are unbiased even if these interactions are present.

A forward selection process was used to identify a unique set of covariates which predicted future exposure to each CSA. This was used to form a single model predicting CSA across all waves (appendix 8.1). The odds of exposure were calculated for each time point using logistic regression.(Robins et al., 2000) Stabilised inverse probability of treatment weights were calculated as the odds of exposure dependent on past exposure to CSA and baseline covariates only divided by odds of exposure dependent on full covariate and exposure history for each wave.(Fewell et al., 2004; Robins et al., 2000) Each wave specific exposure weight for waves 2-6 was multiplied to give the overall IPTW for each CSA.

The ELSA study dataset provides longitudinal inverse probability of censoring weights (wave-IPCW) for core members participating in each ELSA wave from wave 1 onwards.(James Banks et al., 2016; Fewell et al., 2004) Dropout was treated as monotone. Additional weights are given for non-response to the self-completion questionnaire (self-completion-IPCW). The final inverse probability of treatment and censoring weights (IPTCW) were given by the product of the IPTW with the wave-IPCW only for employment and volunteering, and both wave-IPCW and self-completion-IPCW for the remaining CSA. To estimate the MSM for the primary outcome of risk of probable cognitive impairment the IPTCW was applied to a Poisson regression with robust error variance.(G. Y. Zou, 2009) Results were compared with the same modified Poisson regression using standard regression adjustment for cumulative covariate time-varying exposure and weighted only for non-response. We used Poisson regression to directly estimate relative risks, in preference to odds ratios from logistic regression, to aid interpretability. Additionally, we present statistically significant results as adjusted absolute risk difference (AARD) and number needed to expose (NNE). For the secondary outcome of TICS-27 score the models were run in the same fashion (IPTCW vs regression adjustment) using a linear regression.

The data was analysed using Stata version 13.0.(StataCorp, 2013) For the references on which this was based we refer the reader to Fewell and colleagues for the Stata code for calculating IPTCW and to Bodnar and colleagues for the weighted repeated measures regression.(Bodnar et al., 2004; Fewell et al., 2004)

8.3 Results

The mean TICS-27 score was 15.0 (S.D 5.4). The proportion of participants classified as having probable cognitive impairment or dementia in wave 7 was 21.3% (n=867). Demographics are presented in table 8.1. When compared with the core sample present at wave 1, respondents for wave 7 were more likely to be female, to have degree level education, to be white and to have had parents who smoked. Missingness of each exposure is presented in appendix 8.2 and further detail is in the ELSA wave 7 study report (James Banks et al., 2016). CSA exposure over time is presented in table 8.2. Over the study period employment reduced, volunteering increased slightly, there was a strong trend towards increasing internet use, social club membership fell slightly, evening class attendance was stable and newspaper reading fell substantially, but remained the majority of participants.

In standard regression models volunteering was significantly associated with reduced relative risk (RR) of cognitive impairment at wave 4 (RR 0.801; 95% CI 0.656 to 0.977), wave 5 (RR 0.731; 95% CI 0.588 to 0.909) and wave 6 (RR 0.683; 95% CI 0.547 to 0.853; figure 8.1 and appendix 8.3). In the standard regression models for TICS-27 score, we see a similar pattern with an increase in score for wave for wave 3 (0.432; 95% CI 0.137 to 0.727), wave 4 (0.348; 95% CI 0.047 to 0.649), wave 5 (0.433; 95% CI 0.119 to 0.747) and wave 6 (0.530; 95% CI 0.224 to 0.836; figure 8.2 and appendix 8.4).

After adjustment with IPTCW volunteering was more strongly associated with reduced risk of cognitive impairment than estimated in the standard regression models (figure 8.1). The RRs were 0.649 (95% CI 0.417 to 1.009) for wave 2, 0.633 (95% CI 0.407 to 0.984) for wave 3, 0.626 (95% CI 0.401 to 0.977) for wave 4, 0.516 (95% CI 0.302 to 0.881) for wave 5 and 0.564 (95% CI 0.340 to 0.935) for wave 6. The AARD between individuals who volunteered for all 5 waves, approximately 10 years exposure, and those who did not volunteer at any time was 17.2% (95% CI 2.9% to 19.2%). The NNE to prevent one case of cognitive impairment or dementia for individuals who volunteered for all 5 waves compared to those who did not volunteer at any time was 5.8 (95% CI 5.2 – 34.5). As an example of a single wave, approximating a 2-year period of volunteering, the AARD for wave 6 was 7.6% (95% CI 13% to 11.7%) and NNE 13.2 (95% CI 8.5 to 78.0).

Contrasting with these results, volunteering at any wave was not significantly associated with TICS-27 (figure 8.2). To explore this finding the analysis was restricted to only those participants in the bottom 50% of the TICS-27 distribution. In this subset volunteering was associated with TICS-27 at wave 2 (0.821; 95% CI 0.243 to 1.399), wave 3 (0.732; 95% CI 0.010 to 1.365), wave 4 (0.792; 95% CI 0.212 to 1.372) and wave 6 (0.624; 95% CI 0.004 to 1.245) but not quite significant at wave 5 (0.637; 95% CI -0.038 to 1.313).

Table 8.1. ELSA core sample demographics showing the time-invariant covariates used in this analysis at waves 1 and 7.

		Wave 1 (2002/3)	Wave 7 (2014/15)
n		11992	4062
Female (%)		6676 (55.7%)	2291 (56.4%)
Age at recruitment (S.D)		64.7 (10.7)	61.7 (7.9)
Educational Attainment (%)	No formal qualification	4986 (41.7%)	1197 (29.5%)
	High School	2522 (21.1%)	979 (24.1%)
	6th Form	748 (6.3%)	301 (7.4%)
	Some higher education	1317 (11.0%)	603 (14.9%)
	Degree or higher	1370 (11.5%)	638 (15.7%)
	Foreign Qualification	1014 (8.5%)	343 (8.5%)
Non-white Ethnicity (%)		328 (2.8%)	66 (1.6%)
Parental Smoking	Paternal	9099 (79.7%)	3153 (80.4%)
	Maternal	3923 (33.7%)	1544 (38.6%)
TICS† score			15.0 (S.D 5.4)
Cognitive Status	Non-impaired		3195 (78.7%)
	Cognitive Impairment		549 (13.5%)
	Dementia		318 (7.8%)

†TICS, Telephone Interview for Cognitive Status

Table 8.2 Proportion of ELSA core sample participating in the cognitively stimulating activities by cognitive status.

Study Wave	Wave 1 (2002/3)		Wave 2 (2004/5)		Wave 3 (2006/7)		Wave 4 (2008/9)		Wave 5 (2010/11)		Wave 6 (2012/13)	
Year	TICS≥12	TICS≤11	TICS≥12	TICS≤11	TICS≥12	TICS≤11	TICS≥12	TICS≤11	TICS≥12	TICS≤11	TICS≥12	TICS≤11
Employment	1722 (53.9%)	226 (26.1%)	1479 (46.3%)	187 (21.6%)	1332 (41.7%)	153 (17.7%)	1102 (34.5%)	123 (14.2%)	860 (26.9%)	92 (10.6%)	628 (19.7%)	75 (8.7%)
Volunteering	572 (17.9%)	107 (12.4%)	621 (19.4%)	113 (13.0%)	648 (20.3%)	109 (12.6%)	641 (20.1%)	97 (11.2%)	664 (20.8%)	90 (10.4%)	675 (21.1%)	81 (9.3%)
Internet/Email	1417 (45.8%)	176 (21.8%)	1584 (53.0%)	195 (25.9%)	1665 (57.1%)	195 (25.9%)	1779 (60.5%)	197 (26.84)	1972 (65.2%)	214 (28.5%)	2060 (68.6%)	223 (33.0%)
Social Club	620 (20.3%)	185 (24.2%)	574 (19.8%)	162 (23.5%)	552 (19.3%)	159 (23.0%)	579 (20.2%)	142 (20.8%)	605 (20.4%)	134 (19.4%)	581 (19.7%)	115 (18.1%)
Newspaper	2092 (67.7%)	545 (67.5%)	1967 (65.8%)	479 (63.6%)	1951 (66.9%)	480 (63.8%)	1902 (64.7%)	436 (59.4%)	1919 (63.4%)	464 (61.9%)	1829 (60.4%)	408 (58.2%)
Evening Classes	591 (19.4%)	95 (12.4%)	516 (17.8%)	80 (11.6%)	482 (46.9%)	62 (9.0%)	441 (15.4%)	53 (7.8%)	472 (15.9%)	61 (8.8%)	490 (16.6%)	42 (6.6%)

Figure 8.1 IPTW vs standard regression models for CSA predicting risk of probable cognitive impairment in 2014 (wave 7).

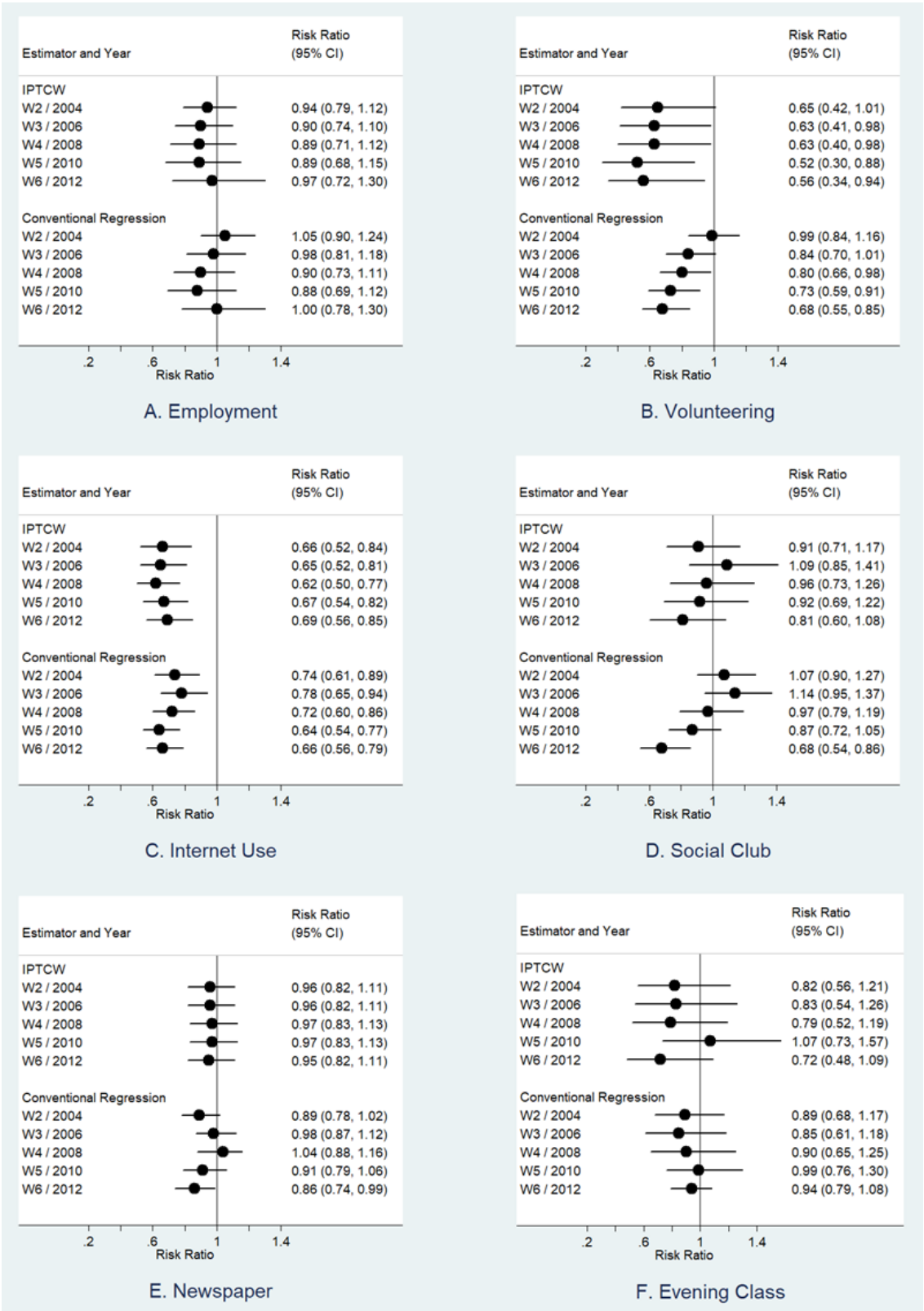
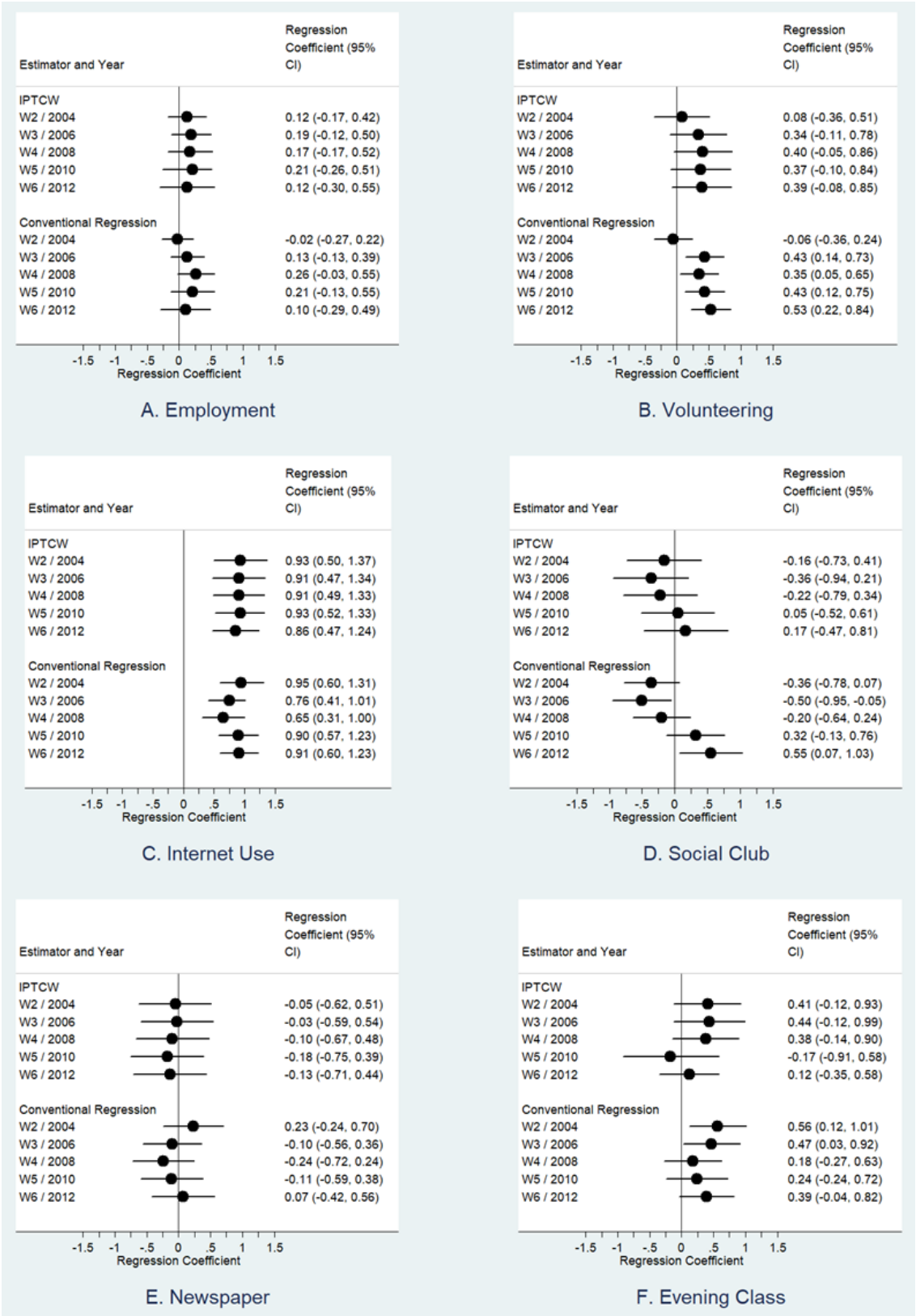


Figure 8.2 IPTW vs standard regression models for CSA predicting TICS-27 score in 2014 (wave 7).



In the standard regression model internet/email use at any wave was associated with lower risk of cognitive impairment. The RRs were 0.737 (5% CI: 0.611 to 0.889) at wave 2, 0.780 (95% CI 0.647 to 0.940) at wave 3, 0.720 (95% CI 0.601 to 0.862) at wave 4, 0.642 (95% CI 0.538 to 0.765) at wave 5 and 0.664 (95% CI 0.562 to 0.785) at wave 6. The IPTCW models estimated similar RR to the standard models with estimates of 0.659 (95% CI 0.518 to 0.839) at wave 2, 0.652 (95% CI 0.523 to 0.814) at wave 3, 0.620 (95% CI 0.502 to 0.767) at wave 4, 0.666 (95% CI 0.540 to 0.821) at wave 5 and 0.691 (95% CI 0.562 to 0.848) at wave 6. The AARD between individuals reporting internet use for all 5 waves and those who did not at any time was 18.2% (95% CI 13.2% – 19.9%). The NNE to prevent one case of cognitive impairment or dementia for individuals who reported internet use for all 5 waves compared to those who did not at any time was 5.5 (95% CI 5.0 to 7.6). As an example of a single wave, approximating a 2-year period of internet/email use, the AARD for wave 6 was 6.4% (95% CI 3.2% to 9.0%) and NNE 15.6 (95% CI 11.1 to 31.7). Estimates of change in TICS-27 score were very similar in all waves (all $p<0.001$) for both standard and IPTCW models (figure 8.2).

Social club membership was associated with reduced risk of cognitive impairment at wave 6 (RR 0.678, 95% CI 0.535 to 0.859) in the standard model. Social club membership at wave 3 was associated with worse TICS-27 (-0.502; 95% CI -0.953 to -0.052) and at wave 6 with improved TICS-27 (0.546; 95% CI 0.067 to 1.026). In the IPTCW models social club membership at any wave was not associated with either risk of cognitive impairment or TICS-27 score.

In the neither the standard or IPTCW models were evening classes associated with risk of cognitive impairment. They were associated with higher TICS-27 at wave 2 (0.564; 95% CI 0.116 to 1.011) and 3 (0.473; 95% CI 0.025 to 0.922) in standard but not IPTCW models. Daily newspaper reading was associated with reduced risk of cognitive impairment at wave 6 (RR 0.855; 95% CI 0.741 to 0.988) in the standard model but not the IPTCW. Employment was not associated with risk of cognitive impairment or TICS-27 score in any model.

8.4 Discussion

Contrary to our hypothesis, we found in the IPTCW models the association between volunteering or internet/email use and risk of cognitive impairment was not substantially different. This implies that their association with cognitive impairment is not due to time-varying confounding affected by prior exposure, at least not from the potential confounders we accounted for. Employment, evening classes, newspaper reading and social clubs were not associated with improved cognitive function. Of these, the results for newspaper reading and social clubs were consistent with our hypothesis, with some associations becoming non-

significant using IPTCW. The effect of volunteering and internet use persisted over several years and, for volunteering, appeared limited to those with poorer cognitive function. That the association between cognitive impairment and social club membership reduced when IPTCW was used, supports the suggestion that the association between social activity and cognitive impairment or dementia may be due to reverse causation. It is consistent with prior research showing that neuropsychiatric manifestations such as apathy, depression and social withdrawal can precede the onset of measurable cognitive impairment.(Cortés, Andrade, & Maccioni, 2018; Sajeev et al., 2016)

8.4.1 Strengths and Limitations

ELSA is a large, nationally representative prospective cohort study with data on many covariates over a long follow-up period. However, the measures of CSA are self-reported, binary measures taken every 2 years. Whilst the long study duration provides good information on participation over time, there is little information on CSA 'dose' received. Additionally, we did not draw other distinctions such as between those in employment due to being unable to work/retire versus preferring not to. We used IPCW to account for dropout, which is a standard approach to missingness at random that is readily integrated into IPTW.(Fewell et al., 2004) However, it does not account for different missingness mechanisms.

The use of MSMs is the main development of the current study on previous research. Having used MSMs it is possible to draw tentative causal conclusions from our analysis, but this is conditional on the strong assumption of no unmeasured confounding.(Robins et al., 2000) The use of baseline CSA at wave 1 to predict subsequent exposure should reduce the effect of confounders which are unobserved because of left censoring, as this is likely to be substantially mediated by baseline CSA. There may be unobserved social stratification or social cognitive deficit which is inadvertently measured by one's propensity to volunteer or take up internet use. However, one might expect such confounders to affect other CSAs similarly, particularly evening classes. The fact it did not gives a degree of support to the no unmeasured confounders assumption.

8.4.2 Relation to other studies

Consistent with our findings, the effect of volunteering seen in the Experience Corps studies seems to be more pronounced in those with poorer cognitive function.(Proulx et al., 2018) Data from other RCTs so far are somewhat conflicting but, considering the lead time and prevalence of neurodegenerative disease, they have been short term or small.(Sakurai et al., 2018) Modest gains in hippocampal size may be the neural substrate of improved cognitive function amongst volunteers in these studies.(Michelle C. Carlson et al., 2015; Sakurai et al., 2018) Volunteering has also been associated with better cognitive function in several cohort studies.(Guiney &

Machado, 2018; Jenkinson et al., 2013; Proulx et al., 2018) According to Guiney and Machado's theoretical framework, volunteering exerts its beneficial effects on cognitive functioning via increased cognitive, social and physical activity leading to improved neurological and mental health which, in turn, improve cognitive functioning.(Michelle C. Carlson et al., 2015; Guiney & Machado, 2018) Should these findings be robustly replicated then this would suggest that volunteering could be promoted as an intervention for those with poorer cognitive function or at risk of dementia.(Jenkinson et al., 2013) Given that this analysis used a nationally representative sample of older adults living in England, this finding may be generalisable to older adults in the rest of Britain as well as comparable western industrialised nations.

General internet use has been associated with improved cognitive function and reduced risk of dementia in several previous observational studies, including in previous analyses of ELSA.(d'Orsi et al., 2017; Liapis & Harding, 2017) This similarity is unsurprising given that our IPTW estimates were similar to the standard regression estimates. It is difficult to contrast our results regarding general internet use with RCTs because they use software designed specifically to target cognitive function.(Liapis & Harding, 2017) Higher levels of internet use are associated with improved access to preventative and treatment health services.(C. S. Clarke et al., 2017) Other mechanisms for a causal association with general computer use have had little specific investigation. It is possible that the exposure being measured is not computer use itself, but acquiring and routinely using a new skill in later life. If this is the case, this finding may be generalisable across similar age cohorts but is unlikely to be generalisable to future cohorts.

