BILINGUALISM AS A PROXY OF COGNITIVE RESERVE

MATTHIAS BERKES

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN PSYCHOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

August 2021

ABSTRACT

Previous studies have reported bilingualism to be a proxy of cognitive reserve (CR) based on evidence that bilinguals express dementia symptoms ~4 years later than monolinguals yet present with greater neuropathology at time of diagnosis when clinical levels are similar. This dissertation presents two studies that provide further evidence for the contribution of bilingualism to CR. The first study uses a novel brain health matching paradigm. Forty cognitively normal bilinguals with diffusion-weighted magnetic resonance images recruited from the community were matched with monolinguals drawn from a pool of 165 individuals in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. White matter integrity was calculated for all participants using fractional anisotropy, axial diffusivity, and radial diffusivity scores. Propensity scores were obtained using white matter measures, sex, age, and education as predictive covariates, and then used in one-to-one matching between language groups, creating a matched sample of 32 participants per group. Matched monolinguals had poorer clinical diagnoses than that predicted by chance from a theoretical null distribution, and poorer cognitive performances than matched bilinguals as measured by scores on the MMSE. The findings support the interpretation that bilingualism acts as a proxy of CR such that monolinguals have poorer clinical and cognitive outcomes than bilinguals for similar levels of white matter integrity even before clinical symptoms appear. The second study examines the role of biomarkers and genetic factors associated with Alzheimer disease in a sample of 641 individuals from the ADNI database. Gradient boosted regression modelling was used to examine the influence of 10 predictive factors on clinical diagnosis in 3 different models. Weighted propensity scores were applied to analyses of white matter integrity and cognitive performance between clinical groups in two models and between language groups in one model. Analyses revealed a strong influence

of biomarkers and genetic factors on clinical diagnosis in monolingual participants, but underrepresentation of bilingual participants in the sample limited interpretations of the findings between language groups. The results of the second study indicate that information about biomarkers and genetic factors improves analyses exploring the role of CR on dementia outcomes.

ACKNOWLEDGEMENTS

I would like to thank the following individuals, without whom this dissertation would not be possible:

First and foremost, my supervisor Dr. Ellen Bialystok, whose brilliance and insight guided me throughout the better part of a decade to produce work of uncompromising quality.

My committee members Dr. Shayna Rosenbaum and Dr. Dale Stevens, whose insightful comments and suggestions shaped this dissertation into its final form.

Dr. Mary Desrocher, Dr. Chandan Narayan, and Dr. Natalie Phillips for their observations and questions that led to stimulating discussion about my research.

Dr. John A. E. Anderson and Dr. Noelia Calvo for their tutelage and assistance, particularly in our 'Brainswap' study.

Dr. Ashley Chung-Fat-Yim for her intellectual support and friendship over the course of my Master's and PhD journeys.

Kornelia Hawrylewicz and Danika Wagner for managerial support and assistance in data collection.

My parents, Noreen Berkes and Dr. Zoltan Berkes, for their patience.

And finally, my partner and greatest cheerleader Julia Song, for her unending emotional and mental support over the years, without whom I would not have survived (quite literally).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GENERAL INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disease of the brain where individuals with the disease develop symptoms over time, including memory loss, language impairment, and in later stages, even loss of some bodily functions such as walking or swallowing (Adams et al., 1997; Chouinard, 2000; Horner et al., 1994). Alzheimer's disease leads to dementia and is ultimately fatal. As populations age and live longer, the prevalence of AD also increases. Globally, dementia (including of the Alzheimer's type but also others) affects ~50 million people, with this number rising to a projected 82 million in 2030 and 152 million by 2050 (World Health Organization, 2020). In the USA, an estimated 5.8 million Americans live with AD, but this number is projected to rise to close to 14 million by the year 2050 (Alzheimer's Association, 2020). As of 2018, AD was the fifth leading cause of death in individuals aged 65 or older in the USA. Care provided by family members was estimated at around $250 billion unpaid hours, without factoring in the associated mental and physical costs of taking care of a loved one with dementia. Total costs associated with health care and long-term care for adults aged 65 and older with AD were an estimated $305 billion in 2020. There is extensive physical and emotional stress on both personal and systemic levels for everyone involved in a diagnosis of AD. Unfortunately, pharmaceutical treatments are largely ineffective at slowing progression of the disease or treating symptoms (Becker et al., 2008; Mehta et al., 2017). There is, however, some evidence that the course of the disease could be delayed by non-pharmaceutical, cognitive-based methods (e.g., Baumgart et al., 2015). It has been estimated that a delay of 5 years of AD onset would lead to an ~50% decrease in overall disease frequency (Alzheimer's Association, 2015). Delaying, if possible, the development of AD symptoms would be the most beneficial

course of action for reducing the associated costs. The objective of this dissertation is to provide support for bilingualism in particular as a lifestyle factor that delays the onset of AD.

Reserve, broadly speaking, is one theory that provides a framework for how to go about delaying AD symptom onset and subsequent dementia. Individual differences exist in cognition, clinical status, and functional ability in aging and brain disease, and reserve is intended to explain how these differences come to be. Multiple potential mechanisms are responsible and involve neuroprotective mechanisms (i.e., factors preventing cognitive decline or neurodegeneration) and compensatory mechanisms (i.e., factors that allow individuals to adapt to declining neural health). These factors work either alone or in tandem to preserve cognitive performance during aging and neural decline (Barulli & Stern, 2013). This disconnect between preserved function and neurodegeneration is the hallmark feature of reserve and is possible through the specific concepts of cognitive reserve, brain reserve, and brain maintenance (N.B.: All definitions and discussions are largely couched in terms of AD and development of dementia, although these terms theoretically also apply to other sources of decline such as Parkinson's disease and Lewy Body disease).

Stern et al. (2020) define these individual concepts in depth to establish common definitions, which are summarised here. Brain reserve is generally thought of as "neurobiological capital"; that is, it refers to cortical thickness, total brain volume, quantity of neurons, or the like, at a given point in time. Individuals with high brain reserve are thought to deal with aging and neurodegeneration better than those with low brain reserve as a result of this built up "capital" prior to decline; there is more neural matter available to lose before cognitive difficulties are observed. In this sense, brain reserve is often thought of as a *passive* model of reserve in which cognitive impairment is imminent once a simple threshold of brain deterioration has occurred.

The reasons for high brain reserve compared to low brain reserve may be impacted by related concepts, brain maintenance and cognitive reserve, which, in turn, are impacted by genetic and lifestyle factors.

Brain maintenance is complementary to brain reserve, but whereas brain reserve refers to neural capital at a given point in time, brain maintenance refers to reduced age- or disease-related development of neural degeneration. Individuals with high brain maintenance will show slower development of neuronal plaques or grey matter deterioration compared to those with low brain maintenance. Thus, whereas brain reserve is measured at a given point in time, brain maintenance is best measured longitudinally by examining deterioration over time. Brain reserve may help protect against the effects of pathology, but brain maintenance is posited to prevent this pathology in the first place. Both genetic factors (e.g., allelic variation in genes) and lifestyle factors (e.g., stimulating leisure activities) are thought to influence brain maintenance.

Cognitive reserve posits that cognitive processes are adaptable, and it is this adaptability that helps explain individual discrepancies in cognitive functioning despite neurodegeneration and pathology. Cognitive reserve is thought of as an *active* process of reserve, such that individuals dynamically cope with or adjust to aging and pathology using compensatory mechanisms or functional brain processes. Theoretically, there are several outcomes to be expected when comparing high cognitive reserve individuals against low cognitive reserve individuals (depicted in Figure 1). First, individuals with high cognitive reserve would show better cognitive performance than low cognitive reserve individuals at similar levels of neuropathology. Second, individuals with high cognitive reserve would show greater amounts of neuropathology than individuals with low cognitive reserve at comparable levels of cognitive performance: high cognitive reserve individuals are better able to cope with the effects of

neurodegeneration than their low cognitive reserve peers. Third, the point of inflection, or where memory begins to be affected by AD, will be later for high cognitive reserve individuals than low cognitive reserve individuals. Lastly, once symptoms of cognitive decline appear, disease progression proceeds faster in high cognitive reserve than low cognitive reserve individuals. Considering these theoretical predictions, an obvious question to ask is: how does one determine who is and is not a high or low reserve individual?

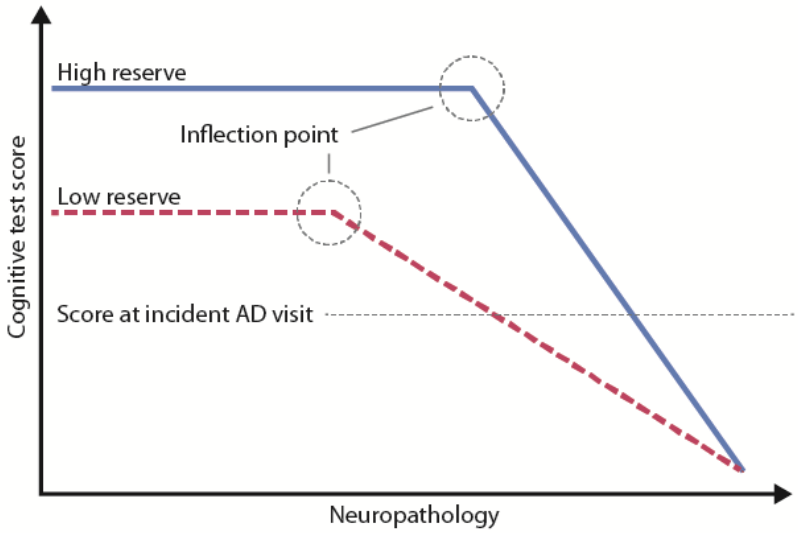


Figure 1. *Theoretical depiction of cognitive performance as a function of increasing neuropathology in high and low cognitive reserve individuals (adapted with permission from Stern, 2012)*

It is not possible to directly assess cognitive reserve, so it can only be investigated through examining proxies that are thought to covary and contribute to reserve. Commonly cited sources of cognitive reserve are socio-behavioural proxies, which include formal education, occupational complexity, stimulating leisure activities, and physical activity (Barulli & Stern, 2013; Stern, 2002; Stern, 2012; Valenzuela & Sachdev, 2006). Two common threads emerge from research involving these factors. The first is that individuals with 'higher' levels of one or

more of these factors (e.g., more formal education and/or more physical exercise and activity) show better clinical and cognitive outcomes in aging than their peers with 'low' levels of these factors. These benefits extend to an individual both when the activity occurred early in life (such as education) and if the activity is currently ongoing (such as stimulating physical and leisure activities). The second common thread is that these factors of cognitive reserve are effortful and engaging. Theoretically, any sufficiently challenging and continuous activity should be a source of cognitive reserve, although what qualifies as "sufficiently challenging" is a matter of discussion (Scarmeas & Stern, 2003).

Bilingualism has been posited as another proxy of cognitive reserve (Bialystok, 2021). Of all engaging activities, language use is the most sustained throughout the day. Both languages in the bilingual mind are jointly activated such that successful language production requires monitoring and selective attention to the required language (Kroll et al., 2014). Language use activates essentially the entire brain, except for some posterior regions (Friederici, 2011). This joint language activation in bilinguals has been posited to have extensive effects (i.e., shaping brain structure and cognitive ability) on brain regions and processes beyond language processing to include nonverbal domains and cognitive performance (Bialystok, 2017). Bilingualism is effortful and contributes to positive brain and cognitive changes; thus, bilingualism is a likely candidate as a lifestyle factor improving cognitive reserve in aging (high reserve). In comparison to bilingualism, single language use, or monolingualism, should not provide any cognitive benefits (low reserve).

This proposition has been tested across multiple studies in relation to the theoretical outcomes as seen in Figure 1. Previous research has shown that bilinguals are diagnosed with clinical impairment, specifically AD, ~4.5 years later than monolinguals (Bialystok et al., 2007).

This has been shown across cultures and after accounting for education and immigration status (Alladi et al., 2013; Chertkow et al., 2010; Woumans et al., 2015; Zheng et al., 2018). Recent research has also shown that bilinguals convert faster from mild cognitive impairment (MCI) to AD than monolinguals (1.9 years versus 2.6 years), a finding that supports both the notion that bilinguals are high reserve, as well as the prediction made by Stern's CR model (Berkes et al., 2020). Finally, there is evidence that bilingualism leads to similar cognitive and clinical outcomes as monolingualism despite greater amounts of neuropathology (Schweizer et al., 2012). In line with Stern's model, this evidence shows that bilinguals (high reserve) are better able to cope with neurodegeneration than monolinguals (low reserve). These studies support the idea that bilingualism is a proxy of cognitive reserve, with results that are in line with many predictions made by the cognitive reserve model.

One prediction made by Stern's model that has yet to be examined when using bilingualism as a proxy of reserve is that individuals with high reserve should show better cognitive and clinical outcomes than low reserve individuals at similar levels of neuropathology. This dissertation will attempt to explore this prediction in Study 1 by comparing bilingual and monolingual older adult participants on clinical and cognitive outcomes in older age while holding brain health (defined by white matter integrity) constant. Study 2 will further explore how bilingualism interacts with cognitive decline in aging by incorporating the role that typical biomarkers of AD play in clinical group diagnosis in participants from a large online database.

Study 1: Poorer Clinical Outcomes for Older Adult Monolinguals when Matched to Bilinguals on Brain Health

Lifelong bilingualism has been shown to confer executive control benefits for older adults, allowing bilinguals on average to outperform monolingual peers (Bialystok et al., 2016). Although positive effects for bilinguals compared to monolinguals are less likely to be found in young adults (e.g., Paap & Greenburg, 2013; Paap & Sawi, 2014; von Bastian et al., 2016) and children (e.g., Dick et al., 2019; Duñabeitia et al., 2014; see Leivada et al., 2020 for a review on the "phantom-like" effects of bilingualism), the positive effects of bilingualism are more reliably found for older adults, particularly when taking into account language proficiency and exposure (see Zhang et al., 2020 for a review). This adaptation in cognitive systems for older bilinguals is thought to result from the demands associated with managing two languages and selecting appropriate responses to satisfy current contextual cues. Managing two languages in one's mind has been likened to "mental juggling" (Kroll. 2008), as each language in a bilingual's repertoire remains simultaneously active while reading, hearing, and speaking, even in single language contexts (Dijkstra, 2005; Marian & Spivey, 2003; Kroll et al., 2006). Further, language selection in bilinguals is modulated by the cingulo-frontoparietal network – the same control network that monolinguals use for performing nonverbal tasks such as Simon or flanker tasks, providing functional neural evidence linking these two activities (e.g., Abutalebi & Green, 2008; Anderson, Chung-Fat-Yim et al., 2018; Luk, Green et al., 2011). Robust evidence also demonstrates that speaking two or more languages is associated with a delay in symptoms of dementia of between 3–5 years compared to monolinguals (e.g., Alladi et al., 2013; Bialystok et al., 2007; Chertkow et al., 2010; Woumans et al., 2015; Zheng et al., 2018). Two recent meta-analyses support the

claims from these studies that bilingualism delays the onset of dementia by 4.7 years (CI: 3.3–6.1; Anderson et al., 2020; Brini et al., 2020) but does not prevent bilingual individuals from developing dementia, a pattern consistent with cognitive reserve (Stern, 2002). However, as it is with behavioral results, some studies have failed to find any differences between monolinguals and bilinguals in clinical diagnoses (Lawton et al., 2015; Sanders et al., 2012; see Mukadam et al., 2017 for a meta-analysis, but see Grundy & Anderson, 2017, for a rebuttal), although continuous bilingual practice and immersion in bilingual environments more accurately predicts positive effects of bilingualism (see Del Maschio et al., 2018, for a review).

In light of the positive behavioral and neuropsychiatric findings, there has been a strong interest in exploring structural and functional brain differences attributable to bilingualism. Perani et al. (2017) used PET to show that in a patient sample matched on disease duration, bilingual patients with AD had more severe cerebral hypometabolism than monolingual patients, a measure that the authors attributed to reduced synaptic function and density. Despite this, bilinguals outperformed monolinguals on short- and long-term verbal memory and visuospatial tasks. Another study compared monolingual and bilingual patients with AD using computed tomography scans (Schweizer et al., 2012). Patients were matched on age, education, occupational status, and clinical level of dementia, yet bilingual patients showed greater medial temporal atrophy than the monolingual group. Importantly, despite this greater atrophy bilinguals were indistinguishable from monolinguals on cognitive status measures derived from standardized tests.

Although grey matter structure has been widely studied, white matter integrity is critical for cognitive functioning, particularly as atrophy occurs with aging (Bennett & Madden, 2014). Diffusion tensor imaging (DTI) is used to measure the directional displacement of water along

neural pathways in the brain and thus provides a measure of microstructural integrity. Fractional anisotropy (FA), axial diffusivity (DA), and radial diffusivity (DR) are measures that reflect the overall health of white matter and respectively correspond to the anisotropic diffusion along an axon, diffusion along the primary axis, and isotropic diffusion perpendicular to the primary axis. A useful heuristic is that higher FA values roughly correspond to greater white matter integrity, while a higher DR value is associated with demyelination of axons and thus poorer integrity (see Madden et al., 2009, for a review). The interpretation of DA, however, is less clear. Conflicting results have been reported in the literature with findings of both DA increases and decreases linked to age-related changes (Burzynska et al., 2010; Cox et al., 2016; Sexton et al., 2014). Notably, increased DA has also been reported as a necessary stage in neuronal loss (Acosta-Cabronero et al., 2012), especially related to microglial processes such that DA decreases as an initial response to axonal loss, but subsequently increases with the clearance of cell debris (Burzynska et al., 2010; see also Michielse et al., 2010, and Sexton et al., 2014, for similar patterns). This pattern of change in DA over time may explain the difference in findings, as the age at which an individual is tested will in part influence DA values and the direction of change.

Only a few studies have compared bilingual and monolingual white matter integrity in older age, with contrasting results. Luk, Bialystok, et al. (2011) showed that older adult bilinguals had greater FA values in the corpus callosum and bilateral superior and inferior longitudinal fasciculi than their monolingual peers. No group differences were found in DA, but monolinguals had greater DR in the body of the corpus callosum — some in areas that overlapped where bilinguals showed greater FA values. This greater white matter integrity in the bilingual group than the monolingual group was found even when both groups were matched on age, education, and gender, with similar neuropsychological performance on standardized tests.

In contrast, a study by Gold et al. (2013) involving older adults reported opposite findings – monolinguals had greater FA values than bilinguals in the corpus callosum, superior and inferior longitudinal fasciculus, and fornix whereas bilinguals showed greater DR in the inferior fronto-occipital fasciculus and corpus callosum than monolinguals. As with Luk, Bialystok, et al. (2011), there were no group differences in DA. However, in a sample of cognitively healthy older adults, Anderson, Grundy, et al. (2018) found that monolinguals had greater FA values, while bilinguals had higher DA and DR values, largely consistent with the results reported by Gold et al. (2013). The two groups were then matched on seven background measures using propensity score matching (PSM), after which only the greater DA findings in bilinguals remained. The higher values were present in a range of white matter tracts including the midbody and splenium of the corpus callosum, and the left superior temporal longitudinal fasciculus. The findings of Anderson, Grundy, et al. replicated those of Gold et al. in the unmatched sample, but more stringent matching criteria led to findings in the same region as that found by Luk, Bialystok, et al.: higher DA values for bilinguals in the left superior longitudinal fasciculus for Anderson, Grundy, et al., and higher FA values for bilinguals in this same region for Luk, Bialystok, et al. The differences in white matter integrity between these studies may possibly be explained by the participants' ages (a mean of approximately 64 years in the study by Gold et al. to a mean of 75 years of age in the study by Anderson, Grundy, et al.), as it has been previously noted that age is a determinant in white matter measures (Burzynska et al., 2010; Michielse et al., 2010; Sexton et al., 2014). However, the scarcity of research investigating this issue in regard to bilingualism means there is currently no consensus. The question is important because it addresses the key tenets of how bilingualism modifies white matter integrity across the lifespan

in particular and the neurological changes associated with increasing cognitive impairment in general.

The concept of cognitive reserve helps to explain the disjunction between preserved cognitive functioning and clinical pathology as has been reported for bilinguals (e.g., Brini et al., 2020). Reserve is thought to be the cumulative improvements to, or maintenance of, neural resources brought about by lifetime exposures like education, occupational complexity, or social engagement, such that individuals are better able to cope with neural decline. Education, in particular, has been extensively studied and posited as a socio-behavioral proxy of reserve, with findings that include higher risk of dementia in those with low education, and slower cognitive and functional decline in those with high educational attainment (for reviews see Meng & D'Arcy, 2012, and Stern, 2009). The findings suggest that education, as a proxy of cognitive reserve, acts to protect against the damaging effects of brain atrophy in both disease and aging. This dissociation between brain state and cognitive level is the signature of cognitive reserve (Bialystok et al., 2018; Stern, 2009).

As noted earlier, bilingualism is associated with a delay in onset of symptoms of dementia by approximately four years and thus has been posited to be another proxy of cognitive reserve. Early life experience in two languages is associated with a lower incidence of MCI than is found for those with minimal second-language learning (Wilson et al., 2015). Recently, bilingualism has also been shown to influence conversion times from MCI to dementia such that bilinguals converted faster to dementia than monolinguals (Berkes et al., 2020). Although this finding seems counterintuitive, faster conversion and decline once cognitive issues appear is in line with predictions made by cognitive reserve theory. Due to the greater accumulation of neuropathology in those with higher levels of reserve (i.e., bilinguals), the inflection point of

decline occurs later than those with low reserve (i.e., monolinguals). The endpoint of cognitive impairment, however, remains similar regardless of reserve. Thus, there is a steeper slope, or faster decline, for those who are able to withstand the detrimental effects of neuropathology for a longer time. This finding of sharper decline is not unique to bilingualism but has also been shown using the previously mentioned proxy of education (e.g., Scarmeas et al., 2006; Stern et al., 1999).

Cognitive reserve — defined in terms of bilingualism for the current study — attenuates age-related decline presumably through the strengthening of neural networks. This strengthening refers both to the accumulation of neural resources prior to decline (through disease or typical age-related decline) as well as compensation in alternate networks in response to task demands (see Cabeza et al., 2018, for a review). Typically, studies match participants on cognitive level and then examine the corresponding brain integrity associated with specific cognitive outcomes. However, this approach does not address what the cognitive outcomes would be for monolinguals in older age who showed the same level of neuropathology. This is the question for the present study.

The present study reverses the usual convention of matching participants on cognitive health to compare brain integrity. Instead, bilinguals and monolinguals were matched on white matter integrity and then cognitive health was evaluated. First, a principal component analysis (PCA) was conducted on white matter parameters to extract a component across each of these correlated measures, which captured the variation in average white matter diffusivity and reduced multicollinearity and multiple comparisons. Then, a sample of cognitively healthy older adult bilinguals were matched on white matter to a subset of monolinguals using PSM. Finally, a randomization analysis was used to compare cognitive health. If bilingualism leads to cognitive

reserve, then monolinguals matched to cognitively healthy bilinguals on white matter integrity will show less favorable cognitive outcomes than bilinguals as measured by clinical diagnoses and cognitive measures. This reversal of the usual approach to matching is novel in the literature. It is also suited to studies that have collected samples of "healthy" older adults that can then be matched to individuals in large databases that include a wider spectrum of cognitive abilities and brain states.

Method

Participants

Forty cognitively healthy older adult bilinguals and 38 cognitively healthy older adult monolinguals were recruited from the community for a prior study (Anderson, Grundy et al., 2018). Screening for language status was conducted via telephone interviews using the *Language and Social Background Questionnaire* (LSBQ; Anderson, Mak et al., 2018). All participants were right-handed with no known neurological impairments or MRI contraindications. Diffusion-weighted scans were subsequently performed, and the resulting images were analyzed. When compared by group, bilingual participants showed lower FA and higher DA values than monolinguals in regions that included, but were not limited to, the anterior corpus callosum, corona radiata, and superior temporal longitudinal fasciculus. That is, when matched for cognitive level, bilinguals showed more neuropathology than monolinguals as found in previous research. To reverse the standard approach, the data from these bilingual individuals were then used as the baseline to match a new group of monolinguals with similar values for white matter integrity.

Data for monolinguals were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (*adni.loni.usc.edu*), specifically from the ADNI-3 study. Detailed

language and social background information were not as readily available as those obtained from the LSBQ, so categorization as monolingual involved some assumptions. Patients were selected for inclusion if their primary language and preferred language of testing were both English. Additionally, they were considered monolingual if they identified as white or African American and were neither Latino nor Hispanic, so that participants with a strong likelihood of using or being exposed to Spanish were not included. It is impossible to rule out the possibility of other language use, but the ethnic, racial, and language criteria provided make it unlikely that these individuals used languages aside from English to a significant degree in their daily lives. After individuals in the database were identified who fit the inclusion criteria, participants for the study were selected through serial search. Participants who had T1-weighted, FLAIR, and axial DTI files available were chosen, for a total of 165 monolingual older adults.

Data Acquisition

Bilingual participants were scanned at York University using a Siemens Trio 3T scanner with a 32-channel head coil. DTI scans were whole-brain 64-direction diffusion-weighted images, with TR = 9200 ms, TE = 86 ms, voxel size of 2.0 mm^3, and FOV = 192 mm.

Monolingual participants taken from the ADNI database were tested at various sites across the United States and Canada, but all used a GE, Siemens, or Phillips scanner. DTI scans were whole-brain 48-direction diffusion-weighted images, obtained with TR = 7200 ms, TE = 56, and voxel size of 2.0 mm^3 for all scanner models. All scans were screened at Mayo Clinic for quality control before being accepted into the ADNI database.

Data Processing

The same protocol for MRI processing was applied separately for the bilingual and monolingual groups. Processing was performed in part using the MRtrix3 package

(www.mrtrix.org), which included the initial step of de-noising using the dwidenoise function. To utilize TOPUP, a synthetic b = 0 image was created following the Synb0-DisCo protocol (Schilling et al., 2019) and subsequently merged with a real b = 0 image obtained during scan acquisition. The results from TOPUP were then used for eddy distortion correction using the eddy function in FSL. Denoising was again done on the eddy output using *dwidenoise*, and residual maps were examined for quality control. Scans with high residual noise were excluded at this point as this indicated that EDDY had done a poor job modeling the data (n = 8 bilinguals, n = 4 monolinguals). Following this, a diffusion tensor was fit using the DTIFIT command from FSL. Finally, Tract-Based Spatial Statistics (TBSS; Smith et al., 2006) included in FSL was utilized on the FA images from the DTIFIT output. Each participant's FA images were aligned to a 1x1x1 mm standard space using nonlinear registration, and subsequently all merged together into a single 4D image. The mean of all FA images was created, as well as a thinned FA skeleton representing the centers of tracts common to all participants. This methodology was then applied to DA and DR data, and tracts were identified post-hoc using the John Hopkins University DTI based probabilistic white matter atlas included in FSL.

The numbers of participants included following these steps were 32 bilinguals and 161 monolinguals after removing participants with scans that were not of sufficient quality (e.g., containing slice drop out). All bilinguals spoke at least two languages, and aside from English, these languages included French (n = 7), German (n = 5), Italian (n = 3), Spanish (n = 3), and others (n = 14), learned at a mean age of M = 5.0.

Analyses and Results

Once whole brain values of FA, DA, and DR were obtained for every participant, principal components analysis (PCA) was conducted. Briefly, PCA uses an orthogonal

transformation to simplify complex data while preserving any trends or patterns (Lever et al., 2017). Furthermore, PCA is "blind" in that it finds patterns without prior knowledge of group inclusion or treatment. FA, DA, and DR measures were the input variables, with close to 80% of the variance in the data accounted for by the first principal component (Figure 2). For the first principal component, a loading score was derived per participant (similar to the concept of "factor scores" from factor analysis) which reflected each person's relative position in this multivariate space, and this was used in the subsequent PSM step. Thus, the PCA provided a single overall assessment of white matter integrity that captured about 80% of the variation.

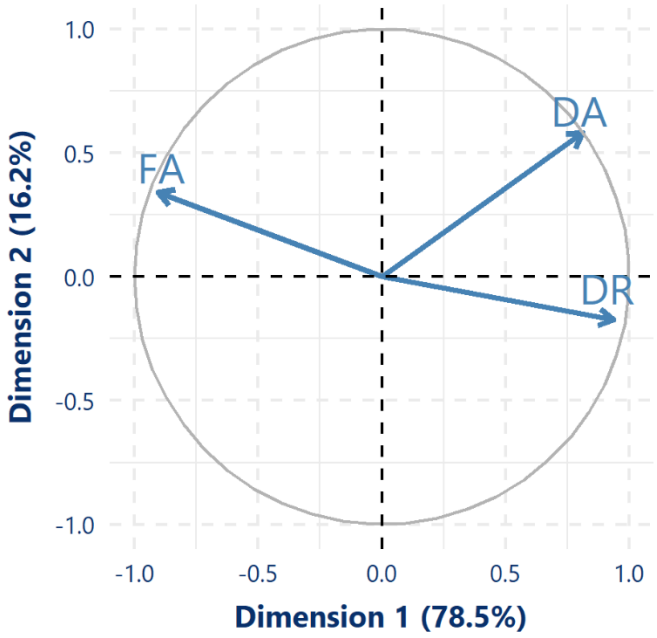


Figure 2. *Variables factor map of principal component analysis using fractional anisotropy (FA), axial diffusivity (DA), and radial diffusivity (DR) measures as input variables.*

To compare cognitive health (i.e., clinical diagnosis) between the two groups, PSM was used to explicitly match bilinguals to monolinguals. PSM is a useful method for sampling from a large reservoir of control participants (monolinguals) to create a smaller subsample with a distribution of covariates that is similar to the distribution in the treated group (bilinguals)

(Rosenbaum & Rubin, 1983). A propensity score is calculated and used as a balancing score wherein participants with similar propensity scores will have similar baseline covariate values regardless of the treatment group. Although there does not seem to be a consensus on which variables to include in the propensity score model, theoretical models suggest including any variables that may influence treatment assignment (Austin, 2011). Additionally, one-to-one matching is the most common implementation of PSM such that pairs of control and treated participants are created with similar propensity scores. In this way, the sample of control participants should be reduced to match the treatment group, in the present case producing 32 participants in each group.

For the analyses, the MatchIt package in R (Ho et al., 2007) was used with the following formula: matchit(Group ~ Sex + Education + Age + PCA, data, method = "nearest", distance = "logit", discard = "treat"). Sex, education, age, and first principal component scores were included as baseline covariates to predict treatment conditions, with the ‘discard’ option specified such that any bilinguals who were sufficiently different from the propensity score model would be excluded. The final sample after this step included 32 bilinguals and 32 monolinguals for which analyses were then performed.

Due to the nature of the current study, comparing clinical diagnoses between bilinguals and monolinguals would yield little of value. Bilingual participants were selected for the original study on the basis of having reported being cognitively normal (CN), whereas monolingual participants were chosen at random from a larger pool which encompassed CN, MCI, and AD diagnoses. As shown in Table 1, the distribution of these diagnostic categories in the unmatched monolingual sample was 117 considered to be CN, 34 with MCI (subtype undefined), and 10 with AD. However, the measurement of interest is the diagnoses of monolinguals *after* they have

been matched on bilingual brain health measures to the original larger monolingual sample that was randomly selected. To accomplish this, a cognitive profile score was created and assigned for each participant. A score of 0 indicates a diagnosis of CN, while a score of 1 indicates impairment (i.e., MCI or AD). MCI and AD were not differentiated in value because any impairment past normal cognition was noteworthy. The unmatched monolingual sample had a mean cognitive profile score of 0.27 compared to a mean score of 0.41 for the matched monolingual sample. That is, ~27% of unmatched monolinguals had a diagnosis of MCI or AD whereas ~41% of matched monolinguals had received a clinical diagnosis of impairment.

To compare the proportion of cognitively impaired individuals in the matched sample to the overall sample, a null distribution was created using the *infer* package in R (https://github.com/tidymodels/infer) by running 1000 permutations of random samples of monolinguals using the "true" proportion of 0.27, against which the sample proportion of 0.41 was compared.

Demographic information, DTI measures, and PCA scores for the full monolingual sample, bilingual sample, and matched monolingual sample are presented in Table 1. The matched dataset had a balance improvement of 62% on propensity scores in that matched monolingual propensity scores (M = 0.34) more closely matched bilingual propensity scores (M = 0.49) than when using those of the full monolingual sample (M = 0.10). Other predictive covariates were also improved from a range of 75% for education levels to 80% for age.

T-tests were performed to compare the bilingual and matched monolingual groups on the covariates entered in the PSM model. There were no significant differences between group means for education, $t(62) = 0.29$, $p = 0.78$, $d = 0.07$, age, $t(62) = 0.33$, $p = 0.74$, $d = 0.08$, and PCA scores, $t(62) = 1.61$, $p = 0.11$, $d = 0.4$. A chi-square test also revealed no difference in

proportions of sex between the two groups, $p = 1$. Due to possible protocol and scanner differences between testing sites, an additional ANOVA test examining PCA scores in the matched sample as a function of site was conducted. There was no significant difference in the overall model, $F(16, 47) = 1.43$, $p = .17$, $\eta^2 = .33$, nor were any pairwise comparisons between sites significant using a Tukey adjustment, ps > .05.

Table 1. *Demographic information and brain measure means (with standard deviations) for the full unmatched monolingual sample, bilingual sample, and matched monolingual sample.*

Group	N	Age in years	Education in years	MMSE	Fractional Anisotropy	Axial Diffusivity	Radial Diffusivity	PCA scores	Cognitive Profile[1]
Monolinguals (unmatched)	161 (58% F)	71.3 (7.1)	16.8 (2.3)	28.3 (2.7)	0.48 (0.03)	1.2×10^{-3} (0.03×10^{-3})	5.5×10^{-4} (0.4×10^{-4})	-0.36 (1.4)	CN = 117 (73%); MCI = 34 (21%); AD = 10 (6%)
Bilinguals	32 (72% F)	73.5 (3.8)	16.1 (2.8)	29.4 (0.7)	0.42 (0.02)	1.2×10^{-3} (0.02×10^{-3})	6.6×10^{-4} (0.5×10^{-4})	1.79 (1.1)	CN = 32 (100%)
Monolinguals (matched)	32 (69% F)	73.1 (6.5)	16.3 (2.5)	26.7 (4.4)	0.45 (0.02)	1.2×10^{-3} (0.04×10^{-3})	6.1×10^{-4} (0.4×10^{-4})	1.32 (1.2)	CN = 19 (59%); MCI = 8 (25%); AD = 5 (16%)

[1]CN = Clinically normal; MCI = Mild cognitive impairment; AD = Alzheimer's disease

Once predictive covariates and brain health were matched between groups, two variables of interest were considered: cognitive performance as measured by MMSE scores and clinical diagnoses of participants. First, a one-way ANOVA revealed a significant difference in MMSE scores, $F(1, 62) = 12.17$, $p < .001$, $\eta^2 = .16$, where bilinguals had higher mean MMSE scores ($M = 29.4$) than monolinguals ($M = 26.7$).