Our finding that the specific type of activity is important contrasts with the conclusion from some of the prior research regarding holistically cognitively enriched lifestyles.(Sajeev et al., 2016) These differences may have arisen simply because we chose to model activities individually or our use of IPTCW. Studies generally are highly heterogeneous in their conceptualisation of both CSA and cognitive functioning which makes direct comparison across studies challenging.

There remains a large number of unanswered questions in the area of CSA and cognitive function. The type, duration and intensity of CSA required to produce benefits in cognitive functioning remains to be characterised more fully. Whether all older adults could anticipate cognitive benefits or if the effects are indeed limited to those with poorer overall cognitive functioning requires further replication. Whilst application of methods for casual analysis of observational data, such as MSMs used in this study, can advance knowledge and identify the CSA most likely to produce benefits, there is still no replacement for well conducted RCTs for robust causal inference.

9. Discussion and Conclusions

This chapter reflects on the main findings from the analysis and discusses the contributions, limitations, theoretical and policy implications of this study and implications for future research.

9.1 Research Questions and new evidence

In this thesis, four main research questions have been examined. These were: (1) When analysing predictors of cognitive maintenance in later life, do cognitive function factor scores lead to different substantive conclusions than sum-scores? (2) Are Bayesian approximate measurement invariance models of cognitive ageing a suitable alternative to conventional longitudinal measurement invariance models? (3) Is higher education associated with brain, neural or neural cognitive reserve, and does this association vary by latent class of cognitive decline in later life? (4) Does exposure to cognitively stimulating activities in later life reduce the risk of dementia or cognitive impairment once time-varying confounding affected by past exposure is accounted for?

9.1.1 Research Question 1

Factor analysis is a commonly used technique for data reduction and estimation of quantities which are not directly observable, such as cognitive function. This technique can test hypotheses about the structure of a measurement instrument to ensure it represents the data, can handle measurement error present in the individual tests and is readily expanded into the flexible structural equation modelling framework. However, it also adds several stages to any analyses done on the data, can appear to complicate comparability across studies and reduce interpretation by generalist researchers when compared to the standardised sum scores commonly used in studies of cognitive ageing. I sought to compare using factor analysis with the standard cognitive scores from the English Longitudinal Study of Ageing (ELSA) to see if there were meaningful differences in results. These scores, called indices in the ELSA documentation, were derived by simply adding the scores from different tests in the domains of memory and executive function, with a global score from adding those together.

I qualitatively compared the results of an exploratory factor analysis with the structure of the indices. I found that the memory index and a memory factor were similar in structure, however the executive function index appeared to combine two separate factors for attention and a visual task. The global index did not represent the data well and no single factor encompassed all the individual cognitive tests. After using a confirmatory factor analysis to estimate factor scores, I then regressed both standardised factor scores and standardised index scores on a range of common predictors of cognitive function. The memory factor and index showed very similar relationships with predictor variables, with only 1 out of 47 parameters being significantly different. The executive function index and the corresponding factors for attention (9 differences in coverage and 8 in inferences) and visual search (14 differences in coverage and 16

differences in inferences) showed substantial differences in the relationship to predictor variables. For example, the visual search factor was not associated with Parkinson's disease, cigarette consumption or depression. All of these had substantial negative effects on the executive function index despite the visual search task being one of the main tasks used to calculate the executive function index.

This paper presents a factor analysis of the cognitive function tests in ELSA which may prove useful to future researchers wishing to use this dataset. It demonstrates how using factor scores in ELSA can lead to quite different substantive conclusions to sum scores which have been used frequently in study reports and secondary analyses of ELSA.(Huppert et al., 2006; Lang, Wallace, Huppert, & Melzer, 2007; Langa et al., 2009; Shankar, Hamer, McMunn, & Steptoe, 2013b) This paper contributes to the literature more generally by demonstrating from a real-world dataset how pre-defined sum-scores may not reflect underlying data structures well.

9.1.2 Research Question 2

Each test which measures cognitive function relies on a combination of processes which may be variably affected by ageing and disease. This means there is good reason for testing the longitudinal measurement invariance of cognitive function in later life rather than assuming it exists.(Blankson & McArdle, 2013; McAvinue et al., 2012) In studies with large numbers of participants, measurement invariance is typically tested for using alternative fit indices, of which the comparative fit index (CFI) is the most common.(Cheung & Rensvold, 2002; Meade & Bauer, 2007) However, different fit indices may give conflicting results, one observed variable must always be assumed to be invariant and the exact change in CFI which indicates non-invariance changes for any given combination of observed variables and measurement occasions..

When testing for measurement invariance of two factors in ELSA, I found that for a memory factor, the test results fell in between different CFI recommendations from different authors. In order to move forward in the research, I needed to know whether or not the memory factor did possess longitudinal measurement invariance. I identified that Bayesian approximate measurement invariance may provide a viable alternative to testing with the CFI.(van de Schoot et al., 2013) Instead of testing global model fit at different levels of invariance like CFI, approximate measurement invariance allowed me to test for a pre-specified acceptable amount of variation for the factor loading and intercept of each observed variable. Having run this analysis, I found that there were several intercepts which showed evidence of non-invariance in the memory factor. I also confirmed that there was minimal non-invariance in a factor measuring attention. I concluded that there was sufficient non-invariance that the memory factor was not suitable for longitudinal analysis. Due to this result, the next paper in the thesis focussed on two specific memory tests but not an overall memory factor.

As well as informing the next step of my research, this analysis develops existing literature on Bayesian approximate measurement invariance by demonstrating its use in a new context. Previous literature on approximate measurement invariance has focussed on the problem of multiple marginally non-invariant parameters resulting in the rejection of measurement invariance for valid measurement tools because of overly conservative tests.(Bengt Muthén & Asparouhov, 2012; van de Schoot et al., 2013) I develop this by showing that it is not simply a case of using conventional or Bayesian tests, but that the Bayesian approach can be used as a valid alternative to or as a compliment to conventional tests. This may be needed when conventional tests provide equivocal results or where there is no clear cut-off for the number of observed indicator variables at the number of times points you are using. Instead of focussing on the multiple minor non-invariance problem, I emphasise that because approximate measurement invariance tests individual parameters rather than global fit, the Bayesian approach can provide additional diagnostic information about the invariance of a given factor structure.

9.1.3 Research Question 3

Higher levels of education are well known to be associated with reduced risk of developing the clinical syndromes of mild cognitive impairment or dementia.(Beydoun et al., 2014; Livingston et al., 2017) However, higher levels of education are associated with more rapid cognitive decline in dementia, and clinicopathological studies have found that education does not seem to be associated with the burden of Alzheimer's disease pathology.(Brayne et al., 2010; X. Meng & D'Arcy, 2012; Serrano-pozo et al., 2013) There has been disagreement in the research literature regarding whether these effects are due to greater education simply increasing baseline cognitive performance or whether it affects cognitive maintenance longitudinally.(Plassman et al., 2010) Furthermore, nearly all studies have either analysed their data using a single overall trajectory of cognitive maintenance or used education to predict odds of belonging to a particular trajectory. This leaves open the question of whether an association between cognitive maintenance over time and education may be moderated by class of latent trajectory. The theories of brain, neural and neural compensation reserve each lead to different predictions about the effect of education on cognitive reserve and each have evidence to support them.(M. Tucker & Stern, 2011; Stern, 2012) It may be that they are not mutually exclusive but that different types of reserve are predominant in different states, especially in health vs disease. As the progression of disease can precede clinical diagnosis by many years, classifying individuals with declining or stable cognitive trajectories from a population sample allows the hypothesis that education will be associated with different forms of cognitive reserve in different states to be tested.

I used two cognitive tests in ELSA, immediate recall of a 10-word list and the number of animals named in one minute. I estimated their association with educational attainment over 5 waves (8 years) of follow-up using a growth mixture model. 2 latent classes of cognitive trajectory were

identified for both tests, which I termed stable and declining. In the declining class educational attainment was associated with baseline performance but not rate of change. In the stable class degree level education was associated with a slight improvement in maintenance of verbal fluency. However, this is of minimal substantive importance.

This analysis builds on prior research by demonstrating that greater education contributes primarily to passive cognitive reserve with only a slight contribution to neural compensation reserve in cognitively stable individuals. It does this by relaxing the assumption that there is only a single trajectory within the population, and allowing the effect of education to vary by class rather than predict class membership. With the exception of Terrera, Brayne and Matthews (2010), prior estimation of the association between education and cognitive maintenance within class had not been tested. It develops their analysis specifically by using different cognitive measures tests without strong floor or ceiling effects, by using a dataset representative of the whole of England and with a larger sample size. It also implements a method of accounting for informative dropout within class using pattern mixture modelling which has not previously been applied in analyses of cognitive maintenance.(Beunckens et al., 2008)

9.1.4 Research Question 4

Cognitively stimulating activities (CSA) are a promising potentially modifiable factor which may improve cognitive maintenance and prevent dementia or mild cognitive impairment.(Sajeev et al., 2016) Whilst greater exposure to CSA has been found to be associated with reduced risk of dementia or cognitive impairment, reverse causation from time-varying confounding affected by past exposure may explain this association.(Sajeev et al., 2016) This is where past CSA (A1), such as volunteering, affects cognitive function (C1) which then affects the probability of future volunteering (A2).(Jenkinson et al., 2013; Shen, 2017) With conventional regression you cannot unbiasedly estimate the effects of both A1 and A2 on the risk of dementia (Y). If you condition on C1 you block the effect of A1 on Y mediated via C1 (including via A2) and if you don't condition on C1 then C1 confounds the association between C2 and Y. Population studies to date have not used methods to address this and the experimental literature is limited by a small number of studies, with small sample sizes and short follow-up periods.(Iizuka et al., 2019; Sajeev et al., 2016)

I estimated the association between risk of cognitive impairment or dementia and 5 waves (8 years) of exposure to working, volunteering, evening classes, social clubs, newspaper reading and internet use using marginal structural models (MSMs).(Robins et al., 2000) To do this, I first calculated stabilised inverse probability of treatment and censoring weights (IPTCW). These weights were applied to a separate repeated-measures modified Poisson regression for each CSA.(Bodnar et al., 2004; G. Y. Zou, 2009) This estimated the association of exposure at each time-point with risk of dementia or cognitive impairment at the end of the follow-up. I contrasted these results to those found using conventional regression adjustment with weighting only used

to account for non-response. I found that volunteering and internet use were associated with reduced risk in both standard and IPTCW models. Social club membership was associated with reduced risk in standard models but not IPTCW models. Working, newspaper reading and evening classes were not associated with risk in either standard or MSM models.

This paper is the first to present an analysis of a range of CSA using marginal structural models to account for potential reverse causation. This contributes to the evidence from ELSA and elsewhere that internet use and volunteering are associated with reduced risk of cognitive impairment.(d'Orsi et al., 2017; Guiney & Machado, 2018; Liapis & Harding, 2017; André J. Xavier et al., 2014) More specifically, my findings contribute to the existing literature by demonstrating that this association is unlikely to be due to time-varying confounding affected by prior exposure. The inability to account for this form of confounding was a key limitation in the existing literature. Alongside unobserved confounding, reverse causation is one of the major weaknesses of standard regression approaches which prevent causal inferences from being drawn from observational data. The association between social club membership and risk of cognitive impairment found in the standard regression was not seen in the MSM. This suggests this association was due to reverse causation and demonstrates why adjusting for time-varying confounding affected by prior exposure is important. The lack of association between social club membership and cognitive impairment using an IPTCW MSM contributes to the literature suggesting that the social withdrawal seen prior to dementia diagnosis is likely to be an early symptom rather than a cause of the condition.(Ismail et al., 2016; Singh-Manoux et al., 2017)

9.2 Strengths and Limitations

The general strength of the analytic approach taken in this thesis was using structural equation modelling and marginal structural modelling to relax the constraints implicit in more common random effects and growth curve models. This confers specific strengths to each of the individual papers. Through the thesis as a whole this is done in a progressive fashion. By sequentially testing important modelling assumptions in measures of cognitive ageing, and demonstrating the influence of not doing so, the thesis builds its substantive arguments based on rigorous methodological foundations. The ELSA dataset itself confers several strengths to the analyses presented in this thesis. It is a large and nationally representative cohort study with a wide range of covariates and long follow-up period.(Littleford et al., 2016) The representative sample provides a good basis for generalisability to the general British population and to other westernised industrialised nations. The range of covariates reduces the chances of the findings being due to unmeasured confounding. The follow-up period of 8-12 years is a suitable window for analysing cognitive maintenance.(Cavedo, Odile, & Lamari, 2017)

The use of education as a predictor of cognitive maintenance and a proxy for cognitive reserve has well known limitations. Early life education is determined by other factors such as general intelligence and early life social circumstances.(Gow et al., 2012; Stern, 2012) Although

including early life socio-economic variables made little difference to the substantive conclusions drawn, intelligence, genetic effects and other unobserved early life exposures may confound the association observed between education and cognition.(Marden et al., 2017)

The lack of direct measurement of brain anatomy and function, is another limitation of this research.(Perneczky, Kempermann, Korczyn, Matthews, & Ikram, 2019) More direct measures of brain and cognitive reserve are now becoming available in some cohort studies. This will enable future work to disentangle the causal pathways and neuroanatomical substrates of cognitive maintenance in greater detail.(Cabeza et al., 2018; Stern et al., 2018; Weiler et al., 2018) This will also enable additional variables to be used to assist with the definition of latent classes.(Zahodne, Wall, et al., 2015)

The ELSA cohort grew up in a period where higher education was achieved only by a small minority. Methods of teaching and learning as well as physical school environments were very different to modern education in Britain or elsewhere. This means that generalisation of the findings of this research to future generations or across populations relies on the assumption of a broad equivalence in the education experienced at each level.(Gindo Tampubolon, 2015) This problem of cross-cohort generalisation is also present for the later life CSA. This is most apparent in the case of internet use. Internet use expanded rapidly across the population, including amongst older adults during the period of ELSA data collection.(Matthews, Nazroo, & Marshall, 2019) The analysis in this thesis is unable to say whether the reduction in risk of dementia or cognitive impairment is because of the direct or indirect effects of internet use itself. It seems plausible that a better ability to access information, preventative health services and social interaction may lead to an improvement in cognitive maintenance and a reduction in the risk of cognitive impairment.(C. S. Clarke et al., 2017; Andre Junqueira Xavier et al., 2013) It is also plausible that the combination of visual, attention and motor skills employed in using the internet provides an effective combination to stimulate neuroprotective mechanisms and maintain functional brain networks.(Liapis & Harding, 2017; Silbert et al., 2016; Steffener & Stern, 2012) If these are the causal mechanisms involved then the finding of a protective effect for internet use will generalise to future cohorts. However, it is also possible that internet use amongst this cohort of older adults over this specific historical period represents the acquisition and, importantly, routine use of a new skill in later life. This could mean that a protective effect of internet use may not be present in future generations because they have acquired this skill earlier in life. As internet use is likely to be almost ubiquitous in future western cohorts, it will be very difficult to establish a protective effect even if it is present. However, similar contrasts could be adopters and non-adopters of future technological innovations which require skill acquisition and use.

A challenge across the thesis as a whole was that a Bayesian or frequentist approach to estimation was not consistently taken throughout the thesis. However, neither of the two papers presented using Bayesian estimation are fully Bayesian. By fully Bayesian I mean in the sense that conventional informative priors are used. Instead, I have taken a ‘calibrated Bayesian’

approach to the analysis. This approach seeks to combine Bayesian estimation and model assessment with good frequentist properties and ideas.(Little, 2006) In the second paper presented the priors are strongly informative of the difference between factor loadings and thresholds. No prior was specified for the value of the factor loadings and thresholds as would be typical in Bayesian factor analysis.(Robert C. MacCallum et al., 2012) The fourth paper presented marginal structural models and uses frequentist estimation. MSMs have been implemented using Bayesian estimation.(Saarela, Moodie, & Stephens, 2015) However, there are several difficulties with this. The inverse probability of treatment and censoring weights used to estimate the MSM in the fourth paper are a form of high dimensional propensity scores. They have very poor frequentist properties with fully Bayesian estimation because it becomes progressively less likely that the prior will be correctly specified as more covariates are added.(Robins, Hernán, & Wasserman, 2015) More fundamentally, in combining prior and observed likelihood to estimate a posterior distribution there is no place in standard Bayesian logic for a propensity score.(Robins et al., 2015) Resolving this tension was beyond the scope of the thesis and therefore a frequentist estimator was used to estimate the MSM.

9.3 Theoretical Implications

This thesis is informed by, and contributes to theories, regarding the measurement of cognitive function using structural equation modelling, the type of cognitive reserve provided by education and counterfactual models of causation as applied to the effect of CSAs on risk of cognitive impairment. These share the unifying theoretical concern of how I can translate beliefs about complex underlying causal structures into my analysis. The first two main research questions examined in this thesis have primarily theoretical implications regarding how cognitive function is measured. In the process of the thesis as a whole, they were important for rigorously conducting the later longitudinal analysis of cognitive maintenance in ELSA.

The first paper relates to the basic principles of factor analysis. That it is a means of estimating the shared variance of observed variables in order to estimate an underlying latent variable whilst separating out non-shared variance as measurement error.(Meredith & Teresi, 2006)The essence of the first question is to what extent cognitive sum scores and cognitive factors lead to substantively different conclusions about the variables which predict cognitive function. The specific application of the factor analysis in this thesis provides a point of departure for future research using the ELSA data or other studies with a similar combination of cognitive tests. It was surprising to me when I started work on the thesis that no one seemed to have previously published tests the measurement properties of the cognitive battery in ELSA. I feel that this paper demonstrates the importance of this step in an analysis of cognitive function, despite the additional burden placed on researchers in doing so. The executive function index provided with ELSA had quite different measurement structure to that revealed by factor analysis. There were a large number of dissimilar associations with predictor variables suggesting incorrect inferences are likely if using the standard executive function index. The unpredictability of this

effect is demonstrated by the fact the ELSA memory index was actually very similar to the memory factor in this case. This builds upon past research showing through real world examples of how commonly used scoring systems for cognitive function may be a poor fit to the data being used.(Ashford et al., 1989; Brugnolo et al., 2009; Gibbons et al., 2002; McGrory et al., 2014; Mungas & Reed, 2000; Shigemori et al., 2010)

Using a standardised sum score for a battery of cognitive tests saves time for the researcher, but another important reason they might be preferred is that at face value it appears to provide the reassurance of consistency across studies. However, if the measurement properties for a given sample cannot be shown to actually be similar to those found in previous analyses then this is a false reassurance. Past research has demonstrated that the fact that you are using the same cognitive scoring system as another study, for example the mini-mental state exam, does not make it automatically comparable to other analyses using the same scoring system.(Brugnolo et al., 2009; Shigemori et al., 2010) This means that the reassurance from using the standardised score could be false reassurance, unless researchers publish on the measurement properties of the instrument they are using in their sample. Fortunately, with the expansion of publishing options and online appendices, this should be much easier to do in future.

If researchers wish to utilise previous evidence of the factor structure for a specific cognitive battery then it is possible to fix factor loadings to be the same as found previously and test to see if they fit their sample. Alternatively, in Bayesian structural equation modelling, evidence from previous analysis of a cognitive battery could be included in subsequent factor analyses by using informative priors.(Bengt Muthén & Asparouhov, 2012) As I was not aware of any prior factor analysis performed on the specific cognitive battery used in ELSA (which was derived from but not identical to test batteries used in previous studies) neither of these options were possible in this thesis.

As with the first paper presented, the second demonstrates the importance of taking the additional analytic steps to test the measurement properties of batteries of cognitive tests and addresses the problem of testing for longitudinal measurement invariance. I advanced the idea that Bayesian approximate measurement invariance can be used when commonly used criteria for alternative fit indices do not provide a definitive answer. This approach provided additional information to the standard tests and enabled a more informed decision to be made about whether to proceed with longitudinal analysis using the factors.(Bengt Muthén & Asparouhov, 2012; van de Schoot et al., 2013) Previous literature using approximate measurement invariance has focussed primarily on its application to accommodating multiple small non-invariances.(Bengt Muthén & Asparouhov, 2012; van de Schoot et al., 2013) To my knowledge this is the first paper to explicitly propose and demonstrate using approximate measurement invariance for this particular purpose.

Having established the measurement properties of the cognitive factors in ELSA and decided to use single tests I then moved on to longitudinal analysis of cognitive maintenance itself. I was

anticipating more difference between classes in the association of education and cognitive maintenance. The most similar previous analysis to my own by Muniz-Terrera et al. found a larger difference between classes than I did despite a similar modelling approach.(Muniz-Terrera et al., 2010) Clinical, as opposed to population based, research has tended to find that greater education is associated with more rapid decline following dementia diagnosis.(X. Meng & D'Arcy, 2012; Stern, Albert, Tang, & Tsai, 1999) Greater education has been associated with different patterns of neuronal activation in health and disease.(Colangeli et al., 2016) In healthy old age adults with higher cognitive reserve show activation of a complex bilateral frontoparietal network whereas in disease this has been found to be restricted to the left anterior cingulate.(Colangeli et al., 2016) This is thought to represent evidence for neural compensation reserve.(Cabeza et al., 2018; Colangeli et al., 2016; Stern et al., 2018) Under the predictions of neural compensation reserve the difference between individuals with more or less education should diminish as pathology advances. With only age-related changes present more educated individuals are able to draw upon this complex bilateral frontoparietal network. As pathology related changes progress this network is lost and only the anterior cingulate remains accessible. According to the neural compensation theory, as these secondary networks fail decline accelerates more rapidly in individuals with more education. This results in cognitive function converging towards those with less education who had started with poor baseline function but decline less rapidly. Without the loss of secondary networks to pathological change, neural compensation predicts little difference in cognitive maintenance between those with more or less education.

I was therefore anticipating more rapid decline in individuals with greater education in the declining class and little to no difference in the stable class. What I found was a small and statistically significant protective effect in the stable cognitive function class for verbal fluency, but no effect in the cognitively declining class or for recall. This provides only limited support for differing forms of cognitive reserve being present with different underlying trajectories. The main effect of education was on brain/passive reserve, with a substantial increase in baseline performance. This does not necessarily mean I have found evidence against education contributing to neural compensation or neural cognitive reserve. Given that both models have empirical support but generate opposing predictions, it is possible both mechanisms are operating with a net result of minimal differences in cognitive maintenance by education.(Stern, 2012) My findings support the overall conclusion that educational attainment has little substantively important association with cognitive maintenance.(Lenehan et al., 2015) This is consistent with a large number of studies in older adults from a wide range of studies showing small or no different in change in cognition by education level.(Lenehan et al., 2015) This includes analysis of the AHEAD study, Victoria Longitudinal Study, ARIC neurocognitive study and others.(H Christensen & Hofer, 2001; Gottesman et al., 2014; Karlamangla et al., 2009; Tucker-Drob et al., 2009; Zahodne et al., 2011b) Large co-ordinated multi-cohort analyses have also found that education is associated with improved baseline scores, but has little effect on cognitive maintenance.(Lipnicki et al., 2019; Piccinin et al., 2013) My analysis extends this

literature by demonstrating this applies whether the underlying trajectory is stable or declining. This was done by relaxing the assumption of population heterogeneity and allowing latent class to moderate the effect of education.

That is not to say there is not also a body of research which has found that education is associated with improved cognitive maintenance.(Beydoun et al., 2014) There could be several reasons for the discrepancy. Lenehan and colleagues point to the influence of earlier studies with fewer measurement occasions being more likely to find that education affects cognition over time.(Lenehan et al., 2015) An alternative explanation was proposed by Foverskov and colleagues from their analysis linking cohort study data to Danish national registry data.(Foverskov, Glymour, Mortensen, Holm, & Lange, 2018) They found education to be associated with slower cognitive decline and suggest that measurement error in self-reported education may lead to underestimation of the effect of education on cognitive maintenance. Zahodne et al. in their 2015 analysis used multiple indicator growth mixture modelling and found that more education was associated with improved cognitive maintenance.(Zahodne, Stern, & Manly, 2015) They suggest that the ability of confirmatory factor analysis to handle measurement error in cognitive function and provide more precise estimates may explain the difference in their results from the majority of recent studies. Nonetheless, the need for increased precision to identify a statistically significant effect from a cohort of several thousand participants demonstrates that the effect itself is small and of unclear substantive importance.

Multiple indicator growth as used by Zahodne, Stern and Manly was the approach I would have used, had it not been for the measurement non-invariance I identified.(Zahodne, Stern, et al., 2015) This statistical approach was born out of the desire to build a model which, although observational, still sought to represent the most likely causal structure underlying the data. This same motivation was developed into the fourth paper, however quite different methods were needed to examine the role of cognitively stimulating activities (CSA) in later life compared to the cognitive stimulation of education in earlier life. The application of my causal model led to two relatively distinctive features of the analysis. The first difference being the use of marginal structural models. That the association with social club membership and cognitive impairment which was lost in the IPTCW MSM demonstrates that time varying confounding affected by past exposure is a theoretical concern with real consequences in the study of social exposures and cognitive maintenance.

The second difference was the testing of individual CSA rather than a composite score which has been used in much of observational research to date.(Sajeev et al., 2016; Yates et al., 2016) This has potential theoretical implications for the understanding of whether general cognitive enrichment or specific activities are important for maintaining cognitive function in later life. Composites of multiple activities generate significant difficulties for the interpretation of the results of those studies. As all exposures are treated equally, it is not possible to know whether they are in fact equal or whether specific activities have a stronger association with cognitive maintenance than others. In this thesis, I show that some activities are associated with reduced

risk of cognitive impairment. Other activities which would usually be viewed as providing cognitive enrichment, such as attending evening classes, were not. This suggests that future research may wish to focus on specific cognitive activities, or at least to contrast specific activities with more general measures of activity.

9.4 Implications for Future Research

The use of applied examples of measuring cognitive ageing in this thesis could be useful to demonstrate why applied researchers might consider the use of factor analysis to increase the rigour of their use of complex measurement instruments. The first two papers showed how cognitive scores which have been used widely can fail to accurately represent the underlying data structure. Similar problems have previously been demonstrated with the mini-mental state exam and this may also apply to other commonly used tests. (Ashford et al., 1989; Brugnolo et al., 2009; Gibbons et al., 2002; McGrory et al., 2014; Mungas & Reed, 2000; Shigemori et al., 2010) Future researchers may wish to check cognitive scores from other publicly available datasets which have not undergone this scrutiny to date.

This can be done without adopting a full structural equation modelling (SEM) framework for the entirety of the analysis as shown in paper 1. SEM has great strength in its flexibility and statistical power. This flexibility comes at the price of increased modelling assumptions as more covariates are added.(VanderWeele, 2012) Specifically, structural equation modelling requires assumptions to be implicitly made about the relationship between covariates which do not need to be made in conventional regression analysis. This is a part of the reason for the number of covariates in the analysis in the third paper being kept to the minimum necessary. Whilst I wished to utilise the flexibility of SEM to identify latent classes of cognitive trajectory, I also wished to avoid making unrealistic assumptions regarding the relationships between my covariates or casual pathway from education to cognitive impairment. It is also why I used a different set of tools entirely for the fourth paper. In this complex longitudinal setting the number of potential mediating pathways would either have had to be unrealistically restricted or be unidentified. Marginal structural models only require the model for the exposure and outcome relationships to be specified correctly, rather than requiring the full covariance structure to be correctly specified.(VanderWeele, 2012) There is potential for a middle ground where factor or latent class analysis are integrated with the advantages of counterfactual causal inference.(B Muthén & Asparouhov, 2014) Further work is required to determine effective and theoretically coherent ways of combining structural equation modelling ability to estimate latent constructs with the well-defined inferential properties of marginal structural models.

In my own work, presented in the second paper, I demonstrated how Bayesian approximate measurement invariance can be integrated with frequentist tests using alternative fit indices. A natural extension of this would be to undertake simulation studies to identify the relative performance of these measures for a range of factor indicators, measurement occasions and

levels of measurement which may plausibly be seen in studies of cognitive maintenance. This could focus on the unbiased estimation of latent means and cognitive maintenance. It would be more challenging, but potentially as important, to simulate some of the other potential benefits of the Bayesian approach, such as the reduction in the capitalisation on chance and the effect of not having to fix one indicator variable to be invariant, as seen in the conventional frequentist approach.(R. C. MacCallum et al., 1992; Robert C. MacCallum et al., 2012) For example, one could simulate a scenario with indicator variables for a latent cognitive function and manipulate the degree of invariance, number of variables, number of occasions and sample size. It would then be possible to determine how often choosing each of the observed variables to be fixed to 1 leads to an incorrect conclusion about measurement invariance. Bayesian approximate measurement invariance could then be applied to the same scenarios and the relative performance of the approaches in correctly identifying non-invariance compared. By using plausible scenarios for studies of cognitive maintenance such studies could be made more directly relevant to informing researchers in this area. However, it would have broader implications for the assessment of measurement invariance across disciplines.

In my analysis of approximate measurement invariance, I also found changes in latent intercepts that were suggestive of variable practice effects in some of the individual cognitive tests. There are several ways of trying to identify and adjust for practice effects in tests of cognitive function, none of which are fully satisfactory.(Goldberg, Harvey, Wesnes, Snyder, & Schneider, 2015; Jones, 2015; Racine et al., 2018) It seems worth exploring whether approximate measurement invariance could be used as a means of testing for and potentially accommodating differential practice effects in longitudinal studies of cognitive functioning. This would not be able to identify practice effects occurring equally for all observed variables, but could account for variations in the degree of practice effect for different observed variables. Whilst not an anticipated outcome of the research questions addressed in this thesis, this could add to range of techniques available to researchers to manage problems with re-test effects when estimating cognitive maintenance.

Ultimately, I did not use longitudinal factors in my further analysis of cognitive maintenance in the third paper. The most straightforward extension of that work would be to domains other than memory. The large majority of studies using growth mixture models to study multiple classes of cognitive decline have used the mini-mental state examination or a measure of episodic memory, leaving open the possibility education may have a greater effect on the maintenance of other cognitive domains.(Baker et al., 2017; Ding et al., 2019; Downer, Chen, Raji, & Markides, 2017; Hayden et al., 2011; Hochstetler et al., 2016; Seonjoo Lee et al., 2018; Marioni et al., 2014; Min, 2018; Olaya, Bobak, Haro, & Demakakos, 2017; Pietrzak et al., 2014; Royall, Palmer, Chiodo, & Polk, 2014; Small & Bäckman, 2007; Terrera, Brayne, & Matthews, 2010; Zahodne et al., 2015) The fact my findings support the literature suggesting that education primarily contributes to passive brain reserve does not mean that education has no meaningful effect on other cognition related outcomes.(Lenehan et al., 2015) For example, higher levels of crystalized abilities or social and occupational skills may reduce rate of progression to cognitive

disability, even though longitudinal effects on cognitive maintenance are small.(Bendayan et al., 2017; Jokinen et al., 2016) A plausible model would have education is the primary exposure, disability the outcome, and cognition as a mediator of the effect of education on disability. Education would be expected to have a direct effect on risk of disability and an indirect effect via baseline cognitive functioning. This relationship may show an exposure-mediator interaction. This would mean that, as well as the direct and indirect effects, there may be an interaction between education and cognition on disability risk. Only relatively recently have these complex effect decompositions been formally defined in a counterfactual framework and accessible software written for use by applied researchers.(Discacciati, Bellavia, Lee, Mazumdar, & Valeri, 2019; VanderWeele, 2014; Vanderweele, Vansteelandt, & Robins, 2014)

As an application of the counterfactual approach to causal effects and non-parametric structural models to the study of cognitive ageing this type of complex effects decomposition would also build upon the work presented in the fourth paper.(VanderWeele, 2014) Wider implementation of the marginal structural model (MSM) approach used in this analysis could be applied to a range of questions about the effects of social exposures on cognitive maintenance. As demonstrated in the fourth paper, this would make a valuable contribution to understanding these relationships by relaxing the assumption that there is no time varying confounding affected by past exposure. In my research I found that the association between social club membership and dementia was no longer significant in the marginal structural models. Related to this, the relationship between social isolation, loneliness and depression with dementia risk is one example of where the question of multi-directional causal relationships is central and MSMs could offer further insight.(Ismail et al., 2016; Yin, Lassale, & Steptoe, 2019)

Further research is also needed into the exposures found to be associated with lower risk of dementia or cognitive impairment in this thesis. Volunteering appears to be a strong contender for a CSA likely to provide some reduction in risk of cognitive impairment and dementia.(Guiney & Machado, 2018) My findings demonstrate that this association is unlikely to be explained by time-varying confounding affected by past exposure. Nonetheless, unmeasured confounding may still play a role in the association. Additionally, volunteering in one societal context may be quite different from another. For this reason, further replication of this finding from datasets with different sets of confounders and drawn from different societal contexts is needed.

Volunteering itself can include highly heterogenous set of activities. I believe that it is the combination of cognitive and social stimulation alongside increased physical activities which is likely to be why volunteering appears to have its protective effect on cognitive maintenance.(Guiney & Machado, 2018) On this basis I would not advocate further research trying to sub-divide volunteering and identify individual active components. However, arguments could be made for this. An argument could be made that sub-diving the intervention is one way of aiming to deliver the most efficient intervention possible. An alternative approach would be to aim to identify whether specific volunteering activities are associated with greater benefits that others, or what duration and intensity of volunteer activity is required to produce cognitive

benefits. There is a precedent of small studies testing the effect of volunteering on cognition, such as the Experience Corps study.(M. C. Carlson et al., 2008) Larger and longer duration experimental studies randomising those who would not otherwise have engaged in volunteer activity would be the ideal way of demonstrating causation. However, even if all the logistical and ethical challenges to building a more substantial experimental evidence base could be overcome, there would still remain substantial problems with external validity. This means that observational studies will continue to have an important place in answering these questions.

General internet use was the other CSA found in my analysis to have an association with reduced risk of cognitive impairment and dementia. The majority of past research on computer use in older adults has focussed on targeted cognitive interventions. My findings are consistent with the smaller amount of research in the area of general computer use in finding a protective effect and extends this literature by demonstrating that this association is ELSA is not due to time varying confounding affected by prior exposure.(d'Orsi et al., 2017; Liapis & Harding, 2017; André J. Xavier et al., 2014) This suggests general internet use is a potential target for intervention to reduce the risk of cognitive impairment and dementia worth exploring in future research. However, as adults entering later life are becoming increasingly computer literate then this raises the important question regarding whether this will be possible to replicate in future cohorts. Whether this effect is seen cohorts with higher levels of computer literacy at baseline will be an important question for future research.

Intervention studies teaching general computer skills are the definitive way of demonstrating causality. These are very difficult to design with adequate sample size and duration. As information technology use continues to become more widespread reducing the effect of non-trial exposure would be difficult. Moreover, the interventions would need to be carefully designed to ensure that the participants who received the training were demonstrably putting those skills into regular use. This equally applies to other interventions which seek to promote new skills and their use. Both volunteering and internet use have in common that they require the acquisition and regular use of new skills. I believe that emphasis in both observational and experimental research should shift towards skill acquisition and sustained use, rather than activity participation alone.(Iizuka et al., 2019)

9.5 Policy Implications

Improving cognitive maintenance and thereby reducing the risk of mild cognitive impairment and dementia is a priority for public health and many older individuals.(Friedman et al., 2015; Livingston et al., 2017) This is the primary area of policy implications arising from this thesis relating to the 3rd and 4th papers presented. The policy relevance of the first two papers is primarily within the academic field. Taken together the first two articles demonstrate the importance of the evaluation of measurement instruments within a given dataset. Current commonly used appraisal criteria used to judge the methodological quality of research and

improve reporting, such as the STROBE criteria, do not include assessment of measurement properties of psychometric instruments.(von Elm et al., 2007, 2014) It may be time to consider whether a form of evaluation of the measurement properties of instruments used in observational research for cognitive function, or many other latent exposures and outcomes, should be considered a mark of quality of research and included formally in such guidelines or journal submission requirements.

The results from the third paper suggest that education is associated with little substantively important difference in cognitive maintenance in those with either a declining or stable trajectory. This is consistent with the view that education primarily contributes primarily to passive brain reserve.(Lenehan et al., 2015; Stern, 2012) In this case, reduced rates of cognitive impairment and faster progression following diagnosis seen in those with higher levels of education are likely to be due to delay in the development of impairment because of higher baseline functioning alone.(X. Meng & D'Arcy, 2012) On a population level, these findings suggest that the expansion in higher education seen in many western countries is likely to be associated with a long term decline in dementia incidence. This is already being observed in some countries.(Serrano-Pozo & Growdon, 2019; Stephan et al., 2018) However, as life expectancy increases then the effect may be a reduction in age-specific incidence at younger ages without affecting the whole population incidence. Nonetheless, this postponement of morbidity would still be a public health gain overall.

In the 4th paper the analysis found that volunteering and internet use specifically were protective against cognitive impairment. Evening classes, social clubs and working were not associated with reduced risk. This finding has direct implications for decisions regarding which CSA older adults may wish to choose or policymakers to promote if they wish to reduce the risk of cognitive impairment on an individual or population level. Analyses using methods designed for causal inference from observational data such as those presented in this thesis are particularly for important for providing policy evidence in the context of the difficulties in performing adequate experimental studies. It is encouraging that our findings are consistent with prior research on these specific CSA.(Guiney & Machado, 2018; Liapis & Harding, 2017) Whilst awaiting further research, public or individual action to promote volunteering and learn new computer skills appear to be good bets for the CSA most likely to reduce risk of cognitive impairment and dementia. With the results of my analysis in the context of prior research they can be cautiously recommended for this aim. The potential harms from volunteering these activities are largely restricted to opportunity costs, whilst the potential benefits to the individuals and wider society extend well beyond cognition alone.

The finding of no association with evening classes and social clubs may also have implications for practice. Social club membership and evening class attendance as reported by ELSA participants has no strict definition. There are likely to represent quite heterogeneous categories including a wide range of specific activities. This does limit the interpretation of these findings. This analysis finds that the typical type and intensity of evening class or social club activities

being undertaken by older adults in England is not sufficient to be associated with improved cognitive maintenance. Conceptually the marginal structural model is attempting to estimate the effects of fixing an individual's exposure to attendance or non-attendance of those evening classes or social clubs. Consequently, my analysis suggests that interventions aimed at improving older adult's attendance at existing social or educational activities are not be likely to be effective for improving cognition. Naturally, this does not mean there might not be a range of other beneficial non-cognitive outcomes of such an intervention. Nor does it mean that some social programmes or educational activities could not have beneficial effects on cognitive maintenance. The mix of these activities undertaken by ELSA participants is representative of the age-specific English population, so promotion of more of these activities for the prevention of cognitive impairment in the English or similar contexts is not supported by my analysis.

In conclusion, this thesis advances our knowledge of how cognitive maintenance can be measured and the effect of education and other cognitively stimulating activities on cognitive maintenance in later life. For the measurement of cognitive maintenance, I demonstrate how using pre-specified cognitive scores or factor analysis can lead to highly variable effects on later analysis and how Bayesian approximate measurement invariance can be applied to cognitive factors over time. I found that education provides primarily passive cognitive reserve, extending past research by showing that this applies latent classes of cognitively stable or declining older adults, Contrastingly, I found that specific cognitively stimulating activities in later life, volunteering and internet use, were associated with improved cognitive maintenance. This was shown by a reduced risk of dementia or cognitive impairment and contributed to the field by being the first analysis on the effect of CSA on cognition to use marginal structural models to account for reverse causation.

References

Aartsen, M. J., Smits, C. H. M., van Tilburg, T., Knipscheer, K. C. P. M., & Deeg, D. J. H. (2002). Activity in Older Adults: Cause or Consequence of Cognitive Functioning? A Longitudinal Study on Everyday Activities and Cognitive Performance in Older Adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *57*(2), P153–P162. https://doi.org/10.1093/geronb/57.2.P153

Aartsen, Marja J, Martin, M., & Zimprich, D. (2004). Gender Differences in Level and Change in Cognitive Functioning. *Gerontology*, *50*(1), 35–38. https://doi.org/10.1159/000074387

Aevarsson, O., & Skoog, I. (2000). A Longitudinal Population Study of the Mini-Mental State Examination in the Very Old : Relation to Dementia and Education. *Dementia and Geriatric Cognitive Disorders*, *11*(3), 166–175.

Akbaraly, T. N., Portet, F., Fustinoni, S., Dartigues, J. F., Artero, S., Rouaud, O., … Berr, C. (2009). Leisure activities and the risk of dementia in the elderly: Results from the Three-City Study. *Neurology*, *73*(11), 854–861. Retrieved from papers://61aea551-8d9f-4bd3-b82b-d7c3e55e6fa7/Paper/p4269

Alexander, M., Perera, G., Ford, L., Arrighi, H. M., Foskett, N., Debove, C., … Gordon, M. F. (2015). Age-Stratified Prevalence of Mild Cognitive Impairment and Dementia in European Populations: A Systematic Review. *Journal of Alzheimer's Disease*, *48*(2), 355–359. https://doi.org/10.3233/JAD-150168

Alvarado, B. E., Zunzunegui, M. V., Del Ser, T., & Béland, F. (2002). Cognitive decline is related to education and occupation in a Spanish elderly cohort. *Aging Clinical and Experimental Research*, *14*(2), 132–142. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-0036019028&partnerID=tZOtx3y1

Alzheimer's Association. (2014). 2014 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *10*(2), e47–e92. https://doi.org/10.1016/j.jalz.2014.02.001

Arenaza-Urquijo, E. M., & Vemuri, P. (2018). Resistance vs resilience to Alzheimer disease Clarifying terminology for preclinical studies. *Neurology*, *90*, 695–703. https://doi.org/10.1212/WNL.0000000000005303

Arnold, S. E., Louneva, N., Cao, K., Wang, L. S., Han, L. Y., Wolk, D. A., … Bennett, D. A. (2013). Cellular, synaptic, and biochemical features of resilient cognition in Alzheimer's disease. *Neurobiology of Aging*, *34*(1), 157–168. https://doi.org/10.1016/j.neurobiolaging.2012.03.004

Ashford, J. W., Kolm, P., Colliver, J. A., Bekian, C., & Hsu, L. N. (1989). Alzheimer patient evaluation and the Mini-Mental State: Item characteristic curve analysis. *Journals of Gerontology*, *44*(5).

Baker, E., Iqbal, E., Johnston, C., Broadbent, M., Shetty, H., Stewart, R., … Dobson, R. J. B.

(2017). Trajectories of dementia-related cognitive decline in a large mental health records derived patient cohort. *PLoS ONE*, 1–18.

Banks, James, Batty, G. D., Begum, N., Demakakos, P., Oliveira, C. de, Head, J., … Zaninotto, P. (2016). *The Dynamics of Ageing: Evidence from the English Longitudinal Study of Ageing*. (James Banks, G. D. Batty, J. Nazroo, & A. Steptoe, Eds.). London, UK: The Institute for Fiscal Studies. Retrieved from file:///C:/Users/benja/Documents/Work/research_library/ELSA wave 7 study report.pdf

Barendse, M. T., Oort, F. J., & Timmerman, M. E. (2014). Using Exploratory Factor Analysis to Determine the Dimensionality of Discrete Responses. *Structural Equation Modeling: A Multidisciplinary Journal*, *5511*(September), 1–15. https://doi.org/10.1080/10705511.2014.934850

Barnes, D. E., Haight, T. J., Mehta, K. M., Carlson, M. C., Kuller, L. H., & Tager, I. B. (2010). Secondhand smoke, vascular disease, and dementia incidence: findings from the cardiovascular health cognition study. *American Journal of Epidemiology*, *171*(3), 292–302. https://doi.org/10.1093/aje/kwp376

Barulli, D., & Stern, Y. (2013). Efficiency, capacity, compensation, maintenance, plasticity: Emerging concepts in cognitive reserve. *Trends in Cognitive Sciences*, *17*(10), 502–509. https://doi.org/10.1016/j.tics.2013.08.012

Beauducel, A., & Herzberg, P. Y. (2006). On the Performance of Maximum Likelihood Versus Means and Variance Adjusted Weighted Least Squares Estimation in CFA. *Structural Equation Modeling*, *13*(2), 204–228.