Second, the proportion of individuals in each of the three groups described as CN, MCI, and AD are reported in Table 1. A randomization-based test for a single proportion was used to compare the proportion of cognitively impaired participants in the matched sample to the overall sample of monolinguals from the ADNI database (Figure 3). The matched monolingual sample

had significantly poorer clinical outcomes (i.e., higher scores on the cognitive profile score reflecting MCI and AD) than that predicted by a null distribution generated from resampling the unmatched sample, $p < .001$. Thus, the matched monolingual sample was more cognitively impaired than would be expected in a theoretical population of both cognitively normal and impaired individuals.

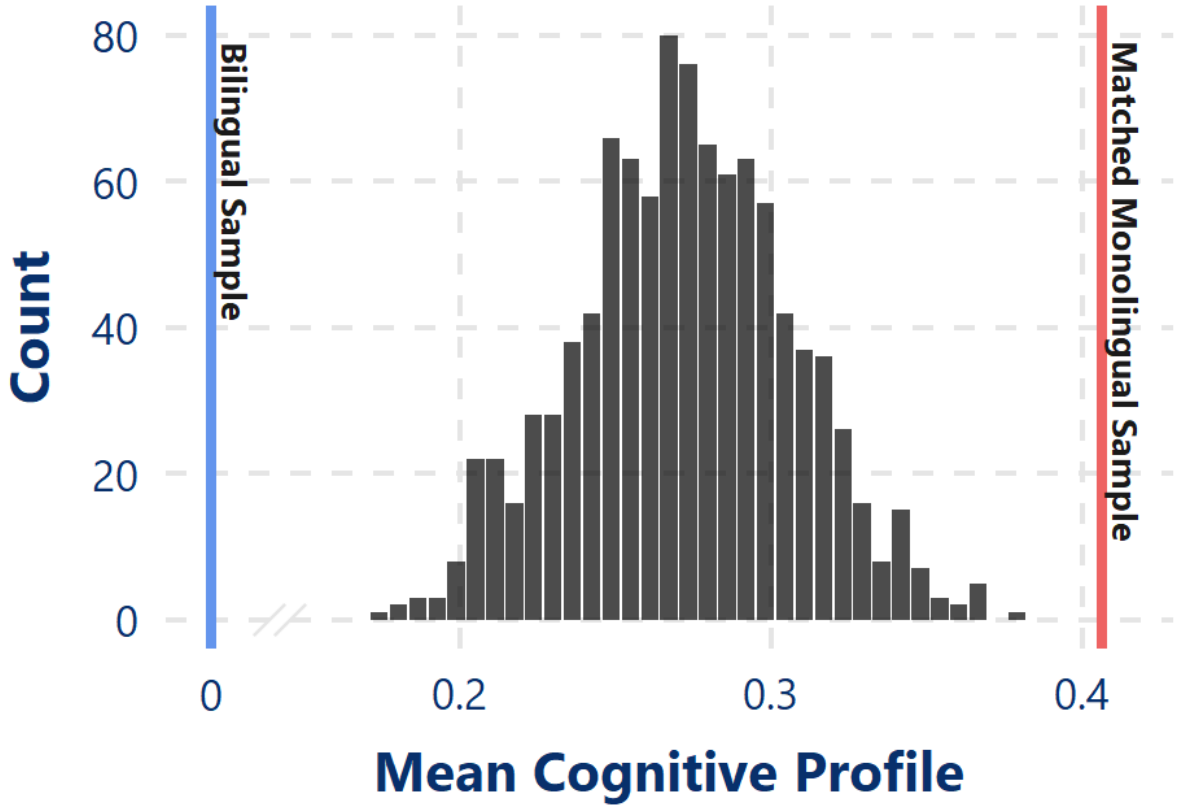


Figure 3. *Randomisation-based null distribution of mean cognitive profiles (where 0 = 'Healthy' and 1 = 'Unhealthy') with matched monolingual sample (red line) and bilingual sample (blue line).*

Discussion

Previous studies have focused on bilingualism as a form of cognitive reserve by using different types of measures including (but not limited to) executive functioning (see Bialystok, 2017, for review), brain imaging (e.g., Abutalebi et al., 2014, 2015), and dementia onset (e.g.,

Alladi et al., 2013; Bialystok et al., 2007; Chertkow et al., 2010; Woumans et al., 2015). These studies typically match participants on age and cognitive level and generally report that, when compared to monolinguals, bilinguals show greater brain atrophy, but better (or equivalent) cognitive outcome. However, this approach leaves unanswered the question about how different levels of brain health correspond to cognitive outcomes in these two groups, the reverse of the question that is usually examined. In other words, how would monolingual older adults cope with the levels of brain integrity found for bilinguals? The present study aimed to fill that gap.

A crucial measure of brain health is white matter integrity, but there is a lack of consistency regarding how these measures and cognitive performance are impacted by bilingualism with aging. In the study by Luk, Bialystok et al. (2011), bilinguals had better white matter integrity than monolinguals as measured by higher FA values in the corpus callosum and superior and inferior longitudinal fasciculi. By contrast, Gold et al. (2013) found that bilinguals had poorer white matter integrity than monolinguals, showing lower FA values in the same regions found by Luk, Bialystok et al., as well as in the fornix. The findings by Anderson, Grundy et al. (2018) are less clear: stricter matching criteria using PSM resulted in higher DA values for bilinguals than monolinguals in parts of the corpus callosum and the left superior temporal longitudinal fasciculus. Unlike the findings of Luk, Bialystok et al. and Gold et al., no differences in FA were found. The interpretation by Anderson, Grundy et al. was that DA enhances white matter integrity as "an index of diffusion along the primary gradient that is associated with positive cognitive outcomes", and thus bilinguals had better white matter integrity than monolinguals. However, studies examining white matter integrity in older age show that DA values increase with age past the 6th decade of life (Michielse et al., 2010), and that groups with AD show higher DA values than those without (Bosch et al., 2012; Salat et al.,

2010). Rather than high DA values being an indicator of maintained neural integrity as claimed by Anderson, Grundy et al., evidence suggests that it may in fact be the reverse; i.e., poorer brain health. The loadings of variables in the factor plot from the current study support the interpretation that DA and DR are associated with cognitive decline, as they correlate together along dimension 1 but negatively correlate with FA in the same dimension, which is presumably associated with cognitive health (Figure 2). The overall PCA scores provide a holistic assessment of white matter integrity while simultaneously accounting for most of the variance in the original variables.

In all three previous studies of white matter and bilingualism, both bilinguals and monolinguals were considered cognitively healthy older adults. This assumption was confirmed, in part, by similar cognitive performances across groups within each of the studies. Yet, despite this cognitive and clinical similarity, bilinguals were more likely to present with poorer white matter integrity than monolinguals (e.g., Anderson, Grundy et al., 2018; Gold et al., 2013) rather than the reverse (e.g., Luk, Bialystok et al., 2011). This was true for the participants in the study by Anderson, Grundy et al., from which the bilingual group in the current study were drawn. What is not addressed by these studies is what would be the cognitive outcomes for monolinguals if their brain integrity were at the level of bilinguals. Put another way, what would the cognitive and clinical outcomes be for monolinguals if they "swapped brains" with bilinguals?

The current study was designed to investigate this question. After matching older adult monolingual participants to bilinguals on sex, age, education, and brain integrity (as measured by a primary PCA score) it was shown that monolinguals whose brain parameters were matched to bilinguals showed more advanced clinical decline. This was reflected in more clinical diagnoses

of MCI and AD than what would be expected by chance within the monolingual matched sample and lower cognitive performance, as seen by poorer MMSE scores. To our knowledge, this is the first study to examine cognitive and clinical outcomes between bilinguals and monolinguals by using this "brain swap" technique to match individuals on brain health rather than the reverse.

Despite the relatively poorer neural health of bilinguals than monolinguals in the studies by Gold et al. (2013) and Anderson, Grundy et al. (2018), bilinguals still had comparable cognitive performance. In the present study, monolinguals showed poorer cognitive performance and poorer clinical outcomes when matched to bilinguals on brain health, a finding consistent with our predictions regarding the contribution of bilingualism to cognitive reserve. Bilinguals performed near ceiling on MMSE scores, while monolinguals' scores in the matched sample were borderline to MCI (despite being matched for age, sex, education, and brain health). These results suggest that monolinguals are less able to cope with neural degeneration than bilinguals.

White matter integrity was selected as the measure by which to judge brain health, in part as it provides an index of connectedness between neural networks, and as a spiritual extension of the work by Anderson, Grundy et al (2018). However, from the perspective of cognitive reserve, using other measures of brain health such as cortical thickness or cerebral atrophy would theoretically lead to a similar pattern of results as seen in the current study. A study by Pettigrew et al. (2017) examined cortical thickness in cognitively normal individuals using cognitive reserve as a factor in predicting progression to MCI. Their findings showed that higher mean cortical thickness at baseline was associated with a reduced risk of clinical symptom onset within 7 years of initial scan, and higher cognitive reserve was similarly associated with reduced symptom onset in general. However, an interaction between the two factors suggested that individuals with low cognitive reserve were more likely to develop clinical symptoms further out

from baseline than those with high cognitive reserve; i.e., high reserve individuals were better able to compensate for cortical atrophy that occurs in the earlier stages of disease progression. If one considers that the participants in this study by Pettigrew et al. were an average age of ~57 years at baseline then their results suggest that testing high reserve against low reserve individuals in later years of life, and consequently atrophy, would lead to similar results as the present study.

Considering that the present results seem to rule out whole-brain white matter integrity as the mechanism by which bilingualism modulates cognitive reserve (i.e., white matter integrity is poor in bilinguals despite normal cognition and thus other measures must be responsible for cognitive maintenance), four studies examining cerebral atrophy in bilingual and monolingual older adults are worth mentioning here. In the first, Abutalebi et al. (2015) found greater grey matter volume in the left and right inferior parietal lobules for cognitively normal bilinguals compared to their monolingual peers. The second study, by Costumero et al. (2020), found reduced parenchymal brain volume for bilinguals than monolinguals in a sample of patients with MCI. The third study, by Schweizer et al. (2012), showed greater cerebral atrophy in bilinguals than monolinguals in a sample of patients with probable AD. The fourth study, by Duncan et al. (2018), showed thicker cortex in language and cognitive control regions for multilinguals than monolinguals in a sample of patients with MCI, but similar or worse thickness in these regions for multilingual than monolingual patients with AD. In all four studies, the language groups were matched on cognitive status. Together, it appears that in older age the stage at which measures are taken could greatly impact the conclusions that are drawn. Older adult bilinguals may have greater grey matter volume in normal cognition but undergo cerebral atrophy at a quicker rate than monolinguals once they progress to MCI and AD, with results during decline (MCI) being

less clear. This is in line with cognitive reserve theories and may point to the mechanism by which bilinguals are better able to cope with decline while maintaining cognitive performance, via an accumulation of neural resources (i.e., brain anatomy and physiology mediating cognitive processes) prior to decline.

For the current study, the MMSE was used as an indicator of cognitive level, as it was available in both the bilingual and ADNI samples. Meta-analyses of studies examining the effect of bilingualism on clinical status and dementia show that a majority use the MMSE as a measure of cognitive performance (Anderson et al., 2020; Brini et al., 2020), although some have argued that there are issues with the MMSE as a diagnostic tool due to its low sensitivity, requiring the use of other tests in tandem for optimal results (e.g., Berkes et al., 2020; Mitchell, 2009). Ideally, the bilingual group in the current study would have been tested on the neuropsychological battery used in the ADNI sample to align cognitive performance to the monolingual group across a wider array of measures, although this was not possible in the present study.

The current study also has other limitations. As mentioned earlier, monolingual patients were selected from the ADNI database and as such did not have objective language measures to fully confirm language usage or proficiency. Future studies would be better suited by having more detailed language information from participants, including but not limited to all languages known or studied along with proficiency, ages of acquisition, and daily language exposure. Another limitation inherent in the ADNI database is the usage of different MRI scanner models across hospital sites. Variability in data between sites could be due to differences in acquisition protocols, scanning parameters, and scanner manufacturers. However, a positive feature of the ADNI dataset is a standardised scanning protocol across collection sites to minimise differences inherent in scanner model, alongside quality control at the Mayo clinic to ensure minimal

differences in scans across sites. Reassuringly, the analyses examining effects of site on PCA scores in the current study did not reveal any significant trends. Regardless, future studies should aim to collect all images using the same scanner model and software, or failing that option, follow the advice of Fortin et al (2017) to harmonize data collected across different sites.

The current study adds unique evidence to support the claim that bilingualism is a cognitive reserve factor. In contrast to typical studies of cognitive reserve in which neural markers are outcome measures, in this case, individuals were matched on *neural parameters* derived from diffusion tensor imaging and diagnostic status was compared. The percentage of monolingual individuals affected by MCI or AD was ~14% more than expected by chance when matched on brain health to a bilingual cognitively normal sample (~41% for matched monolinguals compared to ~27% on an unmatched sample). Furthermore, these results cannot be explained by sex, age, or education, suggesting that bilingualism confers a unique protective benefit. Bilingualism and its associated benefits across neural networks (e.g., Bialystok et al., 2012; Brini et al., 2020) seems, at a minimum, to postpone deleterious effects of aging and poor brain health, whereas monolinguals are more likely to suffer the consequences of earlier cognitive decline. The current findings provide new evidence that bilingualism protects individuals from negative clinical outcomes in the face of aging and neural degeneration.

Study 2: The Effect of Demographics, Biomarkers, and Genetic Factors on Clinical Outcomes for Monolinguals and Bilinguals in the ADNI Database

The previous study investigated bilingualism as a factor of cognitive reserve using a novel paradigm that matched neural health measures between monolingual and bilingual participants. Bilingual status was associated with a reduced chance of MCI or AD diagnosis compared to monolinguals. The conclusion was that bilingualism helps to protect individuals from the negative outcomes associated with aging.

As a brief review, cognitive reserve refers to the concept of preserved cognitive functioning in the face of aging and neural degeneration. This preservation is believed to be accomplished in part through the strengthening of neural networks by accumulation of neural resources prior to decline, as well as compensation in alternate networks once degradation begins (Cabeza et al., 2018). Higher levels of cognitive reserve, specifically as related to bilingualism, have consistently been shown to improve clinical outcomes in older adults by delaying symptoms of AD (e.g., Alladi et al., 2013; Bialystok et al., 2007; Chertkow et al., 2010; Woumans et al., 2015; Zheng et al., 2018). Although some studies have failed to find differences in clinical outcomes between language groups (e.g., Lawton et al., 2015; Sanders et al., 2012), recent meta-analyses have largely confirmed the positive findings (Anderson et al., 2020; Brini et al., 2020).

Bilingualism, and cognitive reserve more broadly, are not the sole determinants of clinical outcomes in aging. Other factors that have been shown to be involved in clinical outcomes include biomarkers, such as the amyloid-β (Aβ) peptide and tau microtubules in cerebrospinal fluid (CSF), and genetic factors, such as the presence of the apolipoprotein E

(APOE) ε4 allele on chromosome 19. However, these biomarkers and genetic risk factors have rarely been discussed in tandem with language use as relating to aging and cognitive decline. Considering that these "hidden" genetic factors may well influence clinical outcomes, studying the interactions between these exogenous and endogenous factors is vital to furthering our understanding of both cognitive and clinical outcomes in aging.

The first biomarker, the Aβ peptide, is a 38- to 43-amino-acid peptide that is derived from the amyloid precursor protein, a protein which is involved in processes such as neuronal development, signalling between neurons, and neuronal homeostasis. Amyloid-β is primarily produced in the brain (Laird et al., 2005) with two major final forms, the 42-amino-acid Aβ42 and the 40-amino-acid Aβ40, which compromise ~5-10% and ~80-90% of total Aβ volumes in the brain, respectively (Murphy & Levine, 2010). Amyloid-β has historically been thought to play a role in neuronal loss and cognitive impairment through the "amyloid cascade hypothesis" (Hensley et al., 1994; Selkoe, 1994), which posits Aβ as the primary cause of AD due to accumulation of senile plaques and intercellular deposition of neurofibrillary tau tangles (the second biomarker of note). As such, levels of Aβ42 found in CSF show an inverse relationship with disease progression – as plaques accumulate in the parenchymal tissue of the brain, less Aβ is found in CSF (Blennow et al., 2010). In contrast, Aβ40 levels remain largely unchanged along the trajectory from normal cognition to AD (e.g., Shoji et al., 1998). Levels of CSF Aβ42 have therefore been thought to aid in the diagnosis of AD, or in predicting future conversion from MCI to AD, although recent research further posits that the ratio of Aβ42/40 is a better diagnostic indicator of AD (see Hansson et al., 2019, for a review).

Despite the central role that Aβ has historically played in the field of AD research, pharmaceutical treatments for AD have either not targeted Aβ as their means of action or have

failed in their attempts when they do target Aβ (Mehta et al., 2017). Furthermore, studies that used levels of plasma Aβ as a predictor of AD (as opposed to CSF Aβ) have been inconsistent in their findings (Rissman et al., 2012). Similarly, studies investigating the effects of cognitive reserve as a modifier of Aβ levels show differing results depending on the medium in which Aβ is examined (i.e., in plasma or CSF). For example, Yaffe et al. (2011) examined the link between cognitive reserve (as measured by education) and plasma Aβ42/40 ratio on cognitive decline in a large prospective cohort of community-dwelling older adults. The authors found that a low Aβ42/40 ratio at intake is associated with greater cognitive decline; however, no association between Aβ levels and cognitive decline was present in individuals with high cognitive reserve. Put another way, a low level of plasma Aβ42/40 was a greater risk factor for cognitive decline in those with low levels of cognitive reserve. This finding provides evidence that cognitive reserve factors can modify the association between decline and biomarkers such as Aβ, which is exactly in line with predictions made by cognitive reserve theory. Conversely, a study also using education as a proxy of cognitive reserve to examine levels of CSF Aβ, rather than plasma Aβ, in cognitively normal and impaired individuals found no effect of reserve nor interaction with age (Almeida et al., 2015), although the authors did not examine cognitive decline as a factor. Soldan et al. (2013) used a composite score of reading, vocabulary, and education to determine cognitive reserve in middle-aged cognitively normal individuals to then determine risk of developing symptoms of preclinical AD. In contrast to Almeida et al., this study examined decline and found main effects of cognitive reserve and baseline Aβ42 levels such that higher levels of reserve were associated with lower risk of developing symptoms, and lower levels of Aβ42 were associated with higher risk of developing symptoms. Another study used bilingualism, rather than education, as a measure of cognitive reserve to see how CSF biomarkers are impacted but

found no differences between cognitively normal monolinguals and bilinguals in CSF Aβ levels (Estanga et al., 2017). This study, however, included middle-aged participants rather than older adults, looked at Aβ42 levels rather than the more promising Aβ42/40 ratio, and did not investigate cognitive decline as a factor. Taken together, these studies provide conflicting reports on how cognitive reserve may modify decline and Aβ levels, although many variables differ between the studies (such as examining plasma or CSF Aβ, and variable selected as a measure of reserve). More research is needed to clarify the findings.

Tau, the major microtubule-associated protein in developed neurons, is another biomarker of interest in AD. Tau functions as structural support and stabilization of microtubules in neurons, which, in turn, allows transport of essential substances, such as neurotransmitters, throughout cells (Weingarten et al., 1975). The presence of phosphate in the brain leads to phosphorylation of tau, which is necessary for tau to bind with microtubules and stimulate their assembly. However, excess levels of phosphate, or hyperphosphorylation, depresses the biological activity of tau. In the brain of someone with AD, tau is three- to four-fold more hyperphosphorylated than that found in a neurotypical adult brain (Iqbal et al., 2010), and, as such, is a prime indicator of brain health and disease progression. This hyperphosphorylated tau, unable to bind with microtubules, falls apart and forms into clumps in the cell body (neurofibrillary tangles), as well as in dendrites and axons (threads). In the brain of someone with AD, this tau neuropathology seems to follow stages as laid out by Braak and Braak (1991), such that tau accumulates in the transentorhinal region first, followed by limbic regions, and neocortical areas last. Tau's function in neurons is of such importance that recent research posits tau pathology, and not Aβ, as the primary driver of AD (Brier et al., 2016; Kametani & Hasegawa, 2018; Ossenkoppele et al., 2016).

Relatively recent research has examined the association between tau and cognitive reserve using a range of measures. Across labs this includes examining tau levels in CSF samples or *in vivo* using positron emission tomography (PET), and considering either education or bilingualism as proxies of cognitive reserve. For example, the study by Almeida et al. (2015) used education as a proxy of cognitive reserve and investigated CSF levels of total tau (t-tau) and phosphorylated tau (p-tau) in both cognitively intact and impaired older adults. Whereas their previously mentioned results regarding Aβ showed no effect of reserve, they found a significant interaction effect between tau and cognitive reserve, such that both t-tau and p-tau levels were higher in older adults with low education than in age-matched peers with high education. Further, tau levels were higher for cognitively impaired adults than for cognitively normal adults. This is in contrast to Aβ findings that usually see the reverse, i.e., lower CSF Aβ values in cognitively impaired individuals due to greater accumulation of neuropathology staying in the brain. The finding by Almeida et al. is somewhat supported by the previously mentioned study by Estanga et al. (2017) who also investigated CSF t-tau and p-tau in middle-aged monolingual and bilingual individuals. Unlike Almeida et al., the study by Estanga et al. used bilingualism as their measure of cognitive reserve. They found lower t-tau values in early bilinguals than in monolinguals and late bilinguals, an effect they attributed to increased cognitive reserve. However, participants in this study were younger, so no symptoms or diagnoses of AD had been met. Because of this, it is hard to compare the results in a meaningful way to studies that use older or clinically impaired adults. Still, it provides a framework upon which future studies can build.

One recent study by Hoenig et al. (2017) used PET to examine tau pathology *in vivo* in high education versus low education AD patients with similar levels of clinical severity.

Compared to the previous studies that examined levels of tau in CSF in relation to cognitive reserve, this study attempted to show greater pathology across regions reflecting different Braak stages. Indeed, the authors found more diverse tau pathology such that high-education AD patients showed pathology in regions associated with later Braak stages V and VI, whereas low-education patients only showed pathology in areas related to Braak stages III or IV. This pathology was seen despite comparable levels of cognitive impairment across groups. These findings support the cognitive reserve hypothesis that cognitive functioning is preserved in the face of increased pathology. However, this finding is at odds with the results and interpretations of Almeida et al. (2015) and Estanga et al. (2017). Previous studies have shown positive correlations between tau pathology in the brain and levels in CSF (Buerger et al., 2006; Tapiola et al., 2009), so given greater pathology in high-reserve individuals in the study by Hoenig et al., one could reasonably expect greater levels of tau in CSF in these individuals compared to low-reserve individuals. This contrasts with the previous studies that claimed the opposite: high reserve leads to reduced levels of CSF tau. A recent study suggests examining tau cleaved at amino acid 368 as a potential biomarker that shows negative correlation with AD pathology (Blennow et al., 2019), but further research is needed to tease apart these interactions.

In addition to Aβ and tau, one of the largest genetic risk factors of AD is the presence of the APOE ε4 allele. APOE exists as three polymorphic alleles, ε2, ε3, and ε4, but only the ε4 allele is associated with an increased risk of developing AD. The ε4 allele has a worldwide frequency of ~14% yet jumps to ~40% frequency in patients diagnosed with AD (Farrer et al., 1997; see Ward et al., 2012, for differences between countries). Additionally, an older study by Seshadri et al. (1995) used a Bayesian calculation to show that adults with at least one APOE ε4 allele had a 29% risk of developing AD in their lifetime compared to a 9% risk in those adults

with no ε4 allele. This predictive aspect of the ε4 allele has been recognised for over two decades and found to be reliable (Ward et al., 2012). One mechanism by which APOE influences development of AD is that the APOE genotype strongly affects deposition rates of Aβ in the brain, such that APOE ε4 carriers show greater abundance of senile plaques compared to noncarriers (Kok et al., 2009; Morris et al., 2010; Polvikoski et al., 1995; Schmechel et al., 1993).

In addition to these increased rates of AD and plaque accumulation, APOE ε4 carriers also show greater rates of cognitive decline in middle- and older age than noncarriers. In one prospective longitudinal study by Caselli et al. (2007), healthy participants aged 50–69 were tested on a range of cognitive domains using a neuropsychological test battery every two years, starting in 1994. Participants in the 60–69 age range who were homozygous for the APOE ε4 allele showed accelerated and earlier decline in one or more domains compared to heterozygous ε4 carriers or non-carriers. This finding follows from an earlier study by the same group that showed verbal memory declines in carriers of the ε4 allele despite being asymptomatic for any diagnostic assessment of MCI or AD (Caselli et al., 2004). Similarly, cognitively normal 80-year-old ε4 carriers showed worse performance on a verbal and nonverbal reasoning task than noncarriers, despite showing no differences on the same task in early childhood (Deary et al., 2002). These declines, along with other studies, highlight the cognitive deficits that APOE ε4 carriers sustain even in the absence of a clinical diagnosis of impairment (Caselli et al., 2011; Izaks et al., 2011).

Given strong evidence that the APOE ε4 allele is a significant risk factor for cognitive decline and subsequent diagnosis of MCI and AD, it follows that research examining how this decline can be mitigated should be at the forefront. Investigating cognitive reserve factors and

their interaction with the effects of APOE is one obvious avenue to pursue. Ferrari et al. (2013) showed in a long-term follow up study of a population-based cohort that higher education and physical activity reduce incidence of AD in APOE ε4 carriers. The effect was significant enough that carriers of the ε4 allele with high education had similar hazard ratios of developing AD as noncarriers. Although the study did not specify education as a proxy of cognitive reserve, the authors posited brain and cognitive reserve as the possible mechanisms by which this effect was seen. This finding follows from previous work done by the same group showing the protective effects of high education on development of dementia and AD, both with and without the context of APOE information (Qiu et al., 2001; Wang et al., 2012). Another longitudinal study defined cognitive reserve as an index consisting of education, reading, and vocabulary abilities, and examined its interaction with the APOE ε4 allele in middle- to older-age adults on time to onset of clinical symptoms of decline (Pettigrew et al., 2013). Cognitive reserve and APOE ε4 status independently predicted clinical symptom onset, but there was no interaction between the two.

To date, very few studies have investigated the effect of APOE and bilingualism as a proxy of cognitive reserve on cognitive outcomes in older age. Crane et al. (2010) examined cognitive decline in Japanese-American men aged ~75 years, and found no effect of speaking or writing Japanese (in addition to English) on rates of decline after accounting for APOE allele status, among other confounding variables. Their conclusion was that multilingualism does not contribute to the cognitive reserve hypothesis. This conclusion was similarly reached by Hack et al. (2019) who investigated dementia onset and the effects of multilingualism using data from the Nun Study (Snowdon et al., 1996). No effect of bilingualism was apparent on dementia onset times after accounting for APOE status, despite independent effects of APOE and written linguistic ability. The authors of this study did find that individuals who spoke four or more

languages were significantly less likely to develop dementia than monolinguals, but this effect was minimised once linguistic ability was included in the model alongside APOE status and education. Cognitive reserve, broadly speaking, has been shown to reduce the rates and odds of developing dementia, but in the presence of APOE, the results are less clear. Interpreting the findings is confounded further depending on how cognitive reserve is measured, be it through education, aerobic exercise, or multilingualism. Given that the presence of APOE ε4 is possibly the largest risk factor for developing dementia, it is entirely possible that the harm of carrying an ε4 allele outweighs any positives that may be gained from higher cognitive reserve. Contributions to the sparse literature surrounding bilingualism and its interaction with APOE on cognitive decline, dementia onset, and dementia rates is needed to help clarify the findings thus far. The current study is an attempt to do so.

As discussed in Study 1, the ADNI database is a remarkable resource for large-scale investigations of AD and its many contributing factors. Aside from basic demographic information for all participants such as age, education, or ethnicity, biomarker and genetic information is also available for those participants willing to undergo blood draws and lumbar punctures. Diffusion magnetic resonance imaging (dMRI) is similarly available for those participants able to participate in the procedure. Together, the information provided in the database allows for a range of questions to be asked and investigated. The current study utilises participant language information to determine bilingual status, which will be considered as a proxy of reserve. Biomarkers and genetic factors including Aβ, tau, and APOE genotype are used in tandem with brain health determined from dMRI as predictors in order to compare clinical outcomes and cognitive decline between monolinguals and bilinguals. Given the previous literature that bilingualism contributes to cognitive reserve, bilinguals should present

with better cognitive outcomes and more favourable diagnoses than monolinguals in the face of comparable, or possibly worse, neuropathology and risk factors. However, when one considers the conflicting results seen in the literature concerning bilingualism, biomarker, and genetic factor effects, it is possible that results will not be so clear cut. Analyses will necessarily have to account for this complexity and will serve as a useful direction for further research.

Methods

Participants

Data for all participants were obtained from the ADNI database (adni.loni.usc.edu), inclusive across all phases of ADNI research (i.e., ADNI-1, ADNI-GO, ADNI-2, and ADNI-3). Subject data are stored in the ADNI database across multiple Excel files corresponding to each variable of interest (e.g., one file for demographic information and another for neuropsychiatric assessment results), so one file per variable was downloaded and subsequently merged based on patient ID. As the current study is not longitudinal, the data from multiple assessment dates per participant were not needed; rather, the earliest date for neuropsychiatric assessments, biomarker assays, and MRI scans relative to participant screening was selected and used for each participant. An additional variable of language group was created to classify participants as monolingual or bilingual. Monolinguals were classified as such if their primary language and preferred language of testing were both English, and participants were neither Latino nor Hispanic to rule out participants with a strong likelihood of using or being exposed to Spanish. There were no racial criteria for monolinguals. Bilinguals were classified as such if the language used for testing was English, but home language was any language other than English. There were no ethnic or racial criteria for bilinguals. There is the possibility that some bilinguals used

more than two languages (so are actually multilinguals), but this was unknown and unable to be examined. Therefore, the term bilingual is used throughout.

The ADNI database is often missing data for participants from one or more variables of interest due to the longitudinal nature of testing in the ADNI protocol, combined with older participant age and diagnoses leading to participant attrition. Monolingual participants were therefore selected based on relative completeness of biomarker and genetic factor data. There were significantly fewer bilinguals in the overall dataset, so all were selected for inclusion. After selection and dataset trimming, there were 577 monolinguals and 64 bilinguals.

Data Acquisition

Details of the procedure for testing are available on the ADNI website, but the general procedure is briefly discussed here. Possible eligible participants were selected from the ADNI database based on diagnoses, medication, and last MMSE score. Participants were contacted and, along with a study partner, attended a screening session. The data obtained during screening that are included in the current study included demographic information, a blood draw for APOE genotyping, testing on the MMSE and the Logical Memory test of immediate and delayed verbal memory, and a screening MRI test. Further appointment dates completed the relevant data collection, including completion of a neuropsychological assessment battery, further MRI and DWI scans, and a lumbar puncture for biomarker testing in at least 20% of participants.

For MRI scanning, participants were tested at various sites across the United States and Canada, but all used a GE, Siemens, or Phillips scanner. All research sites were capable of conducting 1.5 Tesla MRI scans, but 3.0 T scans were performed when possible. DTI scans were whole-brain 48-direction diffusion-weighted images, obtained with TR = 7200 ms, TE = 56, and

voxel size of 2.0 mm^3 for all scanner models. All scans were screened at the Mayo Clinic for quality control before being accepted into the ADNI database.

For biomarker and genetic factor testing, all ADNI sites are provided with the appropriate testing equipment. Blood samples for APOE genotyping are taken at Screening appointment, while lumbar punctures for subsequent Aβ and tau analyses are performed at the Baseline appointment for those participants willing to undergo the procedure. All biological samples are sent to the ADNI Biomarker Laboratory at the University of Pennsylvania for APOE genotyping and biomarker assay. Batch analyses of $A\beta_{1-42}$, t-tau, and p-tau_{181} (i.e., tau that has been phosphorylated at threonine 181) used a fully automated Roche Elecsys and cobas e immunoassay analyser system (Bittner et al., 2016).

DWI Data Processing

Processing was performed in part using the MRtrix3 package (www.mrtrix.org), which included the initial step of de-noising using the dwidenoise function. Eddy distortion correction was then performed using the eddy function in FSL. Following this, a diffusion tensor was fit using the DTIFIT command from FSL. Finally, Tract-Based Spatial Statistics (TBSS; Smith et al., 2006) included in FSL was used on the FA images from the DTIFIT output. Each participant's FA images were aligned to a 1x1x1 mm standard space using nonlinear registration, and subsequently all merged together into a single 4D image. The mean of all FA images was created, as well as a thinned FA skeleton representing the centers of tracts common to all participants. This method was then applied to DA and DR data, and tracts were identified post-hoc using the John Hopkins University DTI based probabilistic white matter atlas included in FSL.

Due to data availability discrepancies between the two language groups, the numbers of participants with DTI values following these steps were 19 bilinguals (selected based on availability of scans) and 82 monolinguals (selected via serial search for participants with Magnetization Prepared Rapid Gradient Echo MRI and axial DTI scans). As was done in Study 1, principal components analysis was conducted using FA, DA, and DR values for these participants. Approximately 72% of brain health variance was accounted for in the first principal component. This loading score per participant was then used in subsequent analyses as a reflection of overall brain health where negative scores reflect higher values in FA and positive scores reflect higher values in DA and DR.