Bell, A., & Jones, K. (2013). The impossibility of separating age, period and cohort effects. *Social Science & Medicine (1982)*, *93*, 163–165. https://doi.org/10.1016/j.socscimed.2013.04.029

Bendayan, R., Piccinin, A. M., Hofer, S. M., Cadar, D., Johansson, B., & Muniz-terrera, G. (2017). Decline in memory, visuospatial ability, and crystalized cognitive abilities in older adults: normative aging or terminal decline? *Journal of Aging Research*, *2017*. https://doi.org/10.1155/2017/6210105

Bennett, D. A., Arnold, S. E., Valenzuela, M. J., Brayne, C., & Schneider, J. A. (2014). Cognitive and social lifestyle: links with neuropathology and cognition in late life. *Acta Neuropathologica*, *127*(1), 137–150. https://doi.org/10.1007/s00401-013-1226-2

Bennett, D. A., Wilson, R. S., Schneider, J. A., Evans, D. A., Mendes de Leon, C. F., Arnold, S. E., … Bienias, J. L. (2003). Education modifies the relation of AD pathology to level of cognitive function in older persons. *Neurology*, *60*(12), 1909–1915. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-0038715432&partnerID=tZOtx3y1

Beunckens, C., Molenberghs, G., Verbeke, G., & Mallinckrodt, C. (2008). A latent-class mixture

model for incomplete longitudinal Gaussian data. *Biometrics*, *64*(1), 96–105. https://doi.org/10.1111/j.1541-0420.2007.00837.x

Beydoun, M. A., Beydoun, H. A., Gamaldo, A. A., Teel, A., Zonderman, A. B., & Wang, Y. (2014). Epidemiologic studies of modifiable factors associated with cognition and dementia: systematic review and meta-analysis. *BMC Public Health*, *14*(1), 643.

Birks, J. (2006). Cholinesterase inhibitors for Alzheimer's disease (Review). *Cochrane Database of Systematic Reviews*, (1).

Bishop, J., Geiser, C., & Cole, D. a. (2015). Modeling Latent Growth With Multiple Indicators: A Comparison of Three Approaches. *Psychological Methods*, *20*(1), 43–62. https://doi.org/10.1037/met0000018

Blankson, a N., & McArdle, J. J. (2013). Measurement Invariance of Cognitive Abilities Across Ethnicity, Gender, and Time Among Older Americans. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *70*, 386–397. https://doi.org/10.1093/geronb/gbt106

Bodnar, L. M., Davidian, M., Siega-Riz, A. M., & Tsiatis, A. A. (2004). Marginal Structural Models for Analyzing Causal Effects of Time-dependent Treatments: An Application in Perinatal Epidemiology. *American Journal of Epidemiology*, *159*(10), 926–934. https://doi.org/10.1093/aje/kwh131

Bollen, K. A., & Curran, P. J. (2006). Unconditional Latent Curve Model. In *Latent Curve Models: A Structural Equation Perspective* (1st ed., pp. 16–57). Hoboken, NJ: John Wiley and Sons Inc.

Boller, B., Mellah, S., Ducharme-Laliberté, G., & Belleville, S. (2017). Relationships between years of education, regional grey matter volumes, and working memory-related brain activity in healthy older adults. *Brain Imaging and Behavior*, *11*, 304–317. https://doi.org/10.1007/s11682-016-9621-7

Bowler, J. V, Munoz, D. G., Merskey, H., & Hachinski, V. (1998). Factors affecting the age of onset and rate of progression of Alzheimer ' s disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *65*, 184–190.

Braak, H., & Del Tredici, K. (2015). The preclinical phase of the pathological process underlying sporadic Alzheimer's disease. *Brain*, *138*(10), 2814–2833. https://doi.org/10.1093/brain/awv236

Brayne, C., Ince, P. G., Keage, H. a D., McKeith, I. G., Matthews, F. E., Polvikoski, T., & Sulkava, R. (2010). Education, the brain and dementia: neuroprotection or compensation? *Brain : A Journal of Neurology*, *133*(Pt 8), 2210–2216. https://doi.org/10.1093/brain/awq185

Bridges, S., Hussey, D., & Blake, M. (2015). The dynamics of ageing : The 2012 English Longitudinal Study of Ageing (Wave 6) Technical Report, (5050), 1–100.

Brodaty, H., Mothakunnel, A., de Vel-Palumbo, M., Ames, D., Ellis, K. a, Reppermund, S., … Sachdev, P. S. (2014). Influence of population versus convenience sampling on sample characteristics in studies of cognitive aging. *Annals of Epidemiology*, *24*(1), 63–71. https://doi.org/10.1016/j.annepidem.2013.10.005

Brugnolo, a, Nobili, F., Barbieri, M. P., Dessi, B., Ferro, A., Girtler, N., … Rodriguez, G. (2009). The factorial structure of the mini mental state examination (MMSE) in Alzheimer's disease. *Archives of Gerontology and Geriatrics*, *49*(1), 180–185.

Byrd, D. a, Touradji, P., Tang, M.-X., & Manly, J. J. (2004). Cancellation test performance in African American, Hispanic, and White elderly. *Journal of the International Neuropsychological Society : JINS*, *10*(3), 401–411.

Cabeza, R., Albert, M., Belleville, S., Craik, F. I. M., Duarte, A., Grady, C. L., … Rajah, N. (2018). Maintainance, reserve and compensation: the cognitive neuroscidence of healthy ageing. *Nature Reviews Neuroscience*, (19), 701–710. https://doi.org/10.1038/s41583-018-0068-2

Cabeza, R., Albert, M., Belleville, S., Craik, F. I. M., Duarte, A., Grady, C. L., … Steffener, J. (2019). Reply to ' Mechanisms underlying resilience in ageing .' *Nature Reviews Neuroscience*, *20*(april), 41583. https://doi.org/10.1038/s41583-019-0139-z

Calamia, M., Markon, K., & Tranel, D. (2012). Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist*, *26*(4), 543–570. https://doi.org/10.1080/13854046.2012.680913

Carlson, M. C., Saczynski, J. S., Rebok, G. W., Seeman, T., Glass, T. A., McGill, S., … Fried, L. P. (2008). Exploring the Effects of an "Everyday" Activity Program on Executive Function and Memory in Older Adults: Experience Corps. *The Gerontologist*, *48*(6), 793–801. https://doi.org/10.1093/geront/48.6.793

Carlson, Michelle C., Kuo, J. H., Chuang, Y. F., Varma, V. R., Harris, G., Albert, M. S., … Fried, L. P. (2015). Impact of the Baltimore Experience Corps Trial on cortical and hippocampal volumes. *Alzheimer's and Dementia*, *11*(11), 1340–1348. https://doi.org/10.1016/j.jalz.2014.12.005

Casey E. Krueger; Joel H. Kramer. (2010). Neurocognitive Assesment. *Continuum Lifelong Learning Neurol*, *16*, 176–190. https://doi.org/10.1039/c1an15761b

Castellano, J. M., Kim, J., Stewart, F. R., Jiang, H., Demattos, R. B., Patterson, B. W., … Holtzman, D. M. (2011). Human apoE Isoforms Differentially Regulate Brain Amyloid- b Peptide Clearance, *3*(89).

Castro-Costa, E., Dewey, M. E., Uchôa, E., Firmo, J., Lima-costa, M. F., & Stewart, R. (2011). Trajectories of cognitive decline over 10 years in a Brazilian elderly population : the Bambuí Cohort Study of Aging. *Cad. Saude Publica*, *27*(Suppl 3), s345-350.

Catheline, G., Amieva, H., Dilharreguy, B., Bernard, C., Duperron, M.-G., Helmer, C., … Allard, M. (2015). Semantic retrieval over time in the aging brain: Structural evidence of hippocampal contribution. *Hippocampus*, *1016*(25), 1008–1016.

Cavedo, E., Odile, M., & Lamari, F. (2017). Preclinical Alzheimer ' s disease : A systematic review of the cohorts underlying the concept, *13*, 454–467. https://doi.org/10.1016/j.jalz.2016.12.003

Cepeda, N. J., Blackwell, K. A., & Munakata, Y. (2013). Speed isn't everything: Complex processing speed measures mask individual differences and developmental changes in executive control. *Developmental Science*, *16*(2), 269–286.

Chan, D. (1998). The Conceptualiszation and Analysis of Change Over Time: An Integrative Approach Incorporating Longitudinal Mean and Covariance Structures Analysis (LMACS) and Multiple Indicator Latent Growth Modeling (MLGM). *Organizational Research Methods*, *1*(4), 421–483. https://doi.org/0803973233

Channon, S., Baker, J. E., & Robertson, M. M. (1993). Working memory in clinical depression: an experimental study. *Psychological Medicine*, *23*, 87–91.

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, Q., & Ibrahim, J. G. (2014). A note on the relationships between multiple imputation, maximum likelihood and fully Bayesian methods for missing responses in linear regression models. *Statistics and Its Interface*, *6*(3), 315–324. https://doi.org/10.4310/SII.2013.v6.n3.a2

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of- Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902

Chou, C., Bentler, P. M., & Pentz, M. A. (1998). Comparisons of two statistical approaches to study growth curves: The multilevel model and the latent curve analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *5*(3), 247–266. https://doi.org/10.1080/10705519809540104

Chow, T. W., Hynan, L. S., & Lipton, A. M. (2006). MMSE scores decline at a greater rate in frontotemporal degeneration than in AD. *Dementia and Geriatric Cognitive Disorders*, *22*(3), 194–199.

Christensen, H, & Hofer, S. (2001). Age is no kinder to the better educated: absence of an association investigated using latent growth techniques in a community sample. *Psychological …*, 15–28. Retrieved from http://journals.cambridge.org/abstract_S0033291799002834

Christensen, H, Korten, A. ., Jorm, A. ., Henderson, A. ., Jacomb, P. ., & Rodgers, B. (1997). Education and Decline in Cognitive Performance: Compensatory but not Protective. *International Journal of Geriatric Psychiatry*, *12*(February 1996), 323–330.

Christensen, Helen, Batterham, P. J., Mackinnon, A. J., Anstey, K. J., Wen, W., & Sachdev, P. S. (2009). Education , Atrophy , and Cognitive Change in an Epidemiological Sample in Early Old Age. *American Journal of Geriatric Psychiatry*, *17*(3), 218–226.

Christensen, Helen, Batterham, P. J., Mackinnon, A. J., Jorm, A. F., Mack, H. a, Mather, K. a, … Easteal, S. (2008). The association of APOE genotype and cognitive decline in interaction with risk factors in a 65-69 year old community sample. *BMC Geriatrics*, *8*, 14. https://doi.org/10.1186/1471-2318-8-14

Clarke, C. S., Round, J., Morris, S., Kharicha, K., Ford, J., Manthorpe, J., … Walters, K. (2017). Exploring the relationship between frequent internet use and health and social care resource use in a community-based cohort of older adults: An observational study in primary care. *BMJ Open*, *7*(7). https://doi.org/10.1136/bmjopen-2017-015839

Clarke, P. J., Ailshire, J. a, House, J. S., Morenoff, J. D., King, K., Melendez, R., & Langa, K. M. (2012). Cognitive function in the community setting: the neighbourhood as a source of “cognitive reserve”? *Journal of Epidemiology and Community Health*, *66*(8), 730–736. https://doi.org/10.1136/jech.2010.128116

Clouston, S. A. P., & Denier, N. (2017a). Mental retirement and health selection: Analyses from the U.S. Health and Retirement Study. *Social Science and Medicine*, *178*, 78–86. https://doi.org/10.1016/j.socscimed.2017.01.019

Clouston, S. A. P., & Denier, N. (2017b). Mental retirement and health selection: Analyses from the U.S. Health and Retirement Study. *Social Science and Medicine*, *178*, 78–86. https://doi.org/10.1016/j.socscimed.2017.01.019

Colangeli, S., Boccia, M., Verde, P., Guariglia, P., Bianchini, F., & Piccardi, L. (2016). Cognitive Reserve in Healthy Aging and Alzheimer’s Disease. *American Journal of Alzheimer’s Disease & Other Dementias®*, *31*(5), 443–449. https://doi.org/10.1177/1533317516653826

Cole, S. R., & Hernán, M. a. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, *168*(6), 656–664. https://doi.org/10.1093/aje/kwn164

Cona, G., Scarpazza, C., Sartori, G., Moscovitch, M., & Bisiacchi, P. S. (2015). Neural bases of

prospective memory: A meta-analysis and the “Attention to Delayed Intention” (AtoDI) model. *Neuroscience & Biobehavioral Reviews*, *52*, 21–37.

Cortés, N., Andrade, V., & Maccioni, R. B. (2018). Behavioral and Neuropsychiatric Disorders in Alzheimer’s Disease. *Journal of Alzheimer’s Disease*, *63*(3), 899–910. https://doi.org/10.3233/JAD-180005

Cossa, F. M., Sala, S. Della, & Spinnler, H. (1995). Alzheimer patients know their date of birth but not their age: a study on disorientation. *International Journal of Geriatric Psychiatry*, *10*, 99–106.

Cotman, C. W., Berchtold, N. C., & Christie, L. (2007). Exercise builds brain health : key roles of growth factor cascades and inflammation, *30*(9), 7–9. https://doi.org/10.1016/j.tins.2007.06.011

Crimmins, E. M., Kim, J. K., Langa, K. M., & Weir, D. R. (2011). Assessment of cognition using surveys and neuropsychological assessment: the Health and Retirement Study and the Aging, Demographics, and Memory Study. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *66 Suppl 1*, i162–i171. https://doi.org/10.1093/geronb/gbr048

Curran, P. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, *38*(4), 37–41. https://doi.org/10.1207/s15327906mbr3804

Cutler, S. J. (2015). Worries about getting Alzheimer’s: Who’s concerned? *American Journal of Alzheimer’s Disease and Other Dementias*, *30*(6), 591–598. https://doi.org/10.1177/1533317514568889

d’Orsi, E., Xavier, A. J., Rafnsson, S. B., Steptoe, A., Hogervorst, E., & Orrell, M. (2017). Is use of the internet in midlife associated with lower dementia incidence? Results from the English Longitudinal Study of Ageing. *Aging and Mental Health*, *0*(0), 1–9. https://doi.org/10.1080/13607863.2017.1360840

Daniel, R. M., Cousens, S. N., De Stavola, B. L., Kenward, M. G., & Sterne, J. a C. (2013). Methods for dealing with time-dependent confounding. *Statistics in Medicine*, *32*(9), 1584–1618. https://doi.org/10.1002/sim.5686

de Azeredo Passos, V. M., Giatti, L., Bensenor, I., Tiemeier, H., Ikram, M. A., de Figueiredo, R. C., … Barreto, S. M. (2015). Education plays a greater role than age in cognitive test performance among participants of the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). *BMC Neurology*, *15*(1), 191.

de Medeiros, K. De, Mosby, A., Hanley, K. B., Pedraza, M. S., & Brandt, J. (2011). A randomized clinical trial of a writing workshop intervention to improve autobiographical memory and well-being in older adults, 803–811. https://doi.org/10.1002/gps.2605

Del Ser, T., Hachinski, V., Merskey, H., & Munoz, D. G. (1999). An autopsy-verified study of the effect of education on degenerative dementia. *Brain*, *122*, 2309–2319.

Depaoli, S., Yang, Y., & Felt, J. (2017). Using Bayesian Statistics to Model Uncertainty in Mixture Models: A Sensitivity Analysis of Priors. *Structural Equation Modeling*, *24*(2), 198–215.

Diggle, P., & Kenward, M. G. (1994). Informative Dropout in Longitudinal Data Analysis (with discussion). In *Applied Statistics* (pp. 49–73).

Ding, X., Charnigo, R. J., Schmitt, F. A., Kryscio, R. J., Abner, E. L., & Neuroimaging, D. (2019). Evaluating trajectories of episodic memory in normal cognition and mild cognitive impairment : Results from ADNI. *PLoS ONE*, 1–16.

Discacciati, A., Bellavia, A., Lee, J. J., Mazumdar, M., & Valeri, L. (2019). Med4way : a Stata command to investigate mediating and interactive mechanisms using the four-way effect decomposition, (November 2018), 15–20. https://doi.org/10.1093/ije/dyy236

Downer, B., Chen, N., Raji, M., & Markides, K. S. (2017). A longitudinal study of cognitive trajectories in Mexican Americans age 75 and older. *International Journal of Geriatric Psychiatry*, *32*, 1122–1130. https://doi.org/10.1002/gps.4575

Dubois, B., Feldman, H. H., Jacova, C., Hampel, H., Molinuevo, J. L., Blennow, K., … De, H. (2014). Advancing research diagnostic criteria for Alzheimer ' s disease : the IWG-2 criteria, *13*(June). https://doi.org/10.1016/S1474-4422(14)70090-0

Dubois, B., Hampel, H., Feldman, H. H., Scheltens, P., Aisen, P., Andrieu, S., … Jack, C. R. (2016). Preclinical Alzheimer ' s disease : Definition , natural history , and diagnostic criteria. *Alzheimer's & Dementia*, *12*, 292–323. https://doi.org/10.1016/j.jalz.2016.02.002

Durazzo, T. C., Mattsson, N., & Weiner, M. W. (2014). Smoking and increased Alzheimer ' s disease risk : A review of. *Alzheimer's & Dementia*, *10*(3), S122–S145. https://doi.org/10.1016/j.jalz.2014.04.009

Emerson, J. A., Smith, C. Y., Long, K. H., Ransom, J. E., Roberts, R. O., Hass, S. L., … Leibson, C. L. (2017). Nursing Home Use Across The Spectrum of Cognitive Decline: Merging Mayo Clinic Study of Aging With CMS MDS Assessments. *Journal of the American Geriatrics Society*, *65*(10), 2235–2243. https://doi.org/10.1111/jgs.15022

Enders, C. K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods*, *16*(1), 1–16. https://doi.org/10.1037/a0022640

Evans, D. A., Beckett, L. A., Albert, M. S., Hebert, L. E., Scherr, P. A., Funkenstein, H. H., & Taylor, J. (1993). Level of Education and Change in Cognitive Function in a Community Population of Older Persons.

Farfel, J. M., Nitrini, R., Suemoto, C. K., Grinberg, L. T., Ferretti, R. E. L., Leite, R. E. P., …

Jacob Filho, W. (2013). Very low levels of education and cognitive reserve: a clinicopathologic study. *Neurology*, *81*(7), 650–657. https://doi.org/10.1212/WNL.0b013e3182a08f1b

Ferreira, L., Ferreira Santos-Galduróz, R., Ferri, C. P., & Fernandes Galduróz, J. C. (2014). Rate of cognitive decline in relation to sex after 60 years-of-age: A systematic review. *Geriatrics and Gerontology International*, *14*(1), 23–31.

Ferrer, E., Balluerka, N., & Widaman, K. F. (2008). Factorial Invariance and The Specification of Second-Order Latent Growth Models. *Methodology*, *4*(1), 22–36. https://doi.org/10.1027/1614-2241.4.1.22.Factorial

Fewell, Z., Hernán, M. A., Wolfe, F., Tilling, K., Choi, H., & Sterne, J. A. C. (2004). Controlling for time-dependent confounding using marginal structural models. *The Stata Journal*, *4*(4), 402–420.

Fischer, F. U., Wolf, D., & Fellgiebel, A. (2019). Connectivity and morphology of hubs of the cerebral structural connectome are associated with brain resilience in AD- and age-related pathology. *Brain Imaging and Behavior*. https://doi.org/https://doi.org/10.1007/s11682-019-00090-y

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466

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.

Foverskov, E., Glymour, M. M., Mortensen, E. L., Holm, A., & Lange, T. (2018). Practice of Epidemiology Education and Cognitive Aging : Accounting for Selection and Confounding in Linkage of Data From the Danish Registry and Survey of Health , Ageing and Retirement in Europe, *187*(11), 2423–2430. https://doi.org/10.1093/aje/kwy162

Franco-marina, F., Garc, J. J., Wagner-echeagaray, F., Gallo, J., Ugalde, O., Sergio, S., … Ju, T. (2019). The Mini-mental State Examination revisited : ceiling and floor effects after score adjustment for educational level in an aging, (2010), 72–81. https://doi.org/10.1017/S1041610209990822

Friedman, D. B., Becofsky, K., Anderson, L. A., Bryant, L. L., Hunter, R. H., Ivey, S. L., … Lin, S. Y. (2015). Public perceptions about risk and protective factors for cognitive health and impairment: A review of the literature. *International Psychogeriatrics*, *27*(8), 1263–1275. https://doi.org/10.1017/S1041610214002877

García-Casal, J. A., Loizeau, A., Csipke, E., Franco-Martín, M., Perea-Bartolomé, M. V., & Orrell, M. (2017). Computer-based cognitive interventions for people living with dementia:

a systematic literature review and meta-analysis. *Aging & Mental Health*, *21*(5), 454–467. https://doi.org/10.1080/13607863.2015.1132677

Giannakopoulos, P., Gold, G., Duc, M., & Patrick, J. M. (2000). Neural substrates of spatial and temporal disorientation in Alzheimer ' s disease. *Acta Neuropathologica*, 189–195.

Gibbons, L. E., van Belle, G., Yang, M., Gill, C., Brayne, C., Huppert, F. A., … Larson, E. (2002). Cross-cultural comparison of the Mini-Mental State examination in United Kingdom and United States participants with Alzheimer's disease. *International Journal of Geriatric Psychiatry*, *17*(8), 723–728. https://doi.org/10.1002/gps.683

Global Council On Brain Health. (2017). *Engage Your Brain: GCBH Recommendations on Cognitively Stimulating Activities*. https://doi.org/DOI: https://doi.org/10.26419/pia.00001.001

Goldberg, T. E., Harvey, P. D., Wesnes, K. A., Snyder, P. J., & Schneider, L. S. (2015). Practice effects due to serial cognitive assessment: Implications for preclinical Alzheimer's disease randomized controlled trials. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, *1*(1), 103–111. https://doi.org/10.1016/j.dadm.2014.11.003

Gottesman, R. F., Rawlings, A. M., Sharrett, a R., Albert, M., Alonso, A., Bandeen-Roche, K., … Mosley, T. H. (2014). Impact of differential attrition on the association of education with cognitive change over 20 years of follow-up: the ARIC neurocognitive study. *American Journal of Epidemiology*, *179*(8), 956–966. https://doi.org/10.1093/aje/kwu020

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

Guerrero-Berroa, E., Luo, X., Schmeidler, J., Rapp, M. A., Dahlman, K., Grossman, H., … Beeri, M. S. (2009). The MMSE orientation for time domain is a strong predictor of subsequent cognitive decline in the elderly. *International Journal of Geriatric Psychiatry*, *24*(12), 1470–1477.

Guiney, H., & Machado, L. (2018). Volunteering in the Community: Potential Benefits for Cognitive Aging. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *73*(3), 399–408. https://doi.org/10.1093/geronb/gbx134

Guye, S., & von Bastian, C. C. (2017). Working memory training in older adults: Bayesian evidence supporting the absence of transfer. *Psychology and Aging*, *32*(8), 732–746. https://doi.org/10.1037/pag0000206

Hancock, G. R., Kuo, W.-L., & Lawrence, F. R. (2001). Teacher's Corner: An Illustration of Second-Order Latent Growth Models. *Structural Equation Modeling*, *8*(3), 470–489. https://doi.org/10.1207/S15328007SEM0803

Hayden, K. M., Reed, B. R., Manly, J. J., Tommet, D., Pietrzak, R. H., Chelune, G. J., … Jones, R. N. (2011). Cognitive decline in the elderly: An analysis of population heterogeneity. *Age and Ageing*, *40*(6), 684–689. https://doi.org/10.1093/ageing/afr101

Helm, J. L., Castro-Schilo, L., & Oravecz, Z. (2017). Bayesian Versus Maximum Likelihood Estimation of Multitrait–Multimethod Confirmatory Factor Models. *Structural Equation Modeling*, *24*(1), 17–30. https://doi.org/10.1080/10705511.2016.1236261

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology*, *18*(2), 284–295.