Data Analyses

The overall dataset of 641 participants had missing data at random, as well as a large difference in participant numbers between language groups. As such, analytic methods were required that could accommodate these difficulties. Functional gradient boosting, or generalised boosted regression modeling (GBM), is one such method. In brief, GBM is a machine-learning algorithm that determines a predictive model of best fit for a given dataset. GBM accomplishes this prediction by building the model in sequential stage-wise fashion from iterative "decision trees" – regression models that are rough and gradually increase in predictive accuracy as they focus on minimising errors from previous models. Because decision trees are added sequentially, GBM is a slow process that increases in accuracy as more trees are considered. Freund and Schapire (1997) introduced one of the first adaptive boosting algorithms, AdaBoost, in which weak hypotheses' predictive probabilities are summed to reach a reliable overall model of fit. As a simple analogy (used by Freund and Schapire), the process is similar to betting on horse races after aggregating "rules-of-thumb" given by a professional gambler to reach an outcome that is

more favourable than following one rule alone. Friedman (2001) developed a similar process he referred to as a Gradient Boosting Machine. As opposed to AdaBoost which sums weak learning algorithms, Friedman's process relied on explicit regression-based models of prediction. In short, "boosting" simply refers to the process of creating a prediction rule about data by building upon weaker hypotheses. This process is highly customisable and shows remarkable success in practical applications (see Natekin & Knoll, 2013, for a review).

The current study makes use of boosting in the R package *Twang* (McCaffrey et al., 2004) to estimate propensity scores. Weighting is then applied using these propensity scores to estimate the average treatment effect on the population (ATE). The ATE essentially measures how the outcome of interest would change if *everyone* in the population received a certain treatment (treatment A) relative to if *everyone* in the population received a different treatment (treatment B). Mathematically this is roughly accomplished by taking the difference between the mean effect of treatment and the mean effect of no treatment on everyone in the sample. For a study with three treatment effects, this equates to 3 ATEs: $\mu_A - \mu_{B;}$ $\mu_A - \mu_C$; and $\mu_B - \mu_C$. The average treatment effect on the treated (ATT) is the other potential estimand where the comparison is mean outcome of, for example, Treatment A on Group A compared to Treatment B on Group A. This differs from ATE in that each group is compared against each other based on potential treatment, rather than an overall population effect of treatment. Given three treatment conditions, this results in 6 ATTs: two comparisons for each treatment group. The overall effect of weighting propensity scores based on either ATE or ATT results in better balanced data from which further analyses (e.g., ANOVAs) can be investigated. This outcome is similar to the aim of one-to-one propensity score matching (used in Study 1), but unlike

matching, weighting accounts for missingness of data and allows more participants to be retained for analyses.

Several GBM analyses were run to investigate different questions. The first analysis examined the effects of demographic variables on clinical diagnoses in only the monolingual participants. This was used as a general proof of concept to ensure the regression models and weighting were functioning as expected, i.e., returning better balanced data. The second model included demographic information, biomarkers, and genetic factors as predictive covariates of clinical diagnosis in monolinguals. The third model compared monolingual and bilingual participants and included demographic information, biomarkers, and genetic factors as predictive covariates of language group inclusion. After GBM analyses were run to create propensity scores that were then used for weighting, linear models were conducted to examine outcome variables of interest such as white matter PCA scores and MMSE. The results of these models are reported in the next section.

Analyses and Results

Participant Information

Demographic, biomarker, genetic factor, and white matter PCA score information for all monolingual and bilingual participants are reported in Table 2. Analyses between groups were not done at this point, but these variables were considered in the following GBM models.

Table 2. *Participant information.*

	Monolinguals			Bilinguals		
	CN	MCI	AD	CN	MCI	AD
Total N	263 (152 F)	188 (75 F)	126 (49 F)	15 (8 F)	28 (18 F)	5 (2 F)
Age	73.4 (7.1)	74.5 (8.0)	74.7 (8.2)	71.7 (6.9)	74.1 (8.5)	79.6 (8.0)
Education	16.6 (2.4)	15.6 (2.8)	15.5 (2.9)	16.8 (2.7)	15.7 (3.6)	17.2 (2.7)
Aβ42 (pg/mL)	1273 (631), N = 263	934 (619), N = 188	658 (365), N = 126	1186 (488), N = 7	1021 (513), N = 7	580 (265), N = 2
Aβ40 (pg/mL)	18847 (5333), N = 201	18679 (6194), N = 56	15433 (4488), N = 16	*NA*	*NA*	*NA*
p-tau$_{181}$ (pg/mL)	21.4 (9.7), N = 263	31.2 (17.0), N = 188	35.4 (15.0), N = 126	18.9 (13.8), N = 7	28.6 (18.7), N = 7	63.4 (1.4), N = 2
t-tau (pg/mL)	234.3 (88.0), N = 263	316.2 (151.6), N = 188	353.5 (127.9), N = 126	214.8 (145.0), N = 7	294.5 (153.8), N = 7	581.2 (8.8), N = 2
APOE4	67 (32.4%), N = 213	88 (49.4%), N = 178	83 (70.8%), N = 120	1 (20.0%), N = 5	2 (40.0%), N = 5	1 (50.0%), N = 2
PCA	-0.41 (1.44), N = 56	-0.067 (1.46), N = 17	1.13 (1.49), N = 9	1.41 (.44), N = 5	.47 (1.24), N = 14	*NA*

Note: Continuous variables are presented as means with standard deviations in parentheses. APOE4 values are expressed as the number of participants with at least one APOE ε4 allele and as a percentage of the diagnostic subsample.

Model 1

Model 1 determined propensity scores for base demographic variables on clinical diagnosis in the monolingual sample. The formula used in R was:

```
GBM1 <- mnps(DIAGNOSIS ~ SEX + AGE + EDU + RACCAT,
                 data = monolinguals,
                 estimand = "ATE",
                 verbose = FALSE,
                 stop.method = c("es.mean"),
                 n.trees = 6500)
```

The input variables included sex, age, education, and race, with 6500 iterative trees run. The final number of trees selected was to minimise the average effect size difference, based on visual inspection of plots of the model. The estimate used is the ATE, and the measure of balance by which the model is fitted is the absolute standardised mean difference (ASMD), otherwise

known as mean effect size ("es.mean"). Balance plots and tables were inspected to determine the quality of propensity score weights and are included in the Appendix (as is the case for all subsequent models) but are discussed here. There was good optimisation of balance measures for all diagnoses after 6500 iterations (Model 1, Plot 1). Plots of the ASMD between groups on the input covariates largely show decreasing values for pairwise comparisons post-weighting compared to pre-weighting (i.e., the effect size of a comparison on age, for example, between CN and MCI subjects was lower post-weighting). Balance tables for each diagnosis were created that gave information on input covariates before and after weighting. For a specific diagnosis, the comparisons would be against the other two diagnoses (e.g., the CN balance table compares CN subjects against MCI and AD subjects combined). For unweighted CN participants there were significant differences in sex ($p < .001$), and education ($p < .001$), with marginal differences in age ($p = .06$). Post-weighting these differences disappeared for all variables (ps > .2). For unweighted MCI participants there were significant differences in sex ($p = .008$) and education ($p = .008$), but these differences disappeared post-weighting (ps > .6). For unweighted AD participants there were significant differences in sex ($p = .023$) and education ($p = .012$), but no differences post-weighting (ps > .4). The relative influence of each variable on the calculation of propensity scores was also calculated, with age accounting for ~67%, education accounting for ~20%, and sex accounting for ~9% in CN participants; age accounting for ~77% and education accounting for ~15% in MCI participants; and age accounting for ~75%, education accounting for ~13%, and sex accounting for ~9% in AD participants.

Once propensity scores were created and variables were weighted based on these scores, between groups analyses examining white matter PCA scores (where higher values reflect larger DA and DR values) were conducted on both unweighted and weighted values. These results

were then compared to see how weighting affects outcomes, with results presented in Table 3. A linear model using the unweighted data to examine PCA scores as a function of diagnosis was significant, $F(2, 79) = 4.41$, $p = .015$. There was a significant difference between CN and AD groups ($p = .004$), and between MCI and AD ($p = .05$). The weighted data, however, had to be examined using a linear model fit to a complex survey design, with inverse-probability weighting and design-based standard errors. PCA scores as a function of diagnosis were examined using a Wald test to account for the weighted survey design, performed with the *survey* package in R. As with the unweighted data, this model was significant, $F(2, 79) = 3.22$, $p = .05$. The estimate for scores of AD subjects was 1.25, SE = .57, which was different from zero ($p = .03$). The difference between CN and AD was significant ($p = .01$), wherein CN subjects had a reduced estimate compared to AD subjects by -1.51, SE = .60. The difference between MCI and AD subjects was not significant ($p = .07$), although MCI subjects had a reduced estimate compared to AD subjects by -1.24, SE = .67.

Table 3. *Unweighted, weighted Model 1, and weighted Model 2 estimates.*

		CN	MCI	AD
Unweighted	PCA Estimates	-0.23 (.51)***	0.098 (.58)*	1.27 (.47)
	MMSE Estimates	29.1 (.2)***	27.0 (.2)***	23.1 (.2)
Model 1	PCA Estimates	-0.27 (.60)**	0.0059 (.67)	1.25 (.57)
	MMSE Estimates	29.1 (.3)***	27.1 (.3)***	23.2 (.3)
Model 2	PCA Estimates	-0.17 (.45)**	0.16 (.57)*	1.30 (.41)
	MMSE Estimates	29.1 (.2)***	27.2 (.3)***	23.2 (.2)

Note: Variables are presented as mean score estimates with standard errors in parentheses. Comparisons are against AD values where * $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$

Scores on MMSE were also investigated using unweighted and weighted values. The unweighted model was significant, $F(2, 570) = 478.7$, $p < .001$. The estimate of AD subjects was 23.1, SE = .16, which was significantly lower than CN patient scores who had an increased

estimate value of 6.0, SE =.19, $p < .001$. MCI subjects similarly had a higher estimate than AD subjects with an increase of 3.9, SE = .21, $p < .001$. The overall model for the weighted values was also significant, $F(2, 570) = 254.4$, $p < .001$. Results were similar to the unweighted model. The estimate of AD subjects was 23.2, SE = .27, which was significantly lower than CN patient scores who had an increased estimate value of 5.9, SE =.28, $p < .001$. MCI subjects similarly had a higher estimate than AD subjects with an increase of 3.9, SE = .31, $p < .001$.

Although the results of PCA and MMSE scores between unweighted and weighted samples seem similar from a simple statistical threshold of $p = .05$, there is a difference in the quality of results. The results of the survey design reflect differences between diagnostic groups after accounting for and weighting demographic variables to reduce any confounding effects. This technique is applied to the subsequent analyses, with implications evaluated in the Discussion. The results of this model showed healthier brains for CN participants than MCI participants, who in turn had healthier brains than AD participants. The same pattern is true for scores on the MMSE where CN participants performed better than MCI participants and AD participants.

Model 2

The second model created propensity scores using demographic variables, as well as biomarker and genetic factor information in the monolingual sample. The formula used in R was:

```
GBM2 <- mnps(DIAGNOSIS ~ SEX + AGE + EDU + RACCAT + ABETA42 + ABRATIO + PTAU
                  + TAU + APOE,
                  data = mono.gbm2,
                  estimand = "ATE",
                  verbose = FALSE,
                  stop.method = c("es.mean"),
                  n.trees = 1100)
```

The input variables again included demographic variables as in model 1, but now also included Aβ42, Aβ ratio (Aβ42/Aβ40), p-tau$_{181}$, t-tau, and APOE ε4 allele information. Balance plots and

tables are again included in the Appendix, with relevant findings presented here. One thousand one hundred iterations were run, which found good optimisation of balance measures for all subjects (Model 2, Plot 1). Plots of the ASMD between clinical groups on the input covariates show a general trend of decreasing values from pre- to post-weighting. This decrease was more pronounced in CN and AD participants compared to MCI subjects, but MCI subjects had lower ASMD values in general (Model 2, Plot 3). As with Model 1, balance tables were assessed for weighting effects on input covariates between clinical groups. For CN subjects, only age was not significantly different in the unweighted values (p = .055; all other comparisons ps < .01). Post-weighting, there were no longer differences in sex (p = .09) or Aβ ratio (p = .14) as well. For MCI subjects, comparisons using unweighted values showed no differences between clinical groups on age (p = .31) or Aβ ratio (p = .07), with all other comparisons significant (ps < .05). Post-weighting, no comparisons were significant (ps > .05). For AD subjects, age comparisons were not significant for unweighted values (p = .30; all other comparisons ps < .05). Post-weighting, education (p = .24) was no longer significant; however, age comparisons were now significant (p =.01). Overall, weighting reduced differences between clinical groups on multiple variables (see Model 2 balance tables for all comparisons). The relative influence of covariates on propensity scores was different across clinical groups. For CN subjects, Aβ ratio contributed ~45%, Aβ42 ~16%, t-tau ~11%, p-tau$_{181}$ ~10%, and age ~9%. For MCI subjects, t-tau contributed ~25%, Aβ42 ~22%, Aβ ratio ~16%, p-tau$_{181}$ ~16%, and age ~11%. For AD subjects, Aβ42 contributed ~33%, t-tau ~24%, Aβ ratio ~21%, p-tau$_{181}$ ~9%, and age ~9%.

As with Model 1, white matter PCA scores as a function of diagnosis were compared using a Wald test to account for the weighted survey design, with AD as the baseline level of predictor. Results are presented in Table 3. The overall model was significant, $F(2, 79) = 5.32$, p

= .007. The intercept, or the effect of AD, was significantly different from zero (p = .002) with an estimate of 1.30, SE = 0.41. The difference in PCA scores between CN and AD participants was significant where CN participants had estimated scores lower than AD by -1.47, SE = 0.45, p = .002. The difference between MCI and AD subjects was significant where MCI participants showed a reduced estimate compared to AD participants by -1.14, SE = 0.57, p = .05. These differences in brain scores reflect the effect of diagnoses after accounting for demographic variables as well as genetic factors and biomarkers.

Scores on the MMSE were investigated next. The overall model was significant, $F(2, 570) = 312.20$, $p < .001$. The estimate for AD participants was 23.2, SE = 0.23. CN participants had a higher estimate by a score of 5.9, SE = 0.25, which was significant, $p < .001$. MCI participants also had higher estimates compared to AD by a score of 4.0, SE = 0.28, $p < .001$. MMSE scores differed between clinical groups as expected after accounting for predictive covariates.

Model 3

The third model attempted to weigh language groups using propensity scores created from covariates including demographic variables, biomarkers, and genetic factors. The formula used in R was:

```
GBM3 <- ps(LANGUSE ~ SEX  + AGE + EDU + ETHCAT + RACCAT + ABETA42 + ABRATIO +
                     PTAU + TAU + APOE,
                     data = as.data.frame(biofull),
                     estimand = "ATE",
                     verbose = FALSE,
                     stop.method = c("es.mean"),
                     n.trees = 1000)
```

Bilinguals and monolinguals are likely classified as such due to different circumstances with unequal probabilities of receiving either “treatment”, so ATT was selected as the estimand. The original model using the ATT estimand failed to converge after 10000 iterations, so ATE was

selected as an alternative investigation. The current model converged using significantly fewer iterations, set here at 1000. However, despite model convergence, overlap in propensity scores was poor (Model 3, Plot 2). This suggests that language group differences are stark enough that even after propensity score estimation, groups are largely separate such that predictive covariates differ between groups. This provides a hint for why model convergence failed using ATT as the estimand in initial attempts. A plot of the ASMD pre- and post-weighting between language groups on covariates shows a general reduction in values post-weighting for comparisons with high unweighted differences. For comparisons with low unweighted differences, weighted values generally had a slight uptick (Model 3, Plot 3). Balance tables revealed no changes in significance between unweighted and weighted means, where only race was significantly different between groups ($p < .01$). The relative influence of covariates revealed that t-tau and p-tau$_{181}$ contributed almost the entirety of propensity score determination, with ~66% and ~20%, respectively.

White matter PCA scores as a function of language use and clinical diagnosis were compared using a Wald test for a weighted survey design, with results displayed in Table 4. The overall model was significant, $F(4, 96) = 19.66$, $p < .001$. A diagnosis of AD was set as the baseline factor for the diagnosis variable, and monolingual status was the baseline factor for language use; therefore, the intercept in the model was monolingual AD participants. Monolingual AD participants had an estimate significantly different from zero with a value of 1.12, SE = .46, p = .02. Monolingual CN participants had significantly lower estimates than monolingual AD participants by -1.52, SE = 0.50, $p = .003$, whereas monolingual MCI participants had lower estimates by -1.18, SE = 0.58, $p = .04$. Bilingual CN participants had significantly higher estimates than their monolingual counterparts by a value of 1.95, SE = .59, *p*

= .001. In contrast, bilingual MCI participants did not have significantly different PCA scores than their monolingual MCI counterparts, differing by a higher estimate of only .11, SE = .53, p = .84. Bilingual PCA scores were only available for 5 CN participants compared to 56 in the monolingual group, 14 MCI participants compared to 17 in the monolingual group, and 0 AD participants compared to 9 in the monolingual group. The implications and possible reasons for this are discussed later.

Cognitive measures were examined next. A weighted survey design analysis of MMSE scores was significant for the overall model, $F(5, 614) = 227.88$, $p < .001$. Monolingual AD participants had a score estimate of 23.1, SE = .22. Monolingual CN participants had significantly higher estimates by a score of 6.0, SE = .23, $p < .001$, as did monolingual MCI participants by a score of 4.0, SE = .29, $p < .001$. There was no observed effect of group: bilingual AD participants were not significantly different than their monolingual counterparts with a decreased estimate of -0.32, SE = 36, $p = .38$. Similarly, bilingual CN participants did not differ from their monolingual counterparts with an estimate of -0.10, SE = .45, $p = .83$, nor did bilingual MCI participants differ from monolingual MCI participants with a slightly lower estimate by -0.29, SE = .53, $p = .57$.

Scores from another measure that looked at diagnosis, ADAS-Cog-13, were also investigated. Higher scores reflect worse performance on the test, and therefore higher estimates are expected for MCI and AD participants compared to CN participants. An overall model examining scores on ADAS-Cog-13 was significant, $F(5, 592) = 135.77$, $p < .001$. Monolingual AD participants scored significantly higher than zero with an estimate of 36.2, SE = 1.1, $p < .001$. Compared to these monolingual AD participants, monolingual CN participants scored significantly better with reduced estimates of -24.7, SE = 1.2, $p < .001$, as did monolingual MCI

participants with reduced estimates by -12.3, SE = 1.4, $p < .001$. Bilingual AD participants did not significantly differ from monolingual AD participants with an estimated increase of 1.8, SE = 2.6, $p = .51$. Bilingual CN participants also did not differ from their monolingual CN counterparts with a lower estimate of -2.4 points, SE = 3.2, $p = .45$. Bilingual MCI participants, however, were significantly different from their monolingual MCI counterparts with lower scores by an estimate of -10.7 points, SE = 3.2, $p < .001$.

The results of these analyses show marked clinical group differences across DTI measures, MMSE scores, and ADAS-Cog-13 scores, but no reliable language group differences aside from PCA scores between CN participants and scores on the ADAS-Cog-13 between MCI participants. After including demographic, biomarker, and genetic factor information in the predictive model, t-tau and p-tau$_{181}$ contributed the greatest influence in predicting these outcomes.

Table 4. *Model 3 estimates for monolingual and bilingual comparisons.*

	Monolinguals			Bilinguals		
	CN	MCI	AD	CN	MCI	AD
PCA Estimates	-0.40 (.50); N = 56***	-0.062 (.58); N = 17	1.12 (.46); N = 9	1.56 (.59); N = 5***	0.047 (.53); N = 14	*NA*
MMSE Estimates	29.1 (.2); N = 262	27.1 (.3); N = 188	23.1 (.2); N = 123	29.0 (.4); N = 15	26.8 (.5); N = 28	22.8 (.4); N = 4
ADAS-Cog13 Estimates	11.5 (1.2); N = 260	23.9 (1.4); N = 179***	36.2 (1.1); N = 120	9.1 (3.2); N = 12	13.2 (3.2); N = 23***	38.0 (2.6); N = 4

Note: Variables are presented as mean score estimates with standard errors in parentheses. Comparisons are between language groups matched on diagnosis where * $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$

Discussion

The current study examined the effects of biomarkers and genetic factors on clinical and cognitive outcomes within the context of bilingualism as a measure of cognitive reserve.

However, difficulties in participant collection and data inclusion necessitated the use of novel exploratory analyses. Although the findings from this methodology can not yet be validated, the methods and results provide useful directions for future research.

Gradient-boosted modeling was used to create propensity scores of clinical diagnostic groups in monolinguals in two different models. The first model examined only demographic factors as predictors of clinical group inclusion. Across all three diagnoses (CN, MCI, and AD), age had the greatest influence on propensity score, ranging from 67% (in CN participants) to 77% (in MCI participants). Education had the next largest influence on propensity score creation, accounting for 13% (in AD participants) to 20% (in CN participants). Sex played only a small part in propensity score creation (~9% for CN and AD participants), with racial categories accounting for less still (<5%). Essentially, in older adults, age itself seems to be the greatest predictor of clinical diagnosis. Education also influences diagnoses, but to a lesser extent than age. These results should come as no surprise: models that attempt to predict risk of dementia in later life often highlight age and education as risk factors (Hou et al., 2019). In the current results, education plays a larger role in determining propensity scores for CN participants than for AD participants. This suggests the possibility of education acting as a mediator between age and cognitive decline, such that education helps ward off or delay the onset of dementia. When cognition is still intact, education plays a large role in maintaining this level of performance. Once the effects of dementia have taken over and decline has occurred, education plays a smaller role. Put another way, education's greatest benefit seems to occur prior to decline and shows diminishing returns after the onset of dementia.

Brain health measures evaluated using white matter PCA scores and MMSE scores for the clinical groups were within expectations. Cognitively normal participants had PCA scores

that reflected greater whole-brain FA values, AD participants had PCA scores that reflected greater DA and DR values, and MCI participants had PCA scores that reflected values somewhere in between these two groups. This pattern held before and after weighting to account for demographic variables, and is consistent with other studies examining white-matter integrity in aging and cognitive decline (e.g., Bennett et al., 2010; Bennett et al., 2017; Bosch et al., 2012; Burzynska et al., 2010). A similar pattern was seen for MMSE scores wherein CN participants had the highest scores, followed by MCI participants, and finally AD participants. As with PCA scores, this pattern held after weighting. Considering that MMSE scores are considered in the diagnostic process itself, concerns would be raised if the pattern showed otherwise. Still, the brain health and MMSE scores provide strong reassurance that the groups, as classified, fall into expected patterns: healthier brains and cognitive outcomes for CN participants compared to their MCI and AD peers.

The second model expanded on Model 1 by including biomarker and genetic factor information, again only in the monolingual subset. Propensity scores generated by this GBM analysis would account for both demographic factors and "hidden" factors that play a role in the development of MCI and AD. In this model, the GBM analysis had good convergence, but comparisons between clinical groups on predictive factors post-weighting showed some amount of variance. For example, CN participants showed differences on some predictive factors such as education and tau levels when compared to MCI and AD participants, and AD participants also showed differences in age and biomarker values when compared to CN and MCI participants. However, MCI participants showed no differences when compared to CN and AD participants on any predictive factors, including age, education, Aβ, or tau levels. These results show that regardless of attempts to match clinical groups perfectly on predictive factors of diagnosis, some

variables are inherently different between groups. MCI participants, who are somewhere along the path between normal cognition and dementia and can thus show profiles similar to either, do not show these differences.

Clinical group propensity scores were determined with varying levels of influence from each of the predictive variables. In Model 1, clinical group status was largely influenced by age followed by education. Including biomarkers in the GBM analysis told a different story. The results are reported in the previous section but restated here for clarity. For CN participants, Aβ42/40 ratio contributed to almost 50% of the propensity score determination, followed by Aβ42 levels (~16%), and then t-tau, p-tau$_{181}$, and age contributing ~10% each. Propensity scores for MCI were determined with lower but more evenly spread values of influence, where t-tau contributed ~25%, Aβ42 ~22%, Aβ ratio ~16%, p-tau$_{181}$ ~16%, and age ~11%. Propensity scores for AD showed greater contribution from Aβ42 (~33%) than other diagnostic groups, followed by t-tau ~24%, Aβ ratio ~21%, p-tau$_{181}$ ~9%, and age ~9%. The results and amount of influence from each variable vary slightly by group, but all show a similar pattern: biomarker levels, i.e., Aβ and tau, influence group classification more strongly than the strongest predictors from Model 1, i.e., age and education. Sex, ethnicity, or racial identification contributed almost no part to classification.

Most of these findings are predictable, but some are surprising. Studies often posit these biomarkers as reliable predictors of AD diagnosis and development (e.g., Cullen et al., 2020; Huang et al., 2020; Schmand et al., 2010), and the data from this study fall into expected patterns. Amyloid-β levels are highest in CN participants and decline in MCI and even more in AD participants. This is the expected pattern considering that, as Aβ plaques accumulate in the brain, less Aβ is found in CSF. Total tau and p-tau$_{181}$ levels also follow expected patterns, where

values for CN, MCI, and AD participants increase in order from lowest to highest. Given these patterns, it is expected that clinical diagnosis would be strongly influenced by a certain level of CSF biomarker. However, the relatively negligible influence that demographic variables have on diagnosis is surprising. When biomarkers are included as predictive factors of diagnosis, their influence outweighs the predictive value of age and education. It could be argued that biomarker levels influence clinician diagnostic decisions, i.e., knowing a participant's CSF biomarker levels plays a role in diagnosis. However, in the ADNI database, clinical diagnoses are determined prior to biomarker assay findings. The results therefore seem to indicate that biomarker levels are the greatest predictors, or indicators, of cognitive impairment, influencing clinical diagnosis more so than cognitive reserve factors such as education or demographic factors such as age or sex.

The absence of APOE's influence on clinical diagnosis was noted, particularly considering its strong predictive value on AD (Seshadri et al., 1995). However, given that APOE ε4 carriers show greater abundance of senile plaques compared to noncarriers, it is possible that the APOE variable was effectively "hidden" and accounted for in biomarker levels. Post hoc exploration of the data showed this to be a strong possibility, with regression models showing significant differences in Aβ42 levels between APOE ε4 carriers and non carriers, $F(1, 498) = 107.75$, $p < .001$, as well as differences in t-tau levels, $F(1, 498) = 20.25$, $p < .001$, and p-tau$_{181}$, $F(1, 498) = 27.16$, $p < .001$. Despite APOE's absence in the model on influencing diagnosis, the interpretation is that it was inherently accounted for in Aβ42 and tau levels.

White matter PCA scores and MMSE scores were again examined as in Model 1. Both measures followed similar patterns as those found in Model 1, i.e., better white matter health for CN participants compared to both MCI and AD participants, and higher scores on the MMSE for

CN and MCI participants compared to AD participants. This is expected, but the quality of results that were obtained in this model are theoretically more reliable than those obtained from Model 1. The results here have accounted for demographic variables as well as biomarker and genetic factor information through propensity score weighting, and support the notion that GBM analyses produce reliable results.

The final model determined propensity scores for language group status after accounting for demographic and biomarker factors. The initial attempt using ATT as the estimand failed to converge even after 10 000 iterations, strongly suggesting the presence of stark group differences. For one, participant numbers were drastically different, with 577 monolinguals but only 48 bilinguals included in the sample. GBM is supposed to be able to handle group size disparities, but the extremely low sampling of data for diagnostic groups within the overall bilingual sample leads to unreliable or missing information that can make it difficult to compare across groups. As one such example, Aβ42 and tau values were only available for 7 CN bilinguals and 2 bilinguals with AD compared to every CN and AD monolingual participant having this information. This vast disparity made it impossible to compare the groups using ATT. Using ATE as the estimand led to model convergence, but evaluation of the propensity score plot also revealed poor overlap between language groups; i.e., even after model convergence, language groups had largely different propensity scores with little overlap. Given the small sample size for bilingual participants, it is impossible to determine if this was a consequence of poor sampling or inherent differences between groups that would be present even with comparable group sizes.

The small sample size of bilingual participants in the ADNI database is unfortunate but not entirely unexpected. Although specific site locations are not revealed, the majority of

participants in the ADNI database are recruited from sites within the USA. Multiple studies have shown that recruitment to clinical trials in the USA – whether for cancer research, therapeutic drugs, or otherwise – largely favours inclusion of white participants over minority racial/ethnic groups (e.g., Baquet et al., 2006; Eshera et al., 2015; Sateren et al., 2002). Considering these findings and the criteria for determining monolingual status (non-Hispanic or Latino, English spoken at test and at home), the overall demographics and language group breakdown for the participants of the current study seem more reasonable. However, the rate at which bilingual participants were willing to undergo biomarker and genetic factor testing compared to monolinguals is an issue with implications to be discussed in the General Discussion.

Although the large language group sample size differences are a concern, results of Model 3 were investigated. Unlike Model 2 that showed a large influence of Aβ on propensity scores, tau and p-tau$_{181}$ influenced almost the entirety of language group classification, at ~66% and ~20%, respectively. There is no discernible reason this should be the case from a theoretical perspective, but from a practical perspective, it may be a consequence of poor sampling of the bilingual group that biases values. Brain health was compared between language groups after weighting by propensity scores. Monolingual participants again showed the same pattern of better brain health for CN participants compared to MCI and AD, but bilingual participants did not. In this sample, bilingual CN participants had brain health scores more similar to monolingual AD participants, while bilingual MCI participants did not significantly differ from their monolingual counterparts. As mentioned previously, bilingual CN participant brain scores were taken from only 5 participants compared to 14 participants in the bilingual MCI group. This low sample makes it difficult, if not impossible, to reliably interpret the brain scores of bilingual CN participants. For MCI participants, however, there were no differences in brain health

between bilinguals and monolinguals after accounting for demographic and biomarker data. This somewhat contrasts with previous studies showing white matter differences between language groups for cognitively healthy individuals (Anderson et al., 2018; Gold et al., 2013; Luk et al., 2011). The inclusion of demographic and biomarker variables in weighting participants and establishing better balance between participants may account for this disparity, but more data from CN and AD participants would be needed to confirm this proposition.

Cognitive performance was measured by MMSE scores as with previous models, and with results from the ADAS-Cog test. Scores on the MMSE followed the pattern seen in previous models, that is, CN participants performed better than MCI participants who, in turn, performed better than AD participants. This pattern was seen for both monolingual and bilingual participants, with no difference between language groups. Previous research that looked at MMSE performance between monolingual and bilingual language groups in older age and with dementia also showed no differences between language groups (e.g., Bialystok et al., 2007; Chertkow et al., 2010; Woumans et al., 2015). Scores on the ADAS-Cog were included in analyses to help dissociate any possible language group differences in cognition. In the USA, the ADAS-Cog and the MMSE are the two primary cognitive outcome measures in all Food and Drug Administration approved clinical studies of AD, with the ADAS-Cog test showing better sensitivity, reliability, and less dependence on education and language skills than the MMSE (Kaufman et al., 2017). There are, however, some concerns that the ADAS-Cog is less effective in detecting early MCI or early stages of dementia than in more severe stages (e.g., Cano et al., 2010; Lezak et al., 2004).

In contrast to the MMSE, higher scores on ADAS-Cog reflect poorer performance. Clinical group performance followed expected patterns – CN participants scored lower (and thus

performed better) than MCI participants, who, in turn, scored lower than AD participants. As with performance on the MMSE, there were no language group differences between monolingual and bilingual CN participants, nor any differences between monolingual and bilingual AD participants. Bilingual MCI participants, however, performed significantly better than their monolingual MCI counterparts such that they scored ~10 points lower. In the overall context, they even performed similarly to monolingual CN participants with a score of ~13 compared to ~12, respectively. If one considers that normal cognition is "pre-decline," and a diagnosis of dementia is the endpoint, or "post-decline," then a diagnosis of MCI should indicate some time point within the stage of "decline" itself. Better performance during this decline, as bilingual participants show, could be another indication of preserved cognitive functioning in the face of neural decline – or simply, cognitive reserve. A recent study comparing bilingual and monolingual MCI patients on global parenchymal atrophy of the brain and overall cognition found no cognitive differences between language groups despite worse atrophy for bilinguals (Costumero et al., 2020). Here, bilingual and monolingual MCI participants show no differences in white matter health, but better performance for bilinguals on at least one cognitive measure, the ADAS-Cog. More research is needed to disentangle the findings at specific time points of cognitive decline.

There are limitations to this study and the interpretations that can be made. As with Study 1, the ADNI database does not provide comprehensive language or culture information about its participants. Classification as "monolingual" or "bilingual" requires some assumptions based on ethnicity, race, and languages spoken at test and at home. Census data from the USA supports these assumptions to a reasonable degree (Rumbaut & Massey, 2013), but there is still a margin of error that cannot be accounted for without more detailed language information. The ADNI

database also provides scores for neuropsychological assessment tests, but only includes raw values. Considering that the tests are individually standardised, no accurate standard scores can be created without access to test-specific documents and standardisation tables. This is an unfortunate loss of potentially useful information to be added to analyses. Sampling rates of bilingual participants was also an issue, with significantly less information available for bilingual participants than for monolinguals. Gradient boosted modeling alleviates some of those concerns, as it is supposed to handle missingness and mismatched sample sizes to create propensity scores for weighting. Despite this, data were either missing entirely for some subsamples of participants (e.g., no DTI information for bilingual AD participants), or provided by too few participants to make meaningful comparisons (e.g., DTI information from only 5 bilingual CN participants, or APOE information from only 2 bilingual AD participants). Databases that introduce more extensive language screening and collect data from a more diverse sample of the population would alleviate some of these issues for future research.

The current study integrated biomarkers, genetic factors of AD, and bilingualism as predictors of brain health and cognitive performance in older aging and dementia. Limitations in the ADNI participant pool demographics and data availability made traditional comparisons between bilinguals and monolinguals difficult. Gradient boosted regression modelling was used to address some of these concerns. Model fit worked well in the monolingual subsample, producing well-balanced weighted data for comparisons between diagnostic groups. Weighting worked less well when comparing bilinguals and monolinguals, where language groups still had propensity scores that were largely distinct from each other. Despite this difference in model fit, biomarkers had the greatest influence across models such that CSF Aβ and tau levels determined clinical group or language group inclusion to a greater degree than other predictors such as

education or age. The results from this study suggest that GBM can be used effectively to better balance data prior to group comparisons, and that future studies examining cognitive decline should, when possible, consider biomarkers in the analyses and interpretations.