Hernán, M. A., Brumback, B. A., & Robins, J. M. (2002). Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. *Statistics in Medicine*, *21*(12), 1689–1709. https://doi.org/10.1002/sim.1144

Hikichi, H., Kondo, K., Takeda, T., & Kawachi, I. (2017). Social interaction and cognitive decline: Results of a 7-year community intervention. *Alzheimer's and Dementia: Translational Research and Clinical Interventions*, *3*(1), 23–32. https://doi.org/10.1016/j.trci.2016.11.003

Hill, N. T. M., Mowszowski, L., Naismith, S. L., Chadwick, V. L., Valenzuela, M., & Lampit, A. (2017). Computerized cognitive training in older adults with mild cognitive impairment or dementia: A systematic review and meta-analysis. *American Journal of Psychiatry*, *174*(4), 329–340. https://doi.org/10.1176/appi.ajp.2016.16030360

Hirni, D. I., Kivisaari, S. L., Monsch, A. U., & Taylor, K. I. (2013). Distinct neuroanatomical bases of episodic and semantic memory performance in Alzheimer's disease. *Neuropsychologia*, *51*(5), 930–937.

Hirono, N., Mori, E., Ishii, K., Ikejiri, Y., Imamura, T., Shimomura, T., … Sasaki, M. (1998). Hypofunction in the posterior cingulate gyrus correlates with disorientation for time and place in Alzheimer's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *64*(4), 552–554.

Hochstetler, H., Trzepacz, P. T., Wang, S., Yu, P., Case, M., & Bayview, J. H. (2016). Empirically Defining Trajectories of Late-Life Cognitive and Functional Decline, *50*, 271–282. https://doi.org/10.3233/JAD-150563

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*(3–4), 117–144. https://doi.org/10.1080/03610739208253916

Hosking, D. E., Sargent-Cox, K. A., & Anstey, K. J. (2015). An Australian survey of cognitive health beliefs, intentions, and behaviours through the adult life course. *Preventive Medicine Reports*, *2*, 498–504. https://doi.org/10.1016/j.pmedr.2015.06.008

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis:

Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Huppert, F. A., Gardener, E., & McWilliams, B. (2006). Cognitive function. In J Banks, E. Breeze, C. Lessof, & J. Nazroo (Eds.), *Retirement, health and relationships of the older population in England; The 2004 English Longitudinal Study of Ageing* (pp. 217–230). London, UK: The Institute for Fiscal Studies.

Iizuka, A., Suzuki, H., Ogawa, S., Kobayashi-cuya, K. E., Kobayashi, M., Takebayashi, T., & Fujiwara, Y. (2019). Can cognitive leisure activity prevent cognitive decline in older adults ? A systematic review of intervention studies, 469–482. https://doi.org/10.1111/ggi.13671

Infurna, F. J., Okun, M. A., & Grimm, K. J. (2016). Volunteering Is Associated with Lower Risk of Cognitive Impairment. *Journal of the American Geriatrics Society*, *64*(11), 2263–2269. https://doi.org/10.1111/jgs.14398

Institute for Digital Research and Education. (2009). Regresssion Diagnostics. In *Regression with Stata Web-Book*. Los Angeles, CA: UCLA. Retrieved from https://stats.idre.ucla.edu/stata/webbooks/reg/chapter2/stata-webbooksregressionwith-statachapter-2-regression-diagnostics/

Ismail, Z., Smith, E. E., Geda, Y., Sultzer, D., Brodaty, H., Smith, G., … Sciences, B. (2016). Neuropsychiatric symptoms as early manifestations of emergent dementia: Provisional diagnostic criteria for mild behavioural impairment. *Alzheimer's & Dementia*, *12*(2), 195–202. https://doi.org/10.1016/j.jalz.2015.05.017.Neuropsychiatric

Jacobs, D., Sano, M., Marder, K., Bell, K., Bylsma, F., Lafleche, G., … Stern, Y. (1994). Age at onset of Alzheimer's disease: Relation to pattern of cognitive dysfunction and rate of decline. *Neurology*, *44*(7), 1215–1220. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-0027930338&partnerID=40&md5=2f3a16731fa5d148f3512198b30b7ee2

Jagust, W. J., & Mormino, E. C. (2011). Lifespan brain activity , b -amyloid , and Alzheimer ' s disease. *Trends in Cognitive Sciences*, *15*(11), 520–526. https://doi.org/10.1016/j.tics.2011.09.004

Jefferson, A. L., Cosentino, S. a, Ball, S. K., Bogdanoff, B., Leopold, N., Kaplan, E., & Libon, D. J. (2002). Errors produced on the mini-mental state examination and neuropsychological test performance in Alzheimer's disease, ischemic vascular dementia, and Parkinson's disease. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *14*, 311–320.

Jenkinson, C. E., Dickens, A. P., Jones, K., Thompson-Coon, J., Taylor, R. S., Rogers, M., … Richards, S. H. (2013). Is volunteering a public health intervention? A systematic review and meta-analysis of the health and survival of volunteers. *BMC Public Health*, *13*, 773.

https://doi.org/10.1186/1471-2458-13-773

Jokinen, H., Melkas, S., Madureira, S., Verdelho, A., Ferro, J. M., Fazekas, F., … Erkinjuntti, T. (2016). Cognitive reserve moderates long-term cognitive and functional outcome in cerebral small vessel disease. *Journal of Neurology, Neurosurgery and Psychiatry*, *87*(12), 1296–1302. https://doi.org/10.1136/jnnp-2016-313914

Jones, R. N. (2015). Practice and retest effects in longitudinal studies of cognitive functioning. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, *1*(1), 101–102. https://doi.org/10.1016/j.dadm.2015.02.002

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsychology Review*, *17*, 213–233.

Kalmijn, S., Feskens, E. J. M., Launer, L. J., Kromhout, D., Bmj, S., Medical, B., … Pd, W. (1997). Longitudinal Study of the Effect of Apolipoprotein e4 Allele on the Association between Education and Cognitive Decline in Elderly Men. *BMJ*, *314*(7073), 34–35.

Kane, R. L., Butler, M., Fink, H. A., Brasure, M., Davila, H., Desai, P., … Barclay, T. (2017). *Interventions to Prevent Age-Related Cognitive Decline, Mild Cognitive Impairment, and Clinical Alzheimer's-Type Dementia*. *Agency for Healthcare Research and Quality (US)* (Vol. AHRQ Publi). Rockville, MD. https://doi.org/https://doi.org/10.23970/AHRQEPCCER188.

Karbach, J., & Verhaeghen, P. (2014). Making Working Memory Work: A Meta-Analysis of Executive-Control and Working Memory Training in Older Adults. *Psychological Science*, *25*(11), 2027–2037. https://doi.org/10.1177/0956797614548725

Karlamangla, A. S., Miller-Martinez, D., Aneshensel, C. S., Seeman, T. E., Wight, R. G., & Chodosh, J. (2009). Trajectories of cognitive function in late life in the United States: demographic and socioeconomic predictors. *American Journal of Epidemiology*, *170*(3), 331–342. https://doi.org/10.1093/aje/kwp154

Kittner, S. J., White, L. R., Farmer, M. E., Wolz, M., Kaplan, E., Moes, E., … Feinleib, M. (1986). Methodological issues in screening for dementia: The problem of education adjustment. *Journal of Chronic Diseases*, *39*(3), 163–170. https://doi.org/10.1016/0021-9681(86)90019-6

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

Koepsell, T. D., Kurland, B. F., Harel, O., Johnson, E. A., Zhou, X.-H., & Kukull, W. A. (2008). Education, cognitive function, and severity of neuropathology in Alzheimer disease. *Neurology*, *70*(19 Pt 2), 1732–1739. https://doi.org/10.1212/01.wnl.0000284603.85621.aa

Kokmen, E., Beard, C. M., Brien, P. C., & Roccu, W. A. (1996). Impact of Referral Bias on Clinical and Epidemiological Studies of Alzheimer ' s Disease, *49*(1), 79–83.

Koncz, R., & Sachdev, P. S. (2018). Are the brain ' s vascular and Alzheimer pathologies additive or interactive ?, 147–152. https://doi.org/10.1097/YCO.0000000000000395

Koster, A., Penninx, B. W. J. H., Bosma, H., Kempen, G. I. J. M., Newman, A. B., Rubin, S. M., … Kritchevsky, S. B. (2005). Socioeconomic differences in cognitive decline and the role of biomedical factors. *Annals of Epidemiology*, *15*(8), 564–571. https://doi.org/10.1016/j.annepidem.2005.02.008

Lampit, A., Hallock, H., & Valenzuela, M. (2014). Computerized Cognitive Training in Cognitively Healthy Older Adults: A Systematic Review and Meta-Analysis of Effect Modifiers. *PLoS Medicine*, *11*(11). https://doi.org/10.1371/journal.pmed.1001756

Lang, I., Wallace, R. B., Huppert, F. A., & Melzer, D. (2007). Moderate alcohol consumption in older adults is associated with better cognition and well-being than abstinence. *Age An*, *36*(March), 256–261. https://doi.org/10.1093/ageing/afm001

Langa, K. M., Kabeto, M. U., & Weir, D. R. (2010). *Report on race and cognitive impairment using HRS, in 2010 Alzheimer's disease facts and figures.* Retrieved from http://www.alz.org/documents_custom/report_alzfacts figures2010.pdf

Langa, K. M., Larson, E. B., Crimmins, E. M., Faul, J. D., Levine, D. A., Kabeto, M. U., & Weir, D. R. (2017). A comparison of the prevalence of dementia in the United States in 2000 and 2012. *JAMA Internal Medicine*, *177*(1), 51–58. https://doi.org/10.1001/jamainternmed.2016.6807

Langa, K. M., Llewellyn, D. J., Lang, I. a, Weir, D. R., Wallace, R. B., Kabeto, M. U., & Huppert, F. a. (2009). Cognitive health among older adults in the United States and in England. *BMC Geriatrics*, *9*, 23. https://doi.org/10.1186/1471-2318-9-23

Leckie, G., & Charlton, C. (2013). runmlwin - A Program to Run the MLwiN Multilevel Modelling Software from within Stata. *Journal of Statistical Software*, *52*(11), 1–40.

Lee, S., Kawachi, I., Berkman, L. F., & Grodstein, F. (2003). Education, Other Socioeconomic Indicators, and Cognitive Function. *American Journal of Epidemiology*, *157*(8), 712–720. https://doi.org/10.1093/aje/kwg042

Lee, S. Y., Song, X. Y., & Tang, N. S. (2007). Bayesian methods for analyzing structural equation models with covariates, interaction, and quadratic latent variables. *Structural Equation Modeling*, *14*(3), 404–434. https://doi.org/10.1080/10705510701301511

Lee, Seonjoo, Zhou, X., Gao, Y., Vardarajan, B., Reyes-, D., Rajan, K. B., … Id, S. B. (2018). Episodic memory performance in a multi- ethnic longitudinal study of 13,037 elderly. *PLoS ONE*, *13*(11), 1–17. https://doi.org/https://doi.org/10.1371/journal. pone.0206803

Lee, Sunmin, Buring, J. E., Cook, N. R., & Grodstein, F. (2006). The relation of education and income to cognitive function among professional women. *Neuroepidemiology*, *26*(2), 93–101. https://doi.org/10.1159/000090254

Lenehan, M. E., Summers, M. J., Saunders, N. L., Summers, J. J., & Vickers, J. C. (2015). Relationship between education and age-related cognitive decline: a review of recent research. *Psychogeriatrics*, (15), 154–162.

Leonards, U., Sunaert, S., Van Hecke, P., & Orban, G. a. (2000). Attention mechanisms in visual search -- an fMRI study. *Journal of Cognitive Neuroscience*, *12 Suppl 2*, 61–75.

Leopold, L., Engelhardt, H., & Engelhartdt, H. (2013). Education and physical health trajectories in old age. Evidence from the Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Public Health*, *58*(1), 23–31. https://doi.org/10.1007/s00038-012-0399-0

Leoutsakos, J. M. S., Muthen, B. O., Breitner, J. C. S., & Lyketsos, C. G. (2012). Effects of non-steroidal anti-inflammatory drug treatments on cognitive decline vary by phase of pre-clinical Alzheimer disease: Findings from the randomized controlled Alzheimer's Disease Anti-inflammatory Prevention Trial. *International Journal of Geriatric Psychiatry*, *27*(4), 364–374. https://doi.org/10.1002/gps.2723

Liapis, J., & Harding, K. E. (2017). Meaningful use of computers has a potential therapeutic and preventative role in dementia care: A systematic review. *Australasian Journal on Ageing*, *36*(4), 299–307. https://doi.org/10.1111/ajag.12446

Lifshitz-Vahav, H., Shrira, A., & Bodner, E. (2017). The reciprocal relationship between participation in leisure activities and cognitive functioning: the moderating effect of self-rated literacy level. *Aging and Mental Health*, *21*(5), 524–531. https://doi.org/10.1080/13607863.2015.1124838

Lin, P.-J., & Neumann, P. J. (2013). The economics of mild cognitive impairment. *Alzheimer's & Dementia : The Journal of the Alzheimer's Association*, *9*(1), 58–62. https://doi.org/10.1016/j.jalz.2012.05.2117

Lin, P. J., Zhong, Y., Fillit, H. M., Chen, E., & Neumann, P. J. (2016). Medicare Expenditures of Individuals with Alzheimer's Disease and Related Dementias or Mild Cognitive Impairment Before and After Diagnosis. *Journal of the American Geriatrics Society*, *64*(8), 1549–1557. https://doi.org/10.1111/jgs.14227

Lipnicki, D. M., Makkar, S. R., Crawford, J. D., Thalamuthu, A., Kochan, N. A., Lima-costa, M. F., … Sachdev, P. S. (2019). Determinants of cognitive performance and decline in 20 diverse ethno-regional groups : A COSMIC collaboration cohort study. *PLoS Medicine*, *16*(7), 1–27. https://doi.org/https://doi.org/10.1371/journal. pmed.1002853

Little, R. J. (2006). Calibrated Bayes : A Bayes / Frequentist Roadmap. *The American*

Statistician, *60*(3), 213–223. https://doi.org/10.1198/000313006X117837

Littleford, C., Hussey, D., Begum, N., & Oskala, A. (2016). *The Dynamics of Ageing: Evidence from the English Longitudinal Study of Ageing*. (James Banks, G. D. Batty, J. Nazroo, & A. Steptoe, Eds.). London, UK: The Institute for Fiscal Studies. Retrieved from file:///C:/Users/benja/Documents/Work/research_library/ELSA wave 7 study report.pdf

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., … Mukadam, N. (2017). Dementia prevention, intervention, and care. *The Lancet*, *390*. https://doi.org/10.1016/S0140-6736(17)31363-6

Lykestos, C., Chen, L., & Anthony, J. C. (1999). Cognitive Decline in Adulthood: An 11.5-Year Follow-Up of the Baltimore Epidemiologic Catchment Area Study. *American Journal of Psychiatry*, *156*(1), 58–65.

M. Tucker, A., & Stern, Y. (2011). Cognitive Reserve in Aging. *Current Alzheimer Research*, *8*(4), 354–360. https://doi.org/10.2174/156720511795745320

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalizing on chance. *Psychological Bulletin*, (111), 490–504.

MacCallum, Robert C., Edwards, M. C., & Cai, L. (2012). Hopes and cautions in implementing Bayesian structural equation modeling. *Psychological Methods*, *17*(3), 340–345. https://doi.org/10.1037/a0027131

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

Marden, J. R., Tchetgen Tchetgen, E. J., Kawachi, I., & Glymour, M. M. (2017). Contribution of Socioeconomic Status at 3 Life-Course Periods to Late-Life Memory Function and Decline: Early and Late Predictors of Dementia Risk. *American Journal of Epidemiology*, *186*(7), 1–10. https://doi.org/10.1093/aje/kwx155

Marengoni, A., Fratiglioni, L., Bandinelli, S., & Ferrucci, L. (2011). Socioeconomic status during lifetime and cognitive impairment no-dementia in late life: the population-based aging in the Chianti Area (InCHIANTI) Study. *Journal of Alzheimer's Disease : JAD*, *24*(3), 559–568. https://doi.org/10.3233/JAD-2011-101863

Marioni, R. E., Proust-Lima, C., Amieva, H., Brayne, C., Matthews, F. E., Dartigues, J.-F., & Jacqmin-Gadda, H. (2014). Cognitive lifestyle jointly predicts longitudinal cognitive decline and mortality risk. *European Journal of Epidemiology*, *29*(3), 211–219. https://doi.org/10.1007/s10654-014-9881-8

Marmot, M., Oldfield, Z., Clemens, S., Blake, M., Phelps, A., Nazroo, J., … Banks, J. (2014).

English Longitudinal Study of Ageing: Waves 0-6, 1998-2013 [computer file] (21st ed.). Colchester, Essex: UK Data Archive (distributor). https://doi.org/http://dx.doi.org/10.5255/UKDA-SN-5050-8

Matthews, K., Nazroo, J., & Marshall, A. (2019). Digital inclusion in later life : cohort changes in internet use over a ten-year period in England. *Ageing and Society*, *39*, 1914–1932.

McArdle, J. J., Fisher, G. G., & Kadlec, K. M. (2007). Latent variable analyses of age trends of cognition in the Health and Retirement Study, 1992-2004. *Psychology and Aging*, *22*(3), 525–545. https://doi.org/10.1037/0882-7974.22.3.525

McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological Methods*, *14*(2), 126–149. https://doi.org/10.1037/a0015857

McAvinue, L. P., Habekost, T., Johnson, K. a., Kyllingsbæk, S., Vangkilde, S., Bundesen, C., & Robertson, I. H. (2012). Sustained attention, attentional selectivity, and attentional capacity across the lifespan. *Attention, Perception, & Psychophysics*, 1570–1582. https://doi.org/10.3758/s13414-012-0352-6

McDaniel, M. a., & Einstein, G. O. (2011). The neuropsychology of prospective memory in normal aging: A componential approach. *Neuropsychologia*, *49*(8), 2147–2155.

McGrory, S., Doherty, J. M., Austin, E. J., Starr, J. M., & Shenkin, S. D. (2014). Item response theory analysis of cognitive tests in people with dementia: a systematic review. *BMC Psychiatry*, *14*(1), 47.

Meade, A. W., & Bauer, D. J. (2007). Power and Precision in Confirmatory Factor Analytic Tests of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 611–635. https://doi.org/10.1080/10705510701575461

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *The Journal of Applied Psychology*, *93*(3), 568–592. https://doi.org/10.5465/AMBPP.2006.27182124

Melby-Lervåg, M., & Hulme, C. (2016). There is no convincing evidence that working memory training is effective: A reply to Au et al. (2014) and Karbach and Verhaeghen (2014). *Psychonomic Bulletin and Review*, *23*(1), 324–330. https://doi.org/10.3758/s13423-015-0862-z

Meng, A., Nexø, M. A., & Borg, V. (2017). The impact of retirement on age related cognitive decline - A systematic review. *BMC Geriatrics*, *17*(1), 1–10. https://doi.org/10.1186/s12877-017-0556-7

Meng, X., & D'Arcy, C. (2012). Education and dementia in the context of the cognitive reserve

hypothesis: a systematic review with meta-analyses and qualitative analyses. *PloS One*, *7*(6), e38268. https://doi.org/10.1371/journal.pone.0038268

Meredith, W., & Teresi, J. a. (2006). An essay on measurement and factorial invariance. *Medical Care*, *44*(11 Suppl 3), S69-77. https://doi.org/10.1097/01.mlr.0000245438.73837.89

Min, J. W. (2018). A longitudinal study of cognitive trajectories and its factors for Koreans aged 60 and over : A latent growth mixture model. *International Journal of Geriatric Psychiatry*, *33*, 755–762. https://doi.org/10.1002/gps.4855

Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., & Hodges, J. R. (2006). The Addenbrooke ' s Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening, (21), 1078–1085.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, a H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Moafmashhadi, P., & Koski, L. (2013). Limitations for interpreting failure on individual subtests of the Montreal Cognitive Assessment. *Journal of Geriatric Psychiatry and Neurology*, *26*(1), 19–28.

Mungas, D., & Reed, B. R. (2000). Application of item response theory for development of a global functioning measure of dementia with linear measurement properties. *Statistics in Medicine*, *19*(11–12), 1631–1644.

Muniz-Terrera, G., Brayne, C., & Matthews, F. (2010). One size fits all? Why we need more sophisticated analytical methods in the explanation of trajectories of cognition in older age and their potential risk factors. *International Psychogeriatrics*, *22*(2), 291–299. https://doi.org/10.1017/S1041610209990937

Muthen, B. (2004). Latent Variable Analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of qantitative methodology for social sciences* (pp. 345–368). Newbury Park, CA: SAGE Publications Ltd.

Muthen, B., Asparouhov, T., Hunter, A., & Leuchter, A. (2011). Growth Modeling with Non-Ignorable Dropout: Alternative Analyses of the STAR*D antidepressant Trial. *Psychological Methods*, *16*(1), 17–33. https://doi.org/10.1037/a0022634.Growth

Muthén, B. O., & Asparouhov, T. (2013). BSEM Measurement Invariance Analysis. *Mplus Web Notes: No. 17*, (17). Retrieved from http://www.statmodel.com/examples/webnotes/webnote17.pdf

Muthén, B, & Asparouhov, T. (2014). *Causal effects in mediation modeling: An introduction with applications to latent variables*. *… in Structural Equation Modeling*. Retrieved from

http://www.statmodel.com/download/Causal.pdf

Muthén, Bengt, & Asparouhov, T. (2011). Bayesian SEM : A more flexible representation of substantive theory Web tables. *Psychological Methods*, *17*(3), 313–335. https://doi.org/10.1037/a0026802

Muthén, Bengt, & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313–335. https://doi.org/10.1037/a0026802

Muthén, L. K., & Muthén, B. O. (2014). MPlus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Nasserinejad, K., Rosmalen, J. Van, De Kort, W., & Lesaffre, E. (2017). Comparison of criteria for choosing the number of classes in Bayesian finite mixture models. *PLoS ONE*, *12*(1), 1–23. https://doi.org/10.1371/journal.pone.0168838

Negash, S., Xie, S., Davatzikos, C., Clark, C. M., Trojanowski, J. Q., Shaw, L. M., … Arnold, S. E. (2013). Cognitive and functional resilience despite molecular evidence of Alzheimer's disease pathology. *Alzheimer's & Dementia : The Journal of the Alzheimer's Association*, *9*(3), e89-95. https://doi.org/10.1016/j.jalz.2012.01.009

Nelson, P. T., Alafuzoff, I., Bigio, E. H., Bouras, C., Braak, H., Cairns, N. J., … Beach, T. G. (2012). Correlation of Alzheimer Disease Neuropathologic Changes With Cognitive Status : A Review of the Literature, *71*(5), 362–381.