General Discussion

This dissertation investigated the factors involved in cognitive reserve in older adults through novel exploratory analyses and methods. The first study explored bilingualism as a factor of cognitive reserve by holding brain health constant between monolingual and bilingual older adults and subsequently comparing clinical outcomes. The results showed poorer clinical diagnostic outcomes for monolinguals than expected by chance from a random sample, strongly suggesting that bilingualism acts as a protective factor in aging. The second study used gradient boosted regression modelling to determine propensity scores for clinical group diagnoses in monolinguals and found that propensity scores were largely influenced by CSF biomarkers rather than demographic factors such as age or education. Results of GBM analyses to determine language group designation also strongly suggested a role of biomarkers, although data availability from bilingual participants were lacking. In general, propensity score weighting accounted for predictive factors and returned better balanced and higher quality results for analyses than what would be expected using the raw data alone. Taken together, the results from these studies contribute to the growing body of evidence supporting bilingualism as a factor of cognitive reserve, and offer new methods of investigating group differences to account for confounding variables and data disparities.

Stern's (2002; 2009) model of cognitive reserve (Figure 1) makes specific predictions that are supported by previous research when setting bilingualism as a proxy of high reserve (see Introduction), and now by the current study as well. The last prediction to be investigated in the context of bilingualism as a proxy of reserve was that bilinguals should show better cognitive and clinical outcomes than monolinguals when neuropathology is held constant. This was shown

in Study 1 by matching white matter health in cognitively normal bilinguals to a larger sample of clinically diverse monolinguals and finding that matched monolinguals showed worse clinical and cognitive outcomes than that predicted by chance. The previously cited literature and current research overwhelmingly support the position that bilingualism is another proxy of cognitive reserve and should be included in the discussion about healthy aging alongside other proxies such as education, physical activity, socioeconomic status, and occupational complexity.

There is, however, debate within the field of bilingualism regarding its role in cognitive outcomes that seems to pull attention away from this argument (Antoniou, 2019). Specifically, bilingualism has been posited to have positive effects on an individual's performance on tests of executive function – the set of cognitive processes that individuals employ to achieve goals in daily life, often by selecting between competing choices, planning behaviour, and ignoring irrelevant stimuli. Executive function is typically examined using lab-based tasks including, but not limited to, the Stroop (Stroop, 1935), Simon (Simon & Rudell, 1967), and Flanker tasks (Eriksen & Eriksen, 1974). For young adults in particular, better performance (i.e., faster reaction times and/or more accurate responses) for bilinguals than monolinguals on a variety of tasks purportedly examining executive function has been reported across a variety of studies from different groups (e.g., Bialystok et al., 2008; Bialystok et al., 2014; Chung-Fat-Yim et al., 2017; Costa et al., 2008; Costa et al., 2009; Hernández et al., 2013; Prior & MacWhinney, 2010; Salvatierra & Roselli, 2011; Singh & Mishra, 2012). Recently, however, many studies have challenged this assertion of a "bilingual advantage" by showing no performance differences between bilingual and monolingual young adults on tasks of executive function (Gathercole et al., 2014; Paap & Greenberg, 2013; von Bastian et al., 2016), or arguing that reported positive effects are a result of publication bias (de Bruin et al., 2015).

Currently, there is no consensus on this topic. Meta-analyses have revealed both positive effects of bilingualism on executive function tasks (Grundy, 2020; van den Noort, 2019; Ware et al., 2020) and weak or null effects between language groups (Donnelly et al., 2019; Lehtonen et al., 2018). These disparate conclusions between meta-analyses are explained by the authors of these and other studies through "differences": differences in statistical analyses (between both individual studies and meta-analyses); differences in how monolingualism and bilingualism are defined; differences in executive function tasks between studies or included studies between meta-analyses; differences between cultural contexts and group demographics. The "true" reason for any observed language group differences is difficult, if not impossible, to ascertain. This focus on executive function differences in young adults distracts from a related point; that is, observing (or not) executive function differences between language groups in young adults does not address the real-life benefits of bilingualism, particularly as it pertains to healthy aging.

Bilingualism reorganizes and recruits executive control brain networks (Bialystok et al., 2012), but it is not surprising that behavioural differences often do not appear in the young adult samples that are at the center of the current disagreement. For one, executive function ability peaks in young adulthood and declines with aging (Park et al., 2002), such that performance of these young undergraduate students would approach ceiling on insufficiently challenging tasks. When discussing reaction times – for example on a Flanker task – any observed group differences (when they occur) are often on the order of tens of milliseconds. This is an effect that has few, if any, correlates outside of a laboratory setting. Furthermore, the effect size of bilingualism on executive function performance is likely either in line with or more than other experience-related mediators such as physical activity (e.g., Chang et al., 2012, $g = 0.097$), or musical training (e.g., Sala & Gobet, 2020, Cohen's $d = 0.16$), but all show small or unreliable

effects regardless (Grundy, 2020). Despite these small effects in lab-based testing, there are intrinsic and extrinsic benefits (such as personal satisfaction and greater social opportunities) of physical activity and musical training, as there are for bilingualism.

Often lost in the debate over behavioural outcomes is the fact that neuroimaging studies reliably show language group differences in brain structure and functional activation even in the absence of behavioural differences (see Pliatsikas, 2020, for a review). Studies with young adults have shown that bilinguals recruit different brain regions and associated networks than monolinguals in the absence of differences in performance on nonverbal cognitive tasks (Garbin et al., 2010; Luk et al., 2010). For example, in the study by Garbin et al. (2010), on a nonverbal switch task monolinguals showed greater activation in the right inferior frontal gyrus than bilinguals, which the authors suggested was evidence for suppression of task inertia. In contrast, bilinguals had greater activation in the left inferior frontal gyrus, which the authors associated with inhibition of incorrect responses. Interpreting the findings in this manner leads to the issue of reverse inference (i.e., inferring a specific cognitive process based on brain activation), but the primary takeaway is that in the absence of behavioural differences, there is a difference in quality of processing. Sceptics of the theory of a bilingual advantage accept that there are neuroanatomical and neural processing differences between the language groups, but qualify this acceptance with the assertion that failure to align with performance means only evidence of brain plasticity but not an advantage (Paap et al., 2016). In this context, however, ‘performance’ and ‘advantage’ refer to outcomes on executive function tasks and not clinical or other cognitive outcomes.

Disregarding neural differences between language groups due to a lack of consensus on behavioural measures when executive function performance is at its peak ignores the reliable

neural and behavioural effects seen when cognition is in decline, i.e., in older age. Studies examining executive function performance in older adults often find language group differences even in the absence of effects in younger adults (e.g., Bialystok et al., 2004; Bialystok et al., 2008; Gold, Kim, et al., 2013; Salvatierra & Rosselli, 2010). There are, naturally, some studies with findings that call into question a bilingual advantage in older age as well. A meta-analysis by Mukadam et al. (2017) attempted to show that there was no protective effect of bilingualism on dementia outcomes in older adults, yet there were some issues such as excluding all retrospective studies from the analyses, conflating the difference between prospective and retrospective outcomes, and inconsistency in their criteria of study quality (see Grundy & Anderson, 2017, for a rebuttal). Sceptics of a bilingual advantage also point to prospective studies that show no difference in incidence rates of dementia between monolinguals and bilinguals (Lawton et al., 2015; Sanders et al., 2012; Yeung et al., 2014; Zahodne et al., 2014). However, cognitive reserve theory only predicts cognitive compensation in the face of neuropathology and a delay in symptom onset, not that decline disappears entirely. Thus far, there has been no alternative explanation or rebuttal for the results that have been shown in retrospective studies comparing dementia onset between language groups (e.g., Alladi et al., 2013; Bialystok et al., 2007), nor for neuroimaging studies showing maintained behavioural performance in bilinguals in the face of increased neurodegeneration (e.g., Perani et al., 2017; Schweizer et al., 2012).

Study 2 provides some avenues for further research in this field. Propensity score matching is becoming more common as a means of matching participants on background measures (used in Study 1), but at the expense of losing participants if there is not a perfect match between groups. Instead, gradient boosted regression modelling creates propensity scores

that can then be used to weight participant data, allowing greater retention of participant information for analyses. One criticism levied against studies showing effects of bilingualism on executive function performance is that Type I errors may occur more commonly due to small sample sizes (Paap & Greenberg, 2013). Propensity score weighting in part addresses this issue by ensuring more participants are included. Theoretically, the use of multi-site databases such as ADNI further alleviates this concern by granting access to hundreds, if not thousands, of participants depending upon the question under investigation.

Databases are unfortunately not without their faults. Just as longitudinal clinical research trials deal with attrition of participants, so do databases – although for ADNI specifically, the rate is significantly lower at ~7% per year, compared to a typical 20-30% for a research trial (Hua et al., 2010). Furthermore, even if standard protocols are followed, multi-site neuroimaging may require the use of data harmonization techniques to ensure accurate analyses across images, especially if images are obtained from a variety of scanner models (Fortin et al., 2017). Despite the wide range of data collected – including background information, genetic information, and cognitive test results – not all the information is useful, usable, or complete. In ADNI specifically, only raw scores are reported for some tests rather than scores standardised to age or education. For the current studies, information about language of participants was minimal and assumptions had to be made regarding language group inclusion. Immigration status, often pointed to as a confounding factor in bilingualism and aging research (Fuller-Thomson & Kuh, 2014), is absent entirely. This lack of information and detail regarding language in the databases unfortunately limits the conclusions that may be drawn from research using these participants.

As mentioned in Study 2, there is a startling lack of participant diversity in research and clinical trials compared to population demographics. One prominent refrain is that ethnically and

racially marginalized groups often have a lack of trust in the medical research community, leading to lower participation in studies (e.g., Ejiogu et al., 2011; Kraft et al., 2018). Some groups have taken steps to explore these issues and to implement recruitment strategies addressing these concerns (e.g., Clark et al., 2019), but when progress is slow, it has substantial implications for data availability and the conclusions that may be drawn. Such was the case for Study 2, wherein only 64 bilinguals were identified after participant review. Of these 64 bilinguals only 48 had clinical diagnoses; of these 48, 19 had diffusion-tensor images; 16 had biomarker data; and only 12 had data about APOE genotyping. Analyses showed a small influence of one cognitive reserve proxy, education, on clinical group diagnoses in monolinguals that should theoretically also be shown with a different proxy (e.g., bilingualism), but this could not be adequately tested because of these issues with participant recruitment and data availability. Considering the benefits of bilingualism in older age, future large-scale research and the conclusions that can be drawn would be best served by taking steps to ensure participant diversity and its associated factors. This involves reducing barriers to entry by, for example, building trust between participants and researchers/health-care providers, providing clear communication and information, and raising awareness of the benefits of participation (see Clark et al., 2019, and U.S. Food and Drug Administration, 2020, for specific recommendations).

The structural and functional brain changes that occur as a result of bilingual language experience lead to benefits beyond delayed dementia onset and (sometimes) faster reaction times in lab-based testing. Bilingualism has also been associated with greater cultural tolerance and open-mindedness (Dewaele & Stavans, 2012), perspective taking in children (Schroeder, 2018), and creative and divergent thinking (Kharkhurin, 2009). Couple these findings with better economic and employment opportunities for those who know multiple languages (Callahan &

Gándara, 2014) and the picture becomes clear: bilingualism as a lifestyle factor benefits individuals to lengths that extend well beyond the current disagreement on executive function tasks. Future educational, economic, social, medical, and research policies would profit from including bilingualism as a factor of interest, with a focus on diversity of participants, novel exploration, and reliable statistical analyses to guide the way.

REFERENCES

Abutalebi, J., Canini, M., Della Rosa, P. A., Green, D. W., & Weekes, B. S. (2015). The neuroprotective effects of bilingualism upon the inferior parietal lobule: A structural neuroimaging study in aging Chinese bilinguals. *Journal of Neurolinguistics*, *33*, 3–13. doi:10.1016/j.jneuroling.2014.09.008

Abutalebi, J., Canini, M., Della Rosa, P. A., Sheung, L. P., Green, D. W., & Weekes, B. S. (2014). Bilingualism protects anterior temporal lobe integrity in aging. *Neurobiology of Aging*, *35*(9), 2126–2133. doi:10.1016/j.neurobiolaging.2014.03.010

Abutalebi, J., & Green, D. W. (2008). Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and Cognitive Processes*, *23*(4), 557–582. doi:10.1080/016909608019206

Acosta-Cabronero, J., Alley, S., Williams, G. B., Pengas, G., & Nestor, P. J. (2012). Diffusion tensor metrics as biomarkers in Alzheimer's disease. *PLoS One*, *7*(11), 1–14. doi:10.1371/journal.pone.0049072

Adams, R. D., Victor, M., Ropper, A. H., & Daroff, R. B. (1997). *Principles of neurology* (6th ed.). McGraw-Hill.

Alladi, S., Bak, T. H., Duggirala, V., Surampudi, B., Shailaja, M., Shukla, A. K., Chaudhuri, J. R., & Kaul, S. (2013). Bilingualism delays age at onset of dementia, independent of education and immigration status. Neurology, 81(22), 1938–1944. doi:10.1212/01.wnl.0000436620.33155.a4

Almeida, R. P., Schultz, S. A., Austin, B. P., Boots, E. A., Dowling, N. M., Gleason, C. E., Bendlin, B. B., Sager, M. A., Hermann, B. P., Zetterberg, H., Carlsson, C. M., Johnson, S. C., Asthana, S., & Okonkwo, O. C. (2015). Effect of cognitive reserve on age-related

changes in cerebrospinal fluid biomarkers of Alzheimer Disease. *JAMA Neurology*, *72*(6), 699–706. doi:10.1001/jamaneurol.2015.0098

Alzheimer's Association. (2015). Changing the trajectory of Alzheimer's disease: How a treatment by 2025 saves lives and dollars. *Chicago: Alzheimer's Association*.

Alzheimer's Association (2020). 2020 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *16*, 391–460. doi:10.1002/alz.12068

Anderson, J.A.E., Chung-Fat-Yim, A., Bellana, B., Luk, G., & Bialystok, E. (2018). Language and cognitive control networks in bilinguals and monolinguals. *Neuropsychologia*, *117*, 352–363. doi:10.1016/j.neuropsychologia.2018.06.023

Anderson, J. A. E., Grundy, J. G., Frutos, J. D., Barker, R. M., Grady, C., & Bialystok, E. (2018). Effects of bilingualism on white matter integrity in older adults. *NeuroImage*, *167*, 143–150. doi:10.1016/j.neuroimage.2017.11.038

Anderson, J. A. E., Hawrylewicz, K., & Grundy, J. G. (2020). Does bilingualism protect against dementia? A meta-analysis. *Psychonomic Bulletin and Review*, *27*(5), 952–965. doi:10.3758/s13423-020-01736-5

Anderson, J. A. E., Mak, L., Chahi, A. K., & Bialystok, E. (2018). The Language and Social Background Questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, *50*(1), 250–263. doi:10.3758/s13428-017-0867-9

Antoniou, M. (2019). The advantages of bilingualism debate. *Annual Review of Linguistics*, *5*, 395–415. doi:10.1146/annurev-linguistics-011718-011820

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*, 399–424. doi:10.1080/00273171.2011.568786

Baquet, C. R., Commiskey, P., Daniel Mullins, C., & Mishra, S. I. (2006). Recruitment and participation in clinical trials: socio-demographic, rural/urban, and health care access predictors. *Cancer Detection and Prevention*, *30*(1), 24–33. doi:10.1016/j.cdp.2005.12.001

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

Baumgart, M., Snyder, H. M., Carrillo, M. C., Fazio, S., Kim, H., & Johns, H. (2015). Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *11*(6), 718–726. doi:10.1016/j.jalz.2015.05.016

Becker, R. E., Greig, N. H., & Giacobini, E. (2008). Why do so many drugs for Alzheimer's disease fail in development? Time for new methods and new practices?. *Journal of Alzheimer's Disease*, *15*(2), 303–325. doi:10.3233/jad-2008-15213

Bennett, I. J., Greenia, D. E., Maillard, P., Sajjadi, S. A., DeCarli, C., Corrada, M. M., & Kawas, C. H. (2017). Age-related white matter integrity differences in oldest-old without dementia. *Neurobiology of Aging*, *56*, 108–114. doi:10.1016/j.neurobiolaging.2017.04.013

Bennett, I. J., & Madden, D. J. (2014). Disconnected aging: Cerebral white matter integrity and age-related differences in cognition. *Neuroscience*, *276*, 187–205. doi:10.1016/j.neuroscience.2013.11.026

Bennett, I. J., Madden, D. J., Vaidya, C. J., Howard, D. V., & Howard, J. H. (2010). Age-related differences in multiple measures of white matter integrity: A diffusion tensor imaging study of healthy aging. *Human Brain Mapping*, *31*(3), 378–390. doi:10.1002/hbm.20872

Berkes, M., Bialystok, E., Craik, F. I. M., Troyer, A., & Freedman, M. (2020). Conversion of mild cognitive impairment to Alzheimer disease in monolingual and bilingual patients. *Alzheimer Disease and Associated Disorders*. doi:10.1097/WAD.0000000000000373

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, *143*(3), 233–262. doi:10.1037/bul0000099

Bialystok, E. (2021). Bilingualism: Pathway to cognitive reserve. *Trends in Cognitive Sciences, 25*(5), 355–364. doi:10.1016/j.tics.2021.02.003

Bialystok, E., Abutalebi, J., Bak, T. H., Burke, D. M., & Kroll, J. F. (2016). Aging in two languages: Implications for public health. *Ageing Research Reviews, 27*, 56–60. doi:10.1016/j.arr.2016.03.003

Bialystok, E., Anderson, J.A.E., & Grundy, J.G. (in press, 2018 online). Interpreting cognitive decline in the face of cognitive reserve: Does bilingualism affect cognitive aging? *Linguistic Approaches to Bilingualism*. doi:10.1075/lab.18040.bia

Bialystok, E., Craik, F. I. M., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, *45*(2), 459–464. doi:10.1016/j.neuropsychologia.2006.10.009

Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, *19*(2), 290–303. doi:10.1037/0882-7974.19.2.290

Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 859–873. doi:10.1037/0278-7393.34.4.859

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–250. doi:10.1016/j.tics.2012.03.001

Bialystok, E., Poarch, G., Luo, L., & Craik, F. I. M. (2014). Effects of bilingualism and aging on executive function and working memory. *Psychology and Aging*, *29*, 696–705. doi:10.1037/a0037254

Bittner, T., Zetterberg, H., Teunissen C. E., Ostlund Jr., R. E., Militello, M., Andreasson, A., Hubeek, I., Gibson, D., Chu, D. C., Eichenlaub, U., Heiss, P., Kobold, U., Leinenbach, A., Madin, K., Manuilova, E., Rabe, C., & Blennow, K. (2016). Technical performance of a novel, fully automated electrochemiluminescence immunoassay for the quantitation of β-amyloid (1-42) in human cerebrospinal fluid. *Alzheimer's & Dementia*, *12*(5), 517–526. doi:10.1016/j.jalz.2015.09.009

Blennow, K., Chen, C., Cicognola, C., Wildsmith, K. R., Manser, P. T., Bohorquez, S. M. S., Zhang, Z., Xie, B., Peng, J., Hansson, O., Kvartsberg, H., Portelius, E., Zetterberg, H., Lashley, T., Brinkmalm, G., Kerchner, G. A., Weimer, R. M., Ye, K., & Höglund, K. (2020). Cerebrospinal fluid tau fragment correlates with tau PET: A candidate biomarker for tangle pathology. *Brain*, *143*(2), 650–660. doi:10.1093/brain/awz346

Blennow, K., Hampel, H., Weiner, M., & Zetterberg, H. (2010). Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nature Reviews Neurology*, *6*(3), 131–144. doi:10.1038/nrneurol.2010.4

Bosch, B., Arenaza-Urquijo, E. M., Rami, L., Sala-Llonch, R., Junqué, C., Solé-Padullés, C., Peña-Gómez, C., Bargalló, N., Molinuevo, J. L., & Bartrés-Faz, D. (2012). Multiple DTI index analysis in normal aging, amnestic MCI and AD. Relationship with

neuropsychological performance. *Neurobiology of Aging*, *33*(1), 61–74. doi:10.1016/j.neurobiolaging.2010.02.004

Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica*, *82*(4), 239–259. doi:10.1007/BF00308809

Brier, M.R., Gordon, B., Friedrichsen, K., McCarthy, J., Stern, A., Christensen, J., Owen, C., Aldea, P., Su, Y., & Hassenstab, J. (2016). Tau and Aβ imaging, CSF measures, and cognition in Alzheimer's disease. *Science Translational Medicine*, *8*(338), 338ra366. doi:10.1126/scitranslmed.aaf2362

Brini, S., Sohrabi, H. R., Hebert, J. J., Forrest, M. R. L., Laine, M., Hämäläinen, H., Karrasch, M., Peiffer, J. J., Martins, R. N., & Fairchild, T. J. (2020). Bilingualism is associated with a delayed onset of dementia but not with a lower risk of developing it: A systematic review with meta-analyses. *Neuropsychology Review*, *30*, 1–24. doi:10.1007/s11065-020-09435-7

Buerger, K., Ewers, M., Pirttila, T., Zinkowski, R., Alafuzoff, I., Teipel, S. J., DeBernardis, J., Kerkman, D., McCulloch, C., Soininen, H., & Hampel, H. (2006). CSF phosphorylated tau protein correlates with neocortical neurofibrillary pathology in Alzheimer's disease. *Brain*, *129*(11), 3035–3041. doi:10.1093/brain/awl269

Burzynska, A. Z., Preuschhof, C., Bäckman, L., Nyberg, L., Li, S.-C., Lindenberger, U., & Heekeren, H. R. (2010). Age-related differences in white matter microstructure: Region-specific patterns of diffusivity. *NeuroImage*, *49*(3), 2104–2112. doi:10.1016/j.neuroimage.2009.09.041

Cabeza, R., Albert, M., Belleville, S., Craik, F., Duarte, A., Grady, C., Lindenberger, U., Nyberg, L., Park, D., Reuter-Lorenz, P. A., Rugg, M. D., Steffener, J., & Rajah, M. N. (2018).

Cognitive neuroscience of healthy aging: Maintenance, reserve, and compensation. *Nature Reviews Neuroscience*, *19*(11), 701-710. doi:10.1038/s41583-018-0068-2

Callahan, R. M. & Gándara, P. C. (Eds.). (2014) *The bilingual advantage: Language, literacy, and the US labor market*. Channel View Publications.

Cano, S. J., Posner, H. B., Moline, M. L., Hurt, S. W., Swartz, J., Hsu, T., & Hobart, J. C. (2010). The ADAS-cog in Alzheimer's disease clinical trials: Psychometric evaluation of the sum and its parts. *Journal of Neurology, Neurosurgery, and Psychiatry, 81*(12), 1363–1368. doi:10.1136/jnnp.2009.204008

Caselli, R. J., Dueck, A. C., Locke, D. E., Hoffman-Snyder, C. R., Woodruff, B. K., Rapcsak, S. Z., & Reiman, E. M. (2011). Longitudinal modeling of frontal cognition in APOE ε4 homozygotes, heterozygotes, and noncarriers. *Neurology*, *76*(16), 1383–1388. doi:10.1212/WNL.0b013e3182167147

Caselli, R. J., Reiman, E. M., Locke, D. E., Hutton, M. L., Hentz, J. G., Hoffman-Snyder, C., Woodruff, B. K., Alexander, G. E., & Osborne, D. (2007). Cognitive domain decline in healthy apolipoprotein E epsilon4 homozygotes before the diagnosis of mild cognitive impairment. *Archives of Neurology*, *64*(9), 1306–1311. doi:/10.1001/archneur.64.9.1306

Caselli, R. J., Reiman, E. M., Osborne, D., Hentz, J. G., Baxter, L. C., Hernandez, J. L., & Alexander, G. G. (2004). Longitudinal changes in cognition and behavior in asymptomatic carriers of the APOE e4 allele. *Neurology*, *62*(11), 1990–1995. doi:10.1212/01.wnl.0000129533.26544.bf

Chang, Y. K., Labban, J. D., Gapin, J. I., & Etnier, J. L. (2012). The effects of acute exercise on cognitive performance: A meta-analysis. *Brain Research*, *1453*, 87–101. doi:10.1016/j.brainres.2012.02.068

Chertkow, H., Whitehead, V., Phillips, N., Wolfson, C., Atherton, J., & Bergman, H. (2010). Multilingualism (but not always bilingualism) delays the onset of Alzheimer disease: Evidence from a bilingual community. *Alzheimer Disease & Associated Disorders*, *24*, 118–125. doi:10.1097/WAD.0b013e3181ca1221

Chouinard, J. (2000). Dysphagia in Alzheimer disease: A review. *The Journal of Nutrition, Health & Aging*, *4*(4), 214–217.

Chung-Fat-Yim, A., Sorge, G. B., & Bialystok, E. (2017). The relationship between bilingualism and selective attention in young adults: Evidence from an ambiguous figures task. *The Quarterly Journal of Experimental Psychology*, *70*, 366–372. doi:10.1080/17470218.2016.1221435

Clark, L. T., Watkins, L., Piña, I. L., Elmer, M., Akinboboye, O., Gorham, M., Jamerson, B., McCullough, C., Pierre, C., Polis, A. B., Puckrein, G., & Regnante, J. M. (2019). Increasing diversity in clinical trials: Overcoming critical barriers. *Current Problems in Cardiology*, *44*(5), 148–172. doi:10.1016/j.cpcardiol.2018.11.002

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, *113*, 135–149. doi:10.1016/j.cognition.2009.08.001

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, *106*, 59–86. doi:10.1016/j.cognition.2006.12.013

Costumero, V., Marin-Marin, L., Calabria, M., Belloch, V., Escudero, J., Baquero, M., Hernandez, M., Ruiz de Miras, J., Costa, A., Parcet, M.-A., & Ávila, C. (2020). A cross-sectional and longitudinal study on the protective effect of bilingualism against dementia

using brain atrophy and cognitive measures. *Alzheimer's Research & Therapy*, *12*(1), 11. doi:10.1186/s13195-020-0581-1

Cox, S. R., Ritchie, S. J., Tucker-Drob, E. M., Liewald, D. C., Hagenaars, S. P., Davies, G., Wardlaw, J. M., Gale, C. R., Bastin, M. E., & Deary, I. J. (2016). Ageing and brain white matter structure in 3,513 UK Biobank participants. *Nature Communications*, *7*(1), 1–13. doi:10.1038/ncomms13629

Crane, P. K., Gruhl, J. C., Erosheva, E. A., Gibbons, L. E., McCurry, S. M., Rhoads, K., Nguyen, V., Arani, K., Masaki, K., & White, L. (2010). Use of spoken and written Japanese did not protect Japanese-American men from cognitive decline in late life. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *65*(6), 654–666. doi:10.1093/geronb/gbq046

Cullen, N. C., Leuzy, A., Palmqvist, S., Janelidze, S., Stomrud, E., Pesini, P., Sarasa, L., Allué, J. A., Proctor, N. K., Zetterberg, H., Dage, J. L., Blennow, K., Mattsson-Carlgren, N., & Hansson, O. (2020). Individualized prognosis of cognitive decline and dementia in mild cognitive impairment based on plasma biomarker combinations. *Nature Aging*. doi:10.1038/s43587-020-00003-5

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias?. *Psychological Science*, *26*(1), 99–107. doi:10.1177/0956797614557866

Deary, I. J., Whiteman, M. C., Pattie, A., Starr, J. M., Hayward, C., Wright, A. F., Carothers, A., & Whalley, L. J. (2002). Cognitive change and the APOE epsilon 4 allele. *Nature*, *418*(6901), 932. doi:10.1038/418932a

Del Maschio, N., Fedeli, D., & Abutalebi, J. (2020). Bilingualism and aging: Why research should continue. *Linguistic Approaches to Bilingualism*, *10*(1), 1–15. doi:10.1075/lab.18032.del

Dewaele, J.-M., & Stavans, A. (2014). The effect of immigration, acculturation and multicompetence on personality profiles of Israeli multilinguals. *International Journal of Bilingualism*, *18*(3), 203–221. doi:10.1177/1367006912439941

Dick, A. S., Garcia, N. L., Pruden, S. M., Thompson, W. K., Hawes, S. W., Sutherland, M. T., Riedel, M. C., Laird, A. R., & Gonzalez, R. (2019). No evidence for a bilingual executive function advantage in the ABCD study. *Nature Human Behaviour*, *3*(7), 692–701. doi:10.1038/s41562-019-0609-3

Dijkstra, T. (2005). Bilingual word recognition and lexical access. In J. F. Kroll & A. M. B. De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 179–201). New York: Oxford University Press.

Donnelly, S., Brooks, P. J., & Homer, B. D. (2019). Is there a bilingual advantage on interference-control tasks? A multiverse meta-analysis of global reaction time and interference cost. *Psychonomic Bulletin & Review*, *26*(4), 1122–1147. doi:10.3758/s13423-019-01567-z

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited: Myth or reality? *Experimental psychology*, *61*(3), 234–251. doi:10.1027/1618-3169/a000243

Duncan, H. D., Nikelski, J., Pilon, R., Steffener, J., Chertkow, H., & Phillips, N. A. (2018). Structural brain differences between monolingual and multilingual patients with mild

cognitive impairment and Alzheimer disease: Evidence for cognitive reserve. *Neuropsychologia*, *109*, 270–282. doi:10.1016/j.neuropsychologia.2017.12.036

Ejiogu, N., Norbeck, J. H., Mason, M. A., Cromwell, B. C., Zonderman, A. B., & Evans, M. K. (2011). Recruitment and retention strategies for minority or poor clinical research participants: Lessons from the Healthy Aging in Neighborhoods of Diversity across the Life Span study. *The Gerontologist*, *51 Suppl 1*(Suppl 1), S33–S45. doi:10.1093/geront/gnr027

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. doi:10.3758/BF03203267

Eshera, N., Itana, H., Zhang, L., Soon, G., & Fadiran, E. O. (2015). Demographics of clinical trials participants in pivotal clinical trials for new molecular entity drugs and biologics approved by FDA from 2010 to 2012. *American Journal of Therapeutics*, *22*(6), 435–455. doi:10.1097/MJT.0000000000000177

Farrer, L. A., Cupples, L. A., Haines, J. L., Hyman, B., Kukull, W. A., Mayeux, R., Myers, R. H., Pericak-Vance, M. A., Risch, N., & van Duijn, C. M. (1997). Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. *JAMA*, *278*(16), 1349–1356. doi:10.1001/jama.1997.03550160069041

Ferrari, C., Xu, W. L., Wang, H. X., Winblad, B., Sorbi, S., Qiu, C., & Fratiglioni, L. (2013). How can elderly apolipoprotein E ε4 carriers remain free from dementia?. *Neurobiology of Aging*, *34*(1), 13–21. doi:10.1016/j.neurobiolaging.2012.03.003

Fortin, J.-P., Parker, D., Tunç, B., Watanabe, T., Elliott, M. A., Ruparel, K., Roalf, D. R., Satterthwaite, T. D., Gur, R. C., Gur, R. E., Schultz, R. T., Verma, R., & Shinohara, R. T.

(2017). Harmonization of multi-site diffusion tensor imaging data. *NeuroImage*, *161*, 149–170. doi:10.1016/j.neuroimage.2017.08.047

Freund, Y., & Schapire, R. E. (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*(1), 119–139. doi:10.1006/jcss.1997.1504

Friederici, A. D. (2011) The brain basis of language processing: From structure to function. *Physiological Reviews*, *91*, 1357–1392. doi:10.1152/physrev.00006.2011

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, *29*(5), 1189–1232.

Fuller-Thomson, E., & Kuh, D. (2014). The healthy migrant effect may confound the link between bilingualism and delayed onset of Alzheimer's disease. *Cortex*, *52*, 128–130. doi:10.1016/j.cortex.2013.08.009

Garbin, G., Sanjuan, A., Forn, C., Bustamante, J. C., Rodriguez-Pujadas, A., Belloch, V., Hernandez, M., Costa, A., & Ávila, C. (2010). Bridging language and attention: Brain basis of the impact of bilingualism on cognitive control. *NeuroImage*, *53*(4), 1272–1278. doi:10.1016/j.neuroimage.2010.05.078

Gathercole, V. C., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Viñas Guasch, N., Roberts, E. J., Hughes, E. K., & Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in Psychology*, *5*, 11. doi:10.3389/fpsyg.2014.00011

Gold, B. T., Johnson, N. F., & Powell, D. K. (2013). Lifelong bilingualism contributes to cognitive reserve against white matter integrity declines in aging. *Neuropsychologia*, *51*(13), 2841–2846. doi:10.1016/j.neuropsychologia.2013.09.037

Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., & Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(2), 387–396. doi:10.1523/JNEUROSCI.3837-12.2013

Grundy, J. G. (2020) The effects of bilingualism on executive functions: An updated quantitative analysis. *Journal of Cultural Cognitive Science*, *4*, 177–199. doi:10.1007/s41809-020-00062-5

Grundy, J. G., & Anderson, J. A. E. (2017). Commentary: The relationship of bilingualism compared to monolingualism to the risk of cognitive decline or dementia: A systematic review and meta-analysis. *Frontiers in aging neuroscience*, *9*, 344. doi:10.3389/fnagi.2017.00344

Hack, E. E., Dubin, J. A., Fernandes, M. A., Costa, S. M., & Tyas, S. L. (2019). Multilingualism and dementia risk: Longitudinal analysis of the Nun Study. *Journal of Alzheimer's Disease : JAD*, *71*(1), 201–212. doi:10.3233/JAD-181302

Han, L., Cole, M., Bellavance, F., McCusker, J., & Primeau, F. (2000). Tracking cognitive decline in Alzheimer's disease using the mini-mental state examination: A meta-analysis. *International Psychogeriatrics*, *12*(2), 231–47. doi:10.1017/s1041610200006359

Hansson, O., Lehmann, S., Otto, M., Zetterberg, H., & Lewczuk, P. (2019). Advantages and disadvantages of the use of the CSF Amyloid β (Aβ) 42/40 ratio in the diagnosis of

Alzheimer's Disease. *