Nelson, P. T., Head, E., Schmitt, F. a, Davis, P. R., Neltner, J. H., Jicha, G. a, … Scheff, S. W. (2011). Alzheimer's disease is not "brain aging": neuropathological, genetic, and epidemiological human studies. *Acta Neuropathologica*, *121*(5), 571–587. https://doi.org/10.1007/s00401-011-0826-y

O'Keeffe, E., Mukhtar, O., & O'Keeffe, S. T. (2011). Orientation to time as a guide to the presence and severity of cognitive impairment in older hospital patients. *Journal of Neurology, Neurosurgery, and Psychiatry*, *82*(5), 500–504.

Olaya, B., Bobak, M., Haro, J. M., & Demakakos, P. (2017). Trajectories of Verbal Episodic Memory in Middle-Aged and Older Adults: Evidence from the English Longitudinal Study of Ageing. *Journal of the American Geriatrics Society*, *65*(6), 1274–1281. https://doi.org/10.1111/jgs.14789

Paradise, M., Mccade, D., Hickie, I. B., Diamond, K., Lewis, S. J. G., & Naismith, S. L. (2015). Caregiver burden in mild cognitive impairment. *Aging and Mental Health*, *19*(1), 72–78. https://doi.org/10.1080/13607863.2014.915922

Perneczky, R., Kempermann, G., Korczyn, A. D., Matthews, F. E., & Ikram, M. A. (2019). Translational research on reserve against neurodegenerative disease : consensus report

of the International Conference on Cognitive Reserve in the Dementias and the Alzheimer ' s Association Reserve , Resilience and Protective Factors Professional Interes, 1–15.

Piccinin, A. M., Muniz-terrera, G., Clouston, S., Reynolds, C. A., Thorvaldsson, V., Deary, I. J., … Hofer, S. M. (2013). Coordinated Analysis of Age , Sex , and Education Effects on Change in MMSE Scores, *68*, 374–390. https://doi.org/10.1093/geronb/gbs077.Advance

Pietrzak, R. H., Lim, Y. Y., Ames, D., Harrington, K., Restrepo, C., Martins, R. N., … Maruff, P. (2014). Trajectories of memory decline in preclinical Alzheimer's disease: results from the Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing. *Neurobiology of Aging*, *36*(3), 1231–1238. https://doi.org/10.1016/j.neurobiolaging.2014.12.015

Plassman, B. L., Williams Jr, J. W., Burke, J. R., Holsinger, T., & Benjamin, S. (2010). Systematic Review : Factors Associated With Risk for and Possible Prevention of Cognitive Decline in Later Life. *Annals of Internal Medicine*, *153*, 182–193.

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., & Ferri, C. P. (2013). The global prevalence of dementia: a systematic review and metaanalysis. *Alzheimer's & Dementia : The Journal of the Alzheimer's Association*, *9*(1), 63-75.e2. https://doi.org/10.1016/j.jalz.2012.11.007

Proulx, C. M., Curl, A. L., & Ermer, A. E. (2018). Longitudinal Associations between Formal Volunteering and Cognitive Functioning. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *73*(3), 522–531. https://doi.org/10.1093/geronb/gbx110

Proust, C., & Jacqmin-Gadda, H. (2005). Estimation of linear mixed models with a mixture of distribution for the random effects. *Computer Methods and Programs in Biomedicine*, *78*(2), 165–173. https://doi.org/10.1016/j.cmpb.2004.12.004

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Racine, A. M., Gou, Y., Fong, T. G., Marcantonio, E. R., Schmitt, E. M., Travison, T. G., … Jones, R. N. (2018). Correction for retest effects across repeated measures of cognitive functioning: A longitudinal cohort study of postoperative delirium. *BMC Medical Research Methodology*, *18*(1), 1–14. https://doi.org/10.1186/s12874-018-0530-x

Rafnsson, S. B., Orrell, M., Orsi, E., Hogervorst, E., & Steptoe, A. (2017). Loneliness , Social Integration , and Incident Dementia Over 6 Years : Prospective Findings From the English Longitudinal Study of Ageing, *00*(00), 1–11. https://doi.org/10.1093/geronb/gbx087

Rasbash, J., Charlton, C., Browne, W., Healy, M., & Cameron, B. (2016). MLwiN Version 2.35. Centre for Multilevel Modelling, University of Bristol.

Reppermund, S., Ising, M., Lucae, S., & Zihl, J. (2009). Cognitive impairment in unipolar

depression is persistent and non-specific: further evidence for the final common pathway disorder hypothesis. *Psychological Medicine*, *39*(4), 603–614.

Reverberi, C., Cherubini, P., Baldinelli, S., & Luzzi, S. (2014). Semantic fluency: Cognitive basis and diagnostic performance in focal dementias and Alzheimer's disease. *Cortex*, *54*(1), 150–164.

Riley, K. P., Jicha, G. A., Davis, D., Abner, E. L., Cooper, G. E., Stiles, N., … Schmitt, F. A. (2011). Prediction of preclinical Alzheimer's disease: longitudinal rates of change in cognition. *Journal of Alzheimer's Disease : JAD*, *25*(4), 707–717. https://doi.org/10.3233/JAD-2011-102133

Robins, J. M., Hernan, M. A., & Brumback, B. (2000). Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology*, *11*, 550–560.

Robins, J. M., Hernán, M. a., & Wasserman, L. (2015). Discussion of "On Bayesian estimation of marginal structural models." *Biometrics*, n/a-n/a. https://doi.org/10.1111/biom.12273

Roe, C. M., Xiong, C., Miller, J. P., & Morris, J. C. (2007). Education and Alzheimer disease without dementia: support for the cognitive reserve hypothesis. *Neurology*, *68*(3), 223–228. https://doi.org/10.1212/01.wnl.0000251303.50459.8a

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, a. R., Schulz, J. B., … Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *NeuroImage*, *60*(1), 830–846.

Rousseau, J., & Mengersen, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society. Series B Statistical Methodology*, *73*(5), 689–710.

Roy, J. (2003). Modeling Longitudinal Data with Nonignorable Dropouts Using a Latent Dropout Class Model. *Biometrics*, *59*(4), 829–836. https://doi.org/10.1111/j.0006-341X.2003.00097.x

Royall, D. R., Palmer, R. F., Chiodo, L. K., & Polk, M. J. (2014). Towards an aging-specific cognitive phenotype: the freedom house study. *Experimental Aging Research*, *40*(3), 245–265. https://doi.org/10.1080/0361073X.2014.896665

Rubin, D. B. (1976). Biometrika Trust Inference and Missing Data Author (s): Donald B . Rubin Published by : Oxford University Press on behalf of Biometrika Trust Stable URL : http://www.jstor.org/stable/2335739 Accessed : 12-06-2016 21 : 34 UTC. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1186/1471-2105-12-432

Ryan, J. J., Glass, L. a, Bartels, J. M., Bergner, C. M., & Paolo, A. M. (2009). Predicting neuropsychological test performance on the basis of temporal orientation. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and*

Cognition, *16*(3), 330–337.

Saarela, O., Moodie, E., & Stephens, D. (2015). On Bayesian estimation of marginal structural models. *Biometrics*. https://doi.org/10.1111/biom.12274

Sajeev, G., Weuve, J., Jackson, J. W., VanderWeele, T. J., Bennett, D. A., Grodstein, F., & Blacker, D. (2016). Late-life Cognitive Activity and Dementia. *Epidemiology*, *27*(5), 732–742. https://doi.org/10.1097/EDE.0000000000000513

Sakurai, R., Ishii, K., Sakuma, N., Yasunaga, M., Suzuki, H., Murayama, Y., … Fujiwara, Y. (2018). Preventive effects of an intergenerational program on age-related hippocampal atrophy in older adults: The REPRINTS study. *International Journal of Geriatric Psychiatry*, *33*(2), e264–e272. https://doi.org/10.1002/gps.4785

Salthouse, T. a. (2004). What and When of Cognitive Aging. *Current Directions in Psychological Science*, *13*(4), 140–144.

Santos, C. Y., Snyder, P. J., Wu, W. C., Zhang, M., Echeverria, A., & Alber, J. (2017). Pathophysiologic relationship between Alzheimer's disease, cerebrovascular disease, and cardiovascular risk: A review and synthesis. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, *7*, 69–87. https://doi.org/10.1016/j.dadm.2017.01.005

Schaie, K. W., Willis, S. L., Jay, G., & Chipuer, H. (1989). Structural invariance of cognitive abilities across the adult life span: A cross-sectional study. *Developmental Psychology*, *25*(4), 652–662. https://doi.org/10.1037/0012-1649.25.4.652

Schmand, B., Smit, J., Lindeboom, J., Hooijer, C., Jonker, C., & Deelman, B. (1997). Low Education Is a Genuine Risk Factor for Accelerated Memory Decline and Dementia. *Journal of Clinical Epidemiology*, *50*(9), 1025–1033.

Seeman, T. E., Huang, M., Bretsky, P., Crimmins, E., Launer, L., & Guralnik, J. M. (2005). Education and APOE-e4 in Longitudinal Cognitive Decline : MacArthur Studies of Successful Aging. *Journal of Gerontology: PSYCHOLOGICAL SCIENCES*, *60*(2), 74–83.

Serra, L., Mancini, M., Cercignani, M., Di, C., Span, B., Giulietti, G., … Bozzali, M. (2017). Network-Based Substrate of Cognitive Reserve in Alzheimer ' s Disease. *Journal of Alzheimer's Disease*, *55*, 421–430. https://doi.org/10.3233/JAD-160735

Serrano-Pozo, A., & Growdon, J. H. (2019). Is Alzheimer's Disease Risk Modifiable? *Journal of Alzheimer's Disease*, *67*(3), 795–819. https://doi.org/10.3233/JAD181028

Serrano-pozo, A., Qian, J., Monsell, S. E., Frosch, M. P., Betensky, R. A., & Hyman, B. T. (2013). Examination of the Clinicopathologic Continuum of Alzheimer Disease in the Autopsy Cohort of the National Alzheimer Coordinating Center, *72*(12), 1182–1192.

Shadlen, M.-F., Larson, E. B., Wang, L., Phelan, E. a, McCormick, W. C., Jolley, L., … van

Belle, G. (2005). Education modifies the effect of apolipoprotein epsilon 4 on cognitive decline. *Neurobiology of Aging*, *26*(1), 17–24. https://doi.org/10.1016/j.neurobiolaging.2004.03.005

Shankar, A., Hamer, M., McMunn, A., & Steptoe, A. (2013a). Social Isolation and Loneliness: Relationships With Cognitive Function During 4 Years of Follow-up in the English Longitudinal Study of Ageing. *Psychosomatic Medicine*, *75*(2). Retrieved from https://journals.lww.com/psychosomaticmedicine/Fulltext/2013/02000/Social_Isolation_and_Loneliness___Relationships.9.aspx

Shankar, A., Hamer, M., McMunn, A., & Steptoe, A. (2013b). Social isolation and loneliness: relationships with cognitive function during 4 years of follow-up in the English Longitudinal Study of Ageing. *Psychosomatic Medicine*, *75*(2), 161–170. https://doi.org/10.1097/PSY.0b013e31827f09cd

Sharp, E. S., & Gatz, M. (2011). Relationship Between Education and Dementia An Updated Systematic Review. *Alzheimer Disease and Associated Disorders*, *25*(4), 289–304.

Shatil, E. (2013). Does combined cognitive training and physical activity training enhance cognitive abilities more than either alone ? A four-condition randomized controlled trial among healthy older adults. *Frontiers in Aging Neuroscience*, *5*(8), 1–12. https://doi.org/10.3389/fnagi.2013.00008

Sheldon, S., & Moscovitch, M. (2012). The nature and time-course of medial temporal lobe contributions to semantic retrieval: An fMRI study on verbal fluency. *Hippocampus*, *22*(6), 1451–1466.

Shen, H. W. (2017). Resource and Strategic Mobilization Model (RSM) of Productive Aging: Examining Older Americans' Participation in Various Productive Activities. *Ageing International*, *42*(1), 23–43. https://doi.org/10.1007/s12126-016-9259-1

Shi, D., Song, H., Liao, X., Terry, R., & Snyder, L. A. (2017). Bayesian SEM for Specification Search Problems in Testing Factorial Invariance. *Multivariate Behavioral Research*, *52*(4), 430–444.

Shigemori, K., Ohgi, S., Okuyama, E., Shimura, T., & Schneider, E. (2010). The factorial structure of the Mini-Mental State Examination (MMSE) in Japanese dementia patients. *BMC Geriatrics*, *10*, 36.

Short, S. D. (2014). *Power of Alternative Fit Indices for Multiple Group Longitudinal Tests of Measurement Invariance*.

Silbert, L. C., Dodge, H. H., Lahna, D., Promjunyakul, N. O., Austin, D., Mattek, N., … Kaye, J. A. (2016). Less Daily Computer Use is Related to Smaller Hippocampal Volumes in Cognitively Intact Elderly. *Journal of Alzheimer's Disease*, *52*(2), 713–717. https://doi.org/10.3233/JAD-160079

Singh-Manoux, A., Dugravot, A., Fournier, A., Abell, J., Ebmeier, K., Kivimäki, M., & Sabia, S. (2017). Trajectories of depressive symptoms before diagnosis of dementia: A 28-year follow-up study. *JAMA Psychiatry*, *74*(7), 712–718. https://doi.org/10.1001/jamapsychiatry.2017.0660

Slavin, H., McManus, J., & Stott, D. J. (2012). Small vessel cerebrovascular disease in older adults. *Reviews in Clinical Gerontology*, *22*(03), 184–194. https://doi.org/10.1017/S0959259812000056

Small, B. J., & Bäckman, L. (2007). Longitudinal Trajectories of Cognitive Change in Preclinical Alzheimer's Disease: A Growth Mixture Modeling Analysis. *Cortex*, *43*(7), 826–834. https://doi.org/10.1016/S0010-9452(08)70682-8

Small, B. J., Rosnick, C. B., Fratiglioni, L., & Backman, L. (2004). Apolipoprotein E and Cognitive. *Psychology and Aging*, *19*(4), 592–600.

Song, X.-Y., Lu, Z.-H., Hser, Y.-I., & Lee, S.-Y. (2011). A Bayesian Approach for Analyzing Longitudinal Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(2), 183–194. https://doi.org/10.1080/10705511.2011.557331

Song, Y.-N., Wang, P., Xu, W., Li, J.-Q., Cao, X.-P., Yu, J.-T., & Tan, L. (2018). Risk factors of rapid cognitive decline in Alzheimer's disease and mild cognitive impairment: A systematic review and meta-analysis. *Journal of Alzheimer's Disease*, *66*(2), 497–515. https://doi.org/10.3233/JAD-180476

Sprague, B. N., Freed, S. A., Webb, C. E., Phillips, C. B., Hyun, J., & Ross, L. A. (2019). The impact of behavioral interventions on cognitive function in healthy older adults : A systematic review. *Ageing Research Reviews*, *52*(April), 32–52. https://doi.org/10.1016/j.arr.2019.04.002

Squire, R. F., Noudoost, B., Schafer, R. J., & Moore, T. (2013). Prefrontal contributions to visual selective attention. *Annual Review of Neuroscience*, *36*, 451–466.

Starke, K. R., Seidler, A., Hegewald, J., Klimova, A., & Palmer, K. (2019). Retirement and decline in episodic memory : analysis from a prospective study of adults in England. *International Journal of Epidemiology*, *0*(0), 1–12. https://doi.org/10.1093/ije/dyz135

StataCorp. (2011). Stata Statistical Software: Release 12 IC. College Station, Texas: StataCorp LP.

StataCorp. (2013). Stata: Release 13 Statistical Software. College Station, Texas: StataCorp LP.

Steel, N., Huppert, F. A., McWilliams, B., & Melzer, D. (2004). Physical and cognitive function. In M. Marmot, J. Banks, R. Blundell, C. Lessof, & J. Nazroo (Eds.), *Health, wealth and lifestyles of the older population in England; The 2002 English Longitudinal Study of*

Ageing. (pp. 249–271). London, UK: The Institute for Fiscal Studies.

Steele, F. (2014). Multilevel Modelling of Repeat Measures Data. In *LEMMA VLE* (pp. 1–62). Bristol, UK: University of Bristol, Centre for Multilevel Modelling.

Steffener, J., & Stern, Y. (2012). Exploring the neural basis of cognitive reserve in aging. *Biochimica et Biophysica Acta*, *1822*(3), 467–473. https://doi.org/10.1016/j.bbadis.2011.09.012

Steffick, D. E. (2000). *Documentation of Affective Functioning Measures in the Health and Retirement Study*. Ann Arbor, MI.

Stephan, B. C. M., Birdi, R., Tang, E. Y. H., Cosco, T. D., Donini, L. M., Licher, S., … Robinson, L. (2018). Secular trends in dementia prevalence and incidence worldwide: A systematic review. *Journal of Alzheimer's Disease*, *66*(2), 653–680. https://doi.org/10.3233/JAD-180375

Steptoe, A., Breeze, E., Banks, J., & Nazroo, J. (2013). Cohort profile: the English Longitudinal Study of Ageing. *International Journal of Epidemiology*, *42*(6), 1640–1648.

Sterba, S. K. (2017). Pattern Mixture Models for Quantifying Missing Data Uncertainty in Longitudinal Invariance Testing. *Structural Equation Modeling*, *24*(2), 283–300. https://doi.org/10.1080/10705511.2016.1250635

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. (2012). Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurology*, *11*(11), 1006–1012. https://doi.org/10.1016/S1474-4422(12)70191-6

Stern, Y., Albert, S., Tang, M.-X., & Tsai, W.-Y. (1999). Rate of memory decline in AD is related to education and occupation: Cognitive reserve? *Neurology*, *53*(9), 1942–1947. https://doi.org/10.1212/WNL.53.9.1942

Stern, Y., Arenaza-urquijo, E. M., Bartr, D., Cantilon, M., Chetelat, G., Ewers, M., … Soldan, A. (2018). Whitepaper : Defining and investigating cognitive reserve , brain reserve , and brain maintenance. *Alzheimer's & Dementia*, 1–7. https://doi.org/10.1016/j.jalz.2018.07.219

Stern, Y., Cantillon, M., Clouston, S. A. P., Elman, J. A., Gold, B. T., Jones, R., … Pettigrew, C. (2019). Mechanisms underlying resilience in ageing. *Nature Reviews Neuroscience*, *20*(april), 2019. https://doi.org/10.1038/s41583-019-0138-0

Sun, J., Feng, X., Liang, D., Duan, Y., & Lei, H. (2012). Down-regulation of energy metabolism in Alzheimer's disease is a protective response of neurons to the microenvironment. *Journal of Alzheimer's Disease*, *28*(2), 389–402. https://doi.org/10.3233/JAD-2011-111313

Suzuki, H., Kuraoka, M., Yasunaga, M., Nonaka, K., Sakurai, R., & Takeuchi, R. (2014). Cognitive intervention through a training program for picture book reading in community-dwelling older adults : a randomized controlled trial. *BMC Geriatrics*, *14*(122), 1–9.

Swaab, D. F. (1991). Brain aging and Alzheimer's disease, "Wear and tear" versus "Use it or lose it." *Neurobiology of Aging*, *12*(4), 317–324. https://doi.org/10.1016/0197-4580(91)90008-8

Tampubolon, G., Nazroo, J., & Pendleton, N. (2017). Trajectories of general cognition and dementia in English older population: An exploration. *European Geriatric Medicine*, *8*(5–6), 454–459. https://doi.org/10.1016/j.eurger.2017.08.001

Tampubolon, Gindo. (2015). Cognitive ageing in Great Britain in the new century: Cohort differences in episodic memory. *PLoS ONE*, *10*(12), 1–17. https://doi.org/10.1371/journal.pone.0144907

Tarbuck, a F., & Paykel, E. S. (1995). Effects of major depression on the cognitive function of younger and older subjects. *Psychological Medicine*, *25*(2), 285–295.

Teipel, S. J., Meindl, T., Wagner, M., Kohl, T., Bürger, K., Reiser, M. F., … Hampel, H. (2009). White matter microstructure in relation to education in aging and alzheimer's disease. *Journal of Alzheimer's Disease*, *17*(3), 571–583. https://doi.org/10.3233/JAD-2009-1077

Tervo, S., Kivipelto, M., Hänninen, T., Vanhanen, M., Hallikainen, M., Mannermaa, A., & Soininen, H. (2004). Incidence and risk factors for mild cognitive impairment: a population-based three-year follow-up study of cognitively healthy elderly subjects. *Dementia and Geriatric Cognitive Disorders*, *17*(3), 196–203. https://doi.org/10.1159/000076356

The Institute for Fiscal Studies. (2015). English Longitudinal Study of Ageing Research Publications. Retrieved October 13, 2015, from http://www.elsa-project.ac.uk/publications/case/related

Todd, J. J., & Marois, R. (2005). Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. *Cognitive, Affective & Behavioral Neuroscience*, *5*(2), 144–155.

Tractenberg, R. E., Weiner, M. F., Aisen, P. S., Kaye, J. a., & Fuh, J. L. (2007). A simple method to rule out dementia with temporal orientation. *Alzheimer's and Dementia*, *3*(1), 28–32.

Tucker-Drob, E. M., Johnson, K. E., & Jones, R. N. (2009). The cognitive reserve hypothesis: a longitudinal examination of age-associated declines in reasoning and processing speed. *Developmental Psychology*, *45*(2), 431–446. https://doi.org/10.1037/a0014012

Tyas, S. L., Salazar, J. C., Snowdon, D. a, Desrosiers, M. F., Riley, K. P., Mendiondo, M. S., & Kryscio, R. J. (2007). Transitions to mild cognitive impairments, dementia, and death:

findings from the Nun Study. *American Journal of Epidemiology*, *165*(11), 1231–1238. https://doi.org/10.1093/aje/kwm085

Uhlmann, R. F., & Larson, E. B. (1991). Effect of Education on the Mini-Mental State Examination as a Screening Test for Dementia. *Journal of the American Geriatrics Society*, *39*(9), 876–880. https://doi.org/10.1111/j.1532-5415.1991.tb04454.x

Uttl, B., & Pilkenton-Taylor, C. (2001a). Letter cancellation performance across the adult life span. *The Clinical Neuropsychologist*, *15*(4), 521–530.

Uttl, B., & Pilkenton-Taylor, C. (2001b). Letter cancellation performance across the adult life span. *The Clinical Neuropsychologist*, *15*(4), 521–530. https://doi.org/10.1076/clin.15.4.521.1881

Valenzuela, M. J., & Sachdev, P. (2006). Brain reserve and cognitive decline: a non-parametric systematic review. *Psychological Medicine*, *36*(8), 1065–1073. https://doi.org/10.1017/S0033291706007744

van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthén, B. (2013). Facing off with Scylla and Charybdis: a comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology*, *4*(MI), 770. https://doi.org/10.3389/fpsyg.2013.00770

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740

van der Lee, S. J., Wolters, F. J., Ikram, M. K., Hofman, A., Ikram, M. A., Amin, N., & van Duijn, C. M. (2018). The effect of APOE and other common genetic variants on the onset of Alzheimer's disease and dementia: a community-based cohort study. *The Lancet Neurology*, *17*(5), 434–444. https://doi.org/10.1016/S1474-4422(18)30053-X

van Erp, S., Mulder, J., & Oberski, D. L. (2018). Prior sensitivity analysis in default bayesian structural equation modeling. *Psychological Methods*, *23*(2), 363–388. https://doi.org/10.1037/met0000162

van Hooren, S. a H., Valentijn, a M., Bosma, H., Ponds, R. W. H. M., van Boxtel, M. P. J., & Jolles, J. (2007). Cognitive functioning in healthy older adults aged 64-81: a cohort study into the effects of age, sex, and education. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, *14*(1), 40–54.

VanderWeele, T. J. (2012). Invited commentary: structural equation models and epidemiologic analysis. *American Journal of Epidemiology*, *176*(7), 608–612. https://doi.org/10.1093/aje/kws213

VanderWeele, T. J. (2014). A unification of mediation and interaction: a 4-way decomposition.

Epidemiology (Cambridge, Mass.), *25*(5), 749–761. https://doi.org/10.1097/EDE.0000000000000121

VanderWeele, T. J., Hawkley, L. C., Thisted, R. a, & Cacioppo, J. T. (2011). A marginal structural model analysis for loneliness: implications for intervention trials and clinical practice. *Journal of Consulting and Clinical Psychology*, *79*(2), 225–235. https://doi.org/10.1037/a0022610

Vanderweele, T. J., Vansteelandt, S., & Robins, J. M. (2014). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology (Cambridge, Mass.)*, *25*(2), 300–306. https://doi.org/10.1097/EDE.0000000000000034

Venneri, A., McGeown, W. J., Biundo, R., Mion, M., Nichelli, P., & Shanks, M. F. (2011). The Neuroanatomical Substrate of Lexical-Semantic Decline in MCI APOE ε4 Carriers and Noncarriers. *Alzheimer Disease and Associated Disorders*, *25*(3), 230–241.

Verghese, J., Lipton, R. B., Katz, M. J., Hall, C. B., Derby, C. A., Kuslansky, G., … Buschke, H. (2003). Leisure Activities and the Risk of Dementia in the Elderly. *New England Journal of Medicine*, *348*, 2508–2516.

Verhagen, a. J., & Fox, J. P. (2013). Bayesian tests of measurement invariance. *British Journal of Mathematical and Statistical Psychology*, *66*(3), 383–401. https://doi.org/10.1111/j.2044-8317.2012.02059.x

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*, *370*, 1453–1457. https://doi.org/10.1016/S0140-6736(07)61602-X

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2014). The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *International Journal of Surgery*, *12*(12), 1495–1499. https://doi.org/10.1016/j.ijsu.2014.07.013

von Oertzen, T., Hertzog, C., Lindenberger, U., & Ghisletta, P. (2010). The effect of multiple indicators on the power to detect inter-individual differences in change. *The British Journal of Mathematical and Statistical Psychology*, *63*(Pt 3), 627–646. https://doi.org/10.1348/000711010X486633

Vos, T., Flaxman, A. D., Naghavi, M., Lozano, R., Michaud, C., Ezzati, M., … Memish, Z. A. (2012). Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, *380*(9859), 2163–2196. https://doi.org/10.1016/S0140-6736(12)61729-2

Weiler, M., Casseb, R. F., Campos, B. M. De, Teixeira, C. V. L., Carletti-Cassani, A. F. M. K., Vicentini, J. E., … Castellano, G. (2018). Cognitive Reserve Relates to Functional Network

Efficiency in Alzheimer ' s Disease. *Frontiers in Aging Neuroscience*, *10*(August), 1–14. https://doi.org/10.3389/fnagi.2018.00255

White, L., Katzman, R., Losonczy, K., Salive, M., Wallace, R., Berkman, L., … Havlik, R. (1994). Association of Education with Incidence of Cognitive Impairment in Three Established Populations for Epidemiologic Studies of the Elderly. *Journal of Clinical Epidemiology*, *47*(4), 363–374.

Wicherts, J. M. (2016). The importance of measurement invariance in neurocognitive ability testing. *The Clinical Neuropsychologist*, *4046*(July), 1–11. https://doi.org/10.1080/13854046.2016.1205136

Wicherts, J. M., & Dolan, C. V. (2010). Measurement Invariance in Confirmatory Factor Analysis : An Illustration Using IQ Test Performance of Minorities. *Educational Measurement: Issues and Practice*, *29*(3), 39–47.

Widaman, K. F., Ferrer, E., & Conger, R. D. (2011). Factorial Invariance within Longitudinal Structual Equation Models: Measuring the Same Construct across Time. *Child Development Perspective*, *4*(1), 10–18. https://doi.org/10.1111/j.1750-8606.2009.00110.x.Factorial

Wiegand, I., Töllner, T., Dyrholm, M., Müller, H. J., Bundesen, C., & Finke, K. (2014). Neural correlates of age-related decline and compensation in visual attention capacity. *Neurobiology of Aging*, *35*(9), 2161–2173. https://doi.org/10.1016/j.neurobiolaging.2014.02.023

Wilson, R. S., Hebert, L. E., Scherr, P. A., Barnes, L. L., & Leon, C. F. M. De. (2009). Educational attainment and cognitive decline in old age. *Neurology*, *72*(2), 460–465.

Winnock, M., Letenneur, L., Jacqmin-Gadda, H., J Dallongeville, J., Amouyel, P., & Dartigues, J. (2002). Longitudinal analysis of the effect of apolipoprotein E e 4 and education on cognitive performance in elderly subjects: the PAQUID study. *Journal of Neurology, Neurosurgery, and Psychiatry*, *72*(6), 794–798.

Wortmann, M., Andrieu, S., Mackell, J., & Knox, S. (2010). EVOLVING ATTITUDES TO ALZHEIMER ' S DISEASE AMONG THE GENERAL PUBLIC AND CAREGIVERS IN EUROPE : FINDINGS FROM THE IMPACT SURVEY, *14*(7), 531–536.

Xavier, André J., D'orsi, E., De Oliveira, C. M., Orrell, M., Demakakos, P., Biddulph, J. P., & Marmot, M. G. (2014). English longitudinal study of aging: Can internet/e-mail use reduce cognitive decline? *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, *69*(9), 1117–1121. https://doi.org/10.1093/gerona/glu105

Xavier, Andre Junqueira, d'Orsi, E., Wardle, J., Demakakos, P., Smith, S. G., & Von Wagner, C. (2013). Internet use and cancer-preventive behaviors in older adults: Findings from a longitudinal cohort study. *Cancer Epidemiology Biomarkers and Prevention*, *22*(11), 2066–

2074. https://doi.org/10.1158/1055-9965.EPI-13-0542

Xie, H., Mayo, N., & Koski, L. (2011). Identifying and characterizing trajectories of cognitive change in older persons with mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders*, *31*, 165–172. https://doi.org/10.1159/000323568

Xu, Y., & Green, S. B. (2015). The Impact of Varying the Number of Measurement Invariance Constraints on the Assessment of Between-Group Differences of Latent Means. *Structural Equation Modeling: A Multidisciplinary Journal*, *5511*(September), 1–12. https://doi.org/10.1080/10705511.2015.1047932

Yaffe, K., Fiocco, A. J., Lindquist, K., Vittinghofff, E., Simonsick, E. M., Newman, A. B., … Study, F. the H. A. (2009). Predictors of maintaining cognitive function in older adults The Health ABC Study. *Neurology*, *72*(23), 2029–2035.

Yamashita, K., Taniwaki, Y., Utsunomiya, H., & Taniwaki, T. (2014). Cerebral Blood Flow Reduction Associated with Orientation for Time in Amnesic Mild Cognitive Impairment and Alzheimer Disease Patients. *Journal of Neuroimaging*, *24*(6), 590–594.

Yao, P., & Meng, C. (2015). Longitudinal Causal Inference of Cognitive Function and Depressive Symptoms in Elderly People. *Epidemiology Biostatistics and Public Health*, *12*(3), 1–8. https://doi.org/10.2427/11262

Yates, L. A., Ziser, S., Spector, A., & Orrell, M. (2016). Cognitive leisure activities and future risk of cognitive impairment and dementia: systematic review and meta-analysis. *Int Psychogeriatr*, (May), 1–16. https://doi.org/10.1017/S1041610216001137

Yin, J., Lassale, C., & Steptoe, A. (2019). Exploring the bidirectional associations between loneliness and cognitive functioning over 10 years : the English longitudinal study of ageing, 1–12. https://doi.org/10.1093/ije/dyz085

Zahodne, L. B., Glymour, M. M., Sparks, C., Bontempo, D., Dixon, R. A., MacDonald, S. W. S., & Manly, J. J. (2011a). Education does not slow cognitive decline with aging: 12-year evidence from the victoria longitudinal study. *Journal of the International Neuropsychological Society : JINS*, *17*(6), 1039–1046. https://doi.org/10.1017/S1355617711001044

Zahodne, L. B., Glymour, M. M., Sparks, C., Bontempo, D., Dixon, R. a, MacDonald, S. W. S., & Manly, J. J. (2011b). Education does not slow cognitive decline with aging: 12-year evidence from the victoria longitudinal study. *Journal of the International Neuropsychological Society : JINS*, *17*(6), 1039–1046. https://doi.org/10.1017/S1355617711001044

Zahodne, L. B., Stern, Y., & Manly, J. J. (2015). Differing effects of education on cognitive decline in diverse elders with low versus high educational attainment. *Neuropsychology*, *29*(4), 649–657. https://doi.org/10.1037/neu0000141

Zahodne, L. B., Wall, M. M., Schupf, N., Mayeux, R., Manly, J. J., Stern, Y., & Brickman, A. M. (2015). Late-life memory trajectories in relation to incident dementia and regional brain atrophy. *Journal of Neurology*, *262*(11), 2484–2490. https://doi.org/10.1007/s00415-015-7871-8

Zhu, Q. Bin, Bao, A. M., & Swaab, D. (2019). Activation of the Brain to Postpone Dementia: A Concept Originating from Postmortem Human Brain Studies. *Neuroscience Bulletin*, *35*(2), 253–266. https://doi.org/10.1007/s12264-019-00340-5

Zhu, X., Qiu, C., Zeng, Y., & Li, J. (2017). Leisure activities, education, and cognitive impairment in Chinese older adults: a population-based longitudinal study. *International Psychogeriatrics*, (May), 1–13. https://doi.org/10.1017/S1041610216001769

Zou, G. (2004). A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *American Journal of Epidemiology*, *159*(7), 702–706. https://doi.org/10.1093/aje/kwh090

Zou, G. Y. (2009). Assessment of risks by predicting counterfactuals. *Statistics in Medicine*, *28*(October 2009), 3761–3781. https://doi.org/10.1002/sim

Appendices.

Appendix 5.1 Correlation Matrix between Cognitive Function Variables Averaged Over All Waves

	Date	Month	Year	Day	Prospective	Fluency	Immediate	Delayed	Correct	Missed	Speed
Date	1										
Month	0.578	1									
Year	0.558	0.788	1								
Day	0.479	0.654	0.684	1							
Prospective	0.253	0.382	0.504	0.389	1						
Fluency	0.233	0.374	0.483	0.366	0.387	1					
Immediate Recall	0.245	0.357	0.458	0.366	0.381	0.479	1				
Delayed Recall	0.273	0.394	0.504	0.398	0.411	0.476	0.729	1			
Letters Correct	0.118	0.148	0.211	0.156	0.204	0.201	0.209	0.226	1		
Letters Missed	-0.142	-0.155	-0.219	-0.170	-0.243	-0.263	-0.266	-0.287	-0.888	1	
Letters Completed	0.126	0.196	0.265	0.206	0.11	0.204	0.19	0.179	-0.281	0.213	1

Appendix 5.2 Exploratory Factor Analysis Loadings by Wave for the 3 Factor Solution.

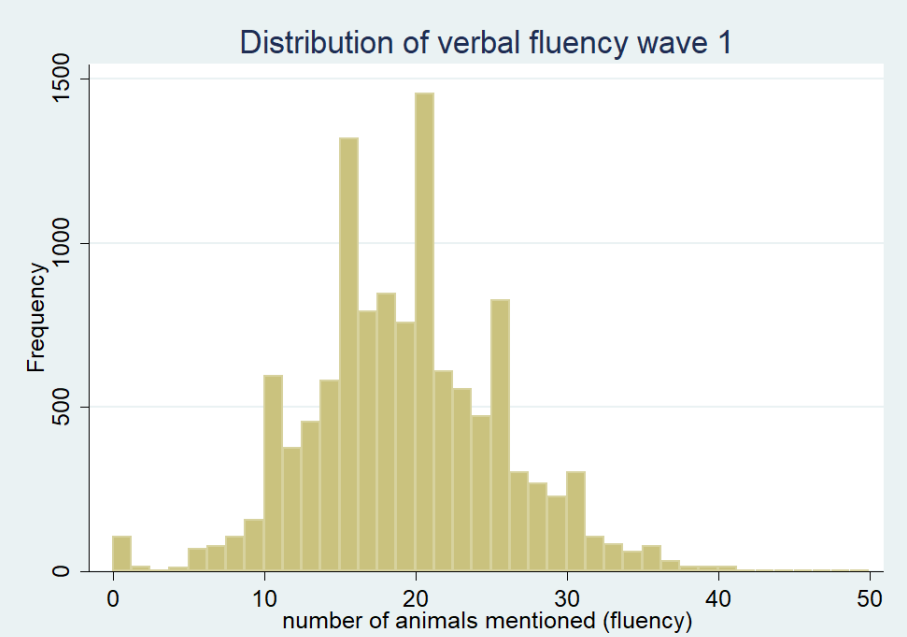
	Wave	1			2			3			4			5		
	Factor	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Variable	Date	0.601*			0.652*		-0.026	0.681*			0.681*		0.014	0.631*		
	Month	0.883*			0.994*		0.021	0.897*			0.921*		0.003	0.929*		
	Year	0.834*			0.910*		-0.024	0.931*			0.898*		-0.021	0.928*		
	Day	0.698*			0.785*		0.024	0.687*			0.826*		0.064*	0.772*		
	Prospective	0.267*	0.341*		0.403*	0.249*	-0.084*	0.348*	0.245*		0.316*	0.255*	-0.091*	0.380*	0.204*	
	Fluency	0.230*	0.465*		0.276*	0.397*	-0.032*	0.187*	0.488*		0.282*	0.404*	-0.022*	0.279*	0.396*	
	Immediate		0.848*			0.831*	0.035*		0.880*			0.883*	0.031*		0.891*	
	Delayed		0.814*			0.805*	0.009		0.790*			0.820*	0.006		0.757*	
	Correct			-0.971*			-1.013*			-1.026*			-1.016*			-0.985*
	Missed			0.897*			0.841*			0.829*			0.839*			0.873*
	Speed	0.202*	0.248*	0.373*	0.217*		0.374*	0.225*		0.365*	0.257*		0.398*	0.285*		0.397*

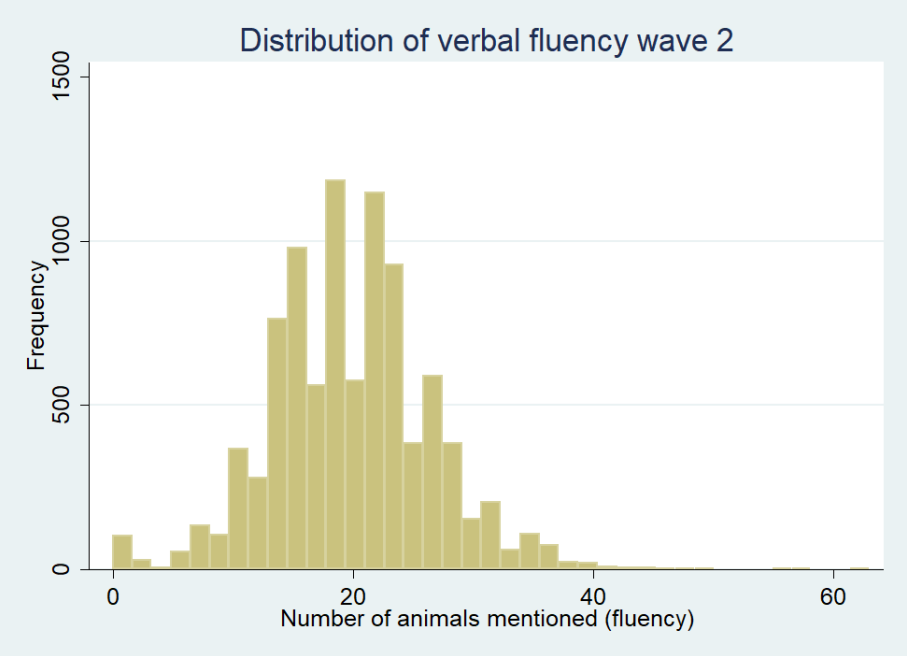
Loadings less than 0.2 not shown.

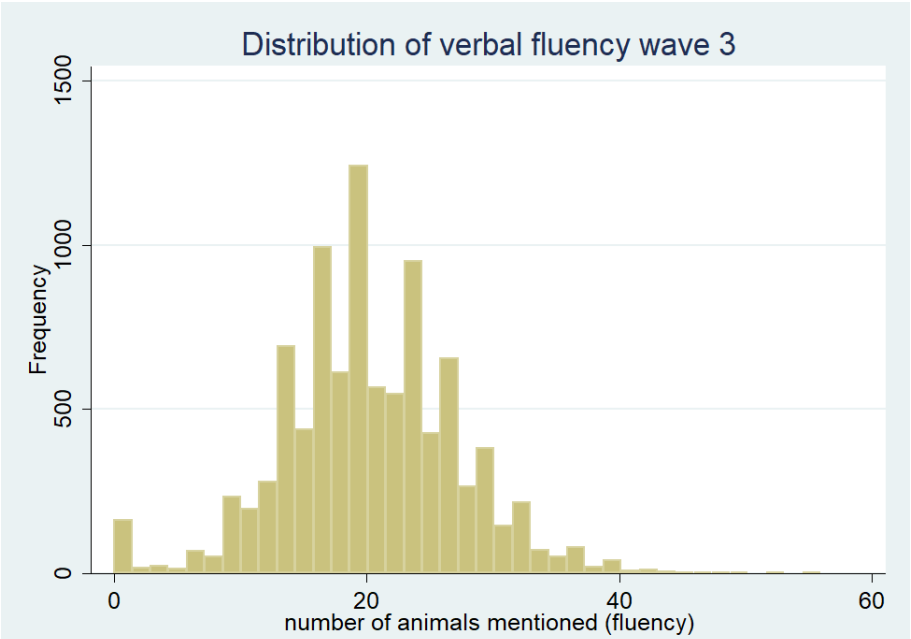
Appendix 5.3 The Correlation between Factor Scores and Index Scores.

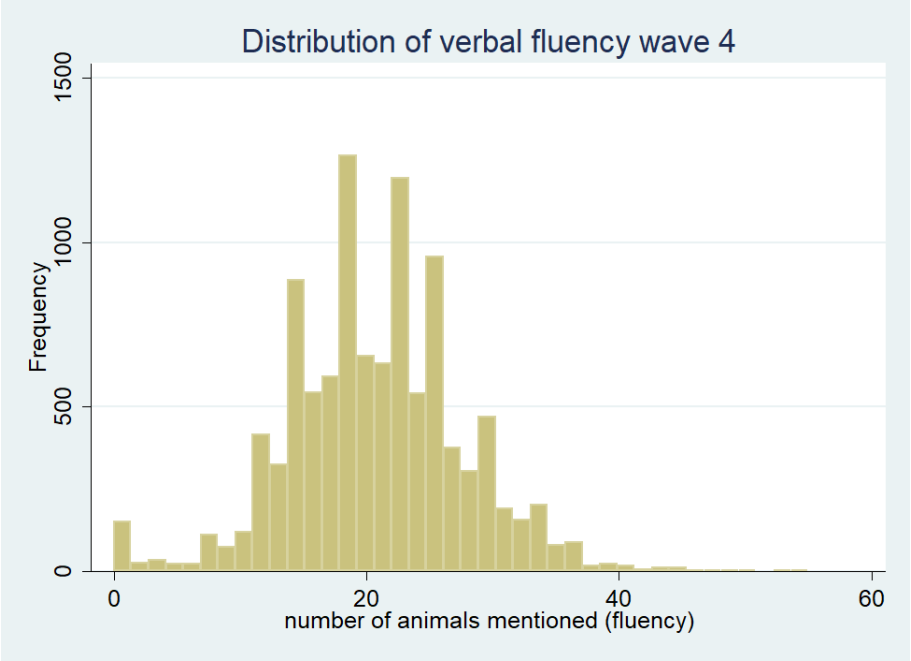
	Attention	Memory	Visual	Global	Memory	Executive
Attention	1.000					
Memory	0.826	1.000				
Visual	0.369	0.354	1.000			
Global	0.798	0.913	0.426	1.000		
Memory	0.824	0.949	0.294	0.897	1.000	
Executive	0.531	0.599	0.464	0.834	0.504	1.000

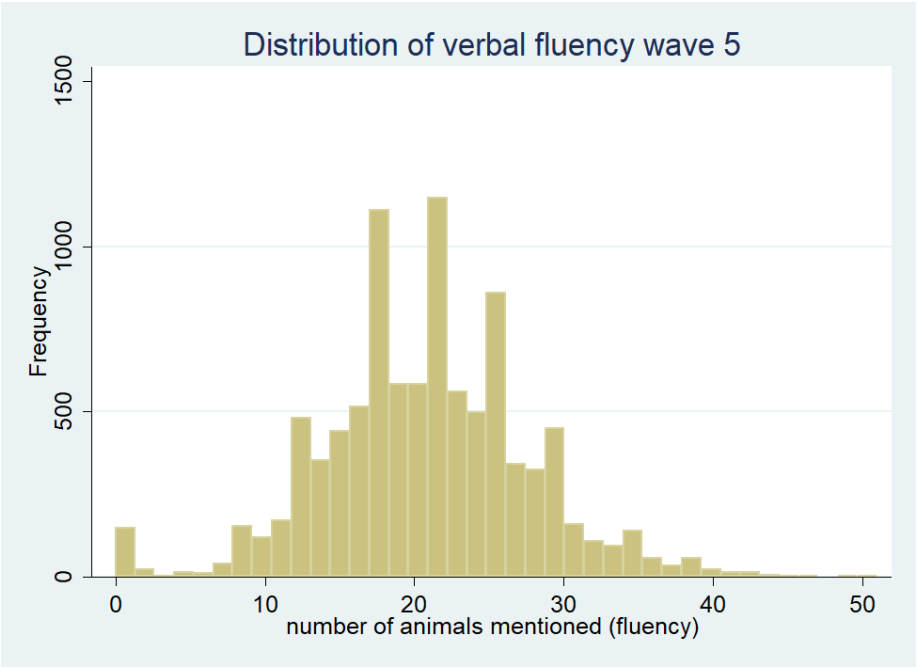
Appendix 7.1 Histograms demonstrating the distribution of verbal fluency scores in waves 1 to 5 of the English Longitudinal Study of Ageing



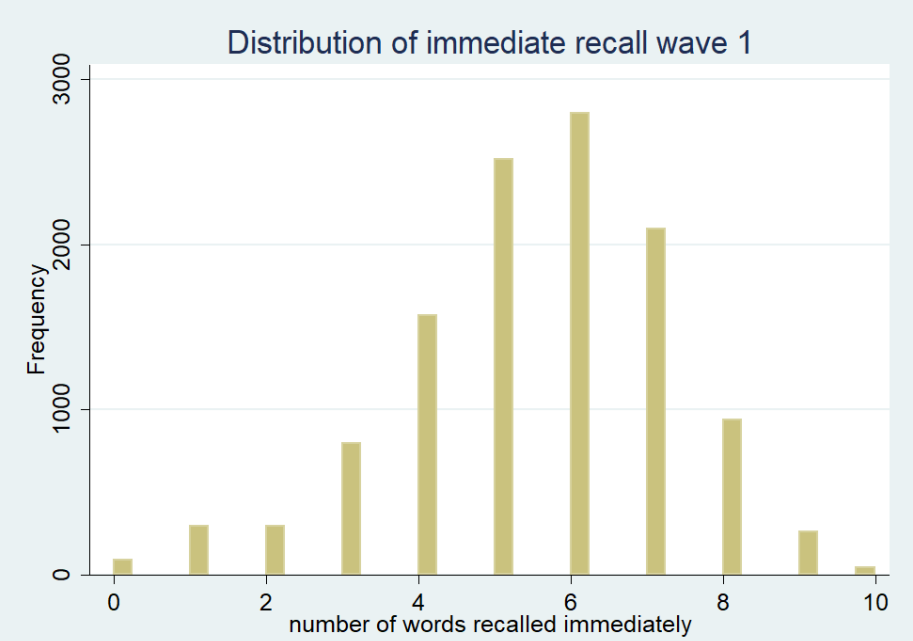


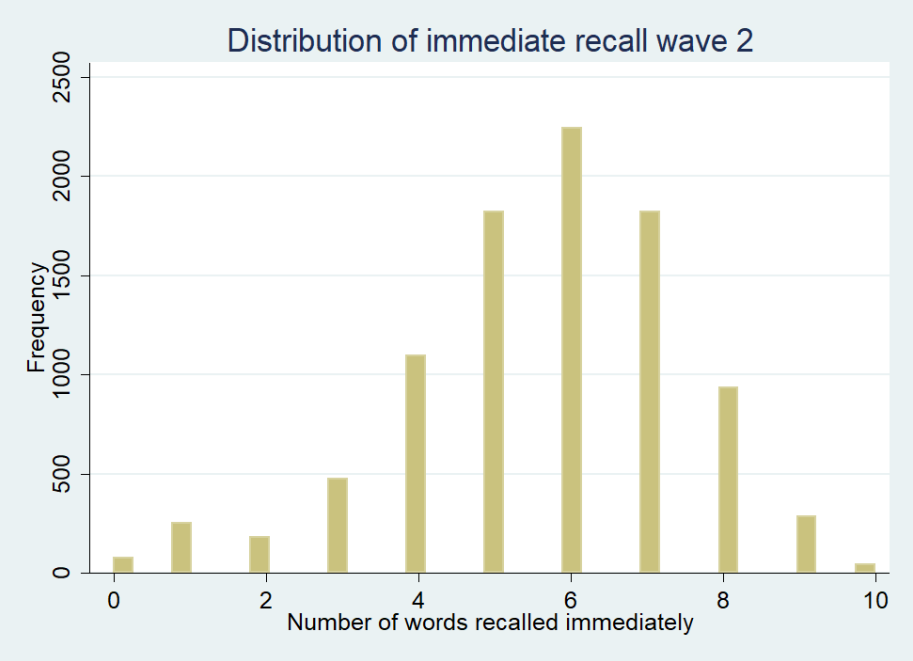

Distribution of verbal fluency wave 3
Frequency
0
500
1000
1500
0
20
40
60
number of animals mentioned (fluency)

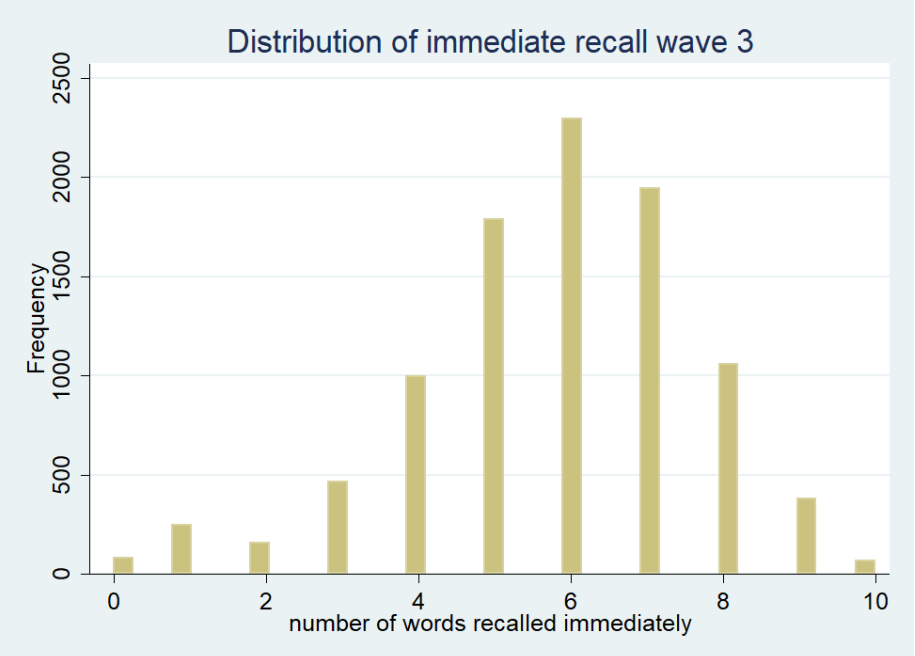

Distribution of verbal fluency wave 4
Frequency
0
500
1000
1500
0
20
40
60
number of animals mentioned (fluency)

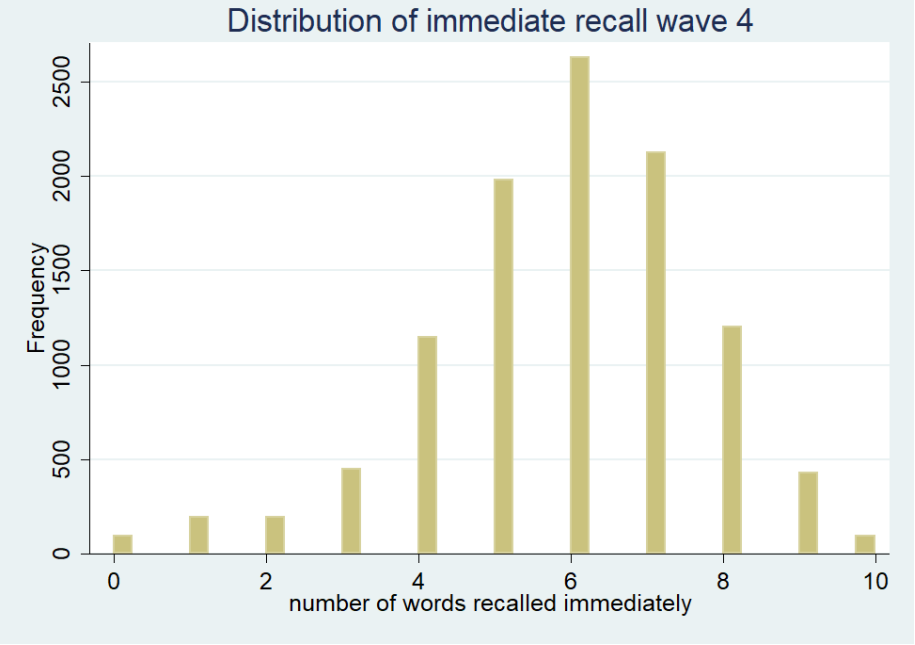

Distribution of verbal fluency wave 5
Frequency
0
500
1000
1500
0
10
20
30
40
50
number of animals mentioned (fluency)

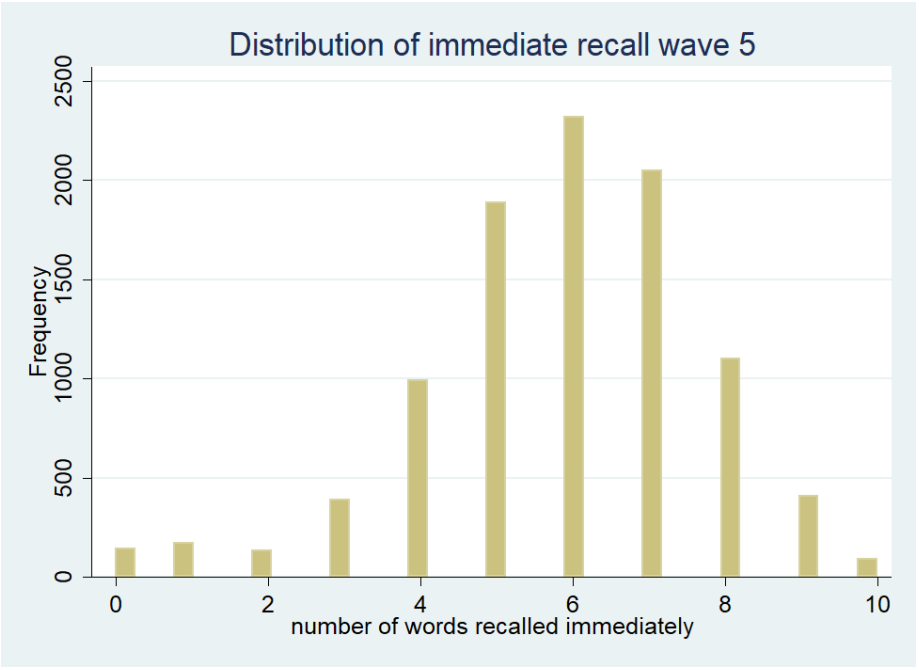
Appendix 7.2 Histograms demonstrating the distribution of immediate scores in waves 1 to 5 of the English Longitudinal Study of Ageing






Distribution of immediate recall wave 3
Frequency
0
500
1000
1500
2000
2500
0
2
4
6
8
10
number of words recalled immediately


Distribution of immediate recall wave 4
Frequency
0
500
1000
1500
2000
2500
0
2
4
6
8
10
number of words recalled immediately


Distribution of immediate recall wave 5
Frequency
0
500
1000
1500
2000
2500
0
2
4
6
8
10
number of words recalled immediately

Appendix 8.1 The covariates used in the estimation of the marginal structural models for each cognitive stimulating activity.

Working

Time invariant variables in the model for working in the month before the interview were age at baseline and gender. Time varying variables were whether the individual was at or past state retirement age, an interaction between age at baseline and being at state retirement age, attending evening classes and homemaking.

Volunteering

Time invariant variables in the model for volunteering were age at baseline and educational attainment. Time varying variables were whether the individual was at or past state retirement age, an interaction between age at baseline and being at state retirement age, verbal fluency, self-rated health, self-rated eyesight, going on day trips, being a member of a church or other religious organisation, charitable work, gym class membership and participation in 'other' activities not specified in the list provided to participants.

Social Club Membership.

Time invariant variables in the model for social club membership were wealth and educational attainment. Time varying variables were other group memberships, working, homemaking, volunteering, verbal fluency, episodic memory and self-rated hearing.

Internet use

Time invariant variables in the model for internet use were age at baseline, gender, 5 category social class, educational attainment, income and wealth. Time varying variables were verbal fluency, episodic memory, working, training, gym class membership and mobile phone use.

Evening class attendance

Time invariant variables in the model for attending evening classes were wealth, educational attainment and being non-married. Time varying variables were volunteering, episodic memory and self-rated health.

Daily Newspaper Reading

Time invariant variables in the model for newspaper reading were age at baseline, being non-married and gender. The time varying variables were verbal fluency, episodic memory and depression.

Appendix 8.2. Wave response rate and item non-response for all cognitively stimulating activity exposure variables.

Non-response		Wave 1	2	3	4	5	6	Final numbers†
	total *n*	11932	9249	7168	5971	5262	4711	
Main survey	Employment / Volunteering	9 (0.1%)	0	1 (0.0%)	2 (0.0%)	2 (0.0%)	0 (0.0%)	3875 2487
Questionnaire	Internet use / Newspaper	1164 (9.7%)	1169 (12.6%)	1010 (14.1%)	807 (13.5%)	531 (10.1%)	555 (11.8%)	2401 2770
	Social Club / Evening Classes	1577 (13.2%)	1597 (17.3%)	1264 (17.6%)	1021 (17.1%)	703 (13.4%)	665 (14.1%)	2460 2452

†The number of participants included in the final regression after accounting for all missing data at all time points required to estimate the inverse probability of treatment and censoring weights.

Appendix 8.3 Inverse probability of treatment and censoring weighted vs standard regression models for cognitively stimulating activities from 2004 to 2012 (waves 2 to 6) predicting risk of probable cognitive impairment in 2014 (wave 7).

Year of Exposure	Employment		Volunteering		Internet Use		Social Club		Newspaper Reading		Evening Classes	
	RR (95% CI)	P>z	RR (95% CI)	P>z	RR (95% CI)	P>z	RR (95% CI)	P>z	RR (95% CI)	P>z	RR (95% CI)	P>z
IPTCW												
W2 / 2004	0.94 (0.79 to 1.12)	0.466	0.65 (0.42 to 1.01)	0.055	**0.66 (0.52 to 0.84)**	**0.001**	0.91 (0.71 to 1.17)	0.46	0.96 (0.82 to 1.11)	0.575	0.82 (0.56 to 1.21)	0.315
W3 / 2006	0.90 (0.74 to1.10)	0.298	**0.63 (0.41 to 0.98)**	**0.042**	**0.65 (0.52 to 0.81)**	**<0.001**	1.09 (0.85 to 1.41)	0.481	0.96 (0.82 to 1.11)	0.554	0.83 (0.54 to 1.26)	0.376
W4 / 2008	0.89 (0.71 to 1.12)	0.322	**0.63 (0.40 to 0.98)**	**0.039**	**0.62 (0.50 to 0.77)**	**<0.001**	0.96 (0.73 to 1.26)	0.768	0.97 (0.83 to 1.13)	0.712	0.79 (0.52 to 1.19)	0.257
W5 / 2010	0.89 (0.68 to 1.15)	0.361	**0.52 (0.30 to 0.88)**	**0.015**	**0.67 (0.54 to 0.82)**	**<0.001**	0.92 (0.69 to 1.22)	0.564	0.97 (0.83 to 1.13)	0.694	1.07 (0.73 to 1.57)	0.736
W6 / 2012	0.97 (0.72 to 1.30)	0.816	**0.56 (0.34 to 0.94)**	**0.026**	**0.69 (0.56 to 0.85)**	**<0.001**	0.81 (0.60 to 1.08)	0.151	0.95 (0.82 to 1.11)	0.53	0.72 (0.48 to 1.09)	0.121
Standard												
W2 / 2004	1.05 (0.90 to 1.24)	0.523	0.99 (0.84 to 1.16)	0.877	**0.74 (0.61 to 0.89)**	**0.001**	1.07 (0.90 to 1.27)	0.458	0.89 (0.78 to 1.02)	0.1	0.89 (0.68 to 1.17)	0.411
W3 / 2006	0.98 (0.81 to 1.18)	0.843	0.84 (0.70 to 1.01)	0.061	**0.78 (0.65 to 0.94)**	**0.009**	1.14 (0.95 to 1.37)	0.162	0.98 (0.87 to 1.12)	0.81	0.85 (0.61 to 1.18)	0.337
W4 / 2008	0.90 (0.73 to 1.11)	0.327	**0.80 (0.66 to 0.98)**	**0.029**	**0.72 (0.60 to 0.86)**	**<0.001**	0.97 (0.79 to 1.19)	0.796	1.01 (0.88 to 1.16)	0.911	0.90 (0.65 to 1.25)	0.524
W5 / 2010	0.88 (0.69 to 1.12)	0.304	**0.73 (0.59 to 0.91)**	**0.005**	**0.64 (0.54 to 0.77)**	**<0.001**	0.87 (0.72 to 1.05)	0.147	0.91 (0.79 to 1.06)	0.222	0.99 (0.76 to 1.30)	0.96
W6 / 2012	1.00 (0.78 to 1.30)	0.978	**0.68 (0.55 to 0.85)**	**0.001**	**0.66 (0.56 to 0.79)**	**<0.001**	**0.68 (0.54 to 0.86)**	**0.001**	**0.86 (0.74 to 0.99)**	**0.033**	0.94 (0.79 to 1.08)	0.142

CI, confidence interval; IPTCW, inverse probability of treatment and censoring weights; RR, Risk Ratio; W, wave.

Appendix 8.4 Inverse probability of treatment and censoring weighted vs standard regression models for cognitively stimulating activities from 2004 to 2012 (waves 2 to 6) predicting telephone interview for cognitive status-27 score in 2014 (wave 7).

Year of	Employment		Volunteering		Internet Use		Social Club		Newspaper Reading		Evening Classes	
Exposure	beta (95% CI)	P>z	beta (95% CI)	P>z	beta (95% CI)	P>z	beta (95% CI)	P>z	beta (95% CI)	P>z	beta (95% CI)	P>z
IPTCW												
W2 / 2004	0.12 (-0.17 to 0.42)	0.420	0.08 (-0.36 to 0.51)	0.732	**0.93 (0.50 to 1.37)**	**<0.001**	-0.16 (-0.73 to 0.41)	0.581	-0.05 (-0.62 to 0.51)	0.851	0.41 (-0.12 to 0.93)	0.130
W3 / 2006	0.19 (-0.12 to 0.50)	0.232	0.34 (-0.11 to 0.78)	0.137	**0.91 (0.47 to 1.34)**	**<0.001**	-0.36 (-0.94 to 0.21)	0.213	-0.03 (-0.59 to 0.54)	0.927	0.44 (-0.12 to 0.99)	0.127
W4 / 2008	0.17 (-0.17 to 0.52)	0.331	0.40 (-0.05 to 0.86)	0.082	**0.91 (0.49 to 1.33)**	**<0.001**	-0.22 (-0.79 to 0.34)	0.434	-0.10 (-0.67 to 0.48)	0.742	0.38 (-0.14 to 0.90)	0.151
W5 / 2010	0.21 (-0.26 to 0.51)	0.536	0.37 (-0.10 to 0.84)	0.126	**0.93 (0.52 to 1.33)**	**<0.001**	0.05 (-0.52 to 0.61)	0.877	-0.18 (-0.75 to 0.39)	0.532	-0.17 (-0.91 to 0.58)	0.661
W6 / 2012	0.12 (-0.30 to 0.55)	0.569	0.39 (-0.08 to 0.85)	0.103	**0.86 (0.47 to 1.24)**	**<0.001**	0.17 (-0.47 to 0.81)	0.595	-0.13 (-0.71 to 0.44)	0.657	0.12 (-0.35 to 0.58)	0.631
Standard												
W2 / 2004	-0.02 (-0.27 to 0.22)	0.843	-0.06 (-0.36 to 0.24)	0.697	**0.95 (0.60 to 1.31)**	**<0.001**	-0.36 (-0.78 to 0.07)	0.099	0.23 (-0.24 to 0.70)	0.331	**0.56 (0.12 to 1.01)**	**0.014**
W3 / 2006	0.13 (-0.13 to 0.39)	0.317	**0.43 (0.14 to 0.73)**	**0.004**	**0.76 (0.41 to 1.01)**	**<0.001**	**-0.50 (-0.95 to -0.05)**	**0.029**	-0.10 (-0.56 to 0.36)	0.672	**0.47 (0.03 to 0.92)**	**0.039**
W4 / 2008	0.26 (-0.03 to 0.55)	0.082	**0.35 (0.05 to 0.65)**	**0.023**	**0.65 (0.31 to 1.00)**	**<0.001**	-0.20 (-0.64 to 0.24)	0.373	-0.24 (-0.72 to 0.24)	0.321	0.18 (-0.27 to 0.63)	0.437
W5 / 2010	0.21 (-0.13 to 0.55)	0.224	**0.43 (0.12 to 0.75)**	**0.007**	**0.90 (0.57 to 1.23)**	**<0.001**	0.32 (-0.13 to 0.76)	0.161	-0.11 (-0.59 to 0.38)	0.673	0.24 (-0.24 to 0.72)	0.333
W6 / 2012	0.10 (-0.29 to 0.49)	0.613	**0.53 (0.22 to 0.84)**	**0.001**	**0.91 (0.60 to 1.23)**	**<0.001**	**0.55 (0.07 to 1.03)**	**0.026**	0.07 (-0.42 to 0.56)	0.782	0.39 (-0.04 to 0.82)	0.076

CI, confidence interval; IPTCW, inverse probability of treatment and censoring weights; RR, Risk Ratio; W, wave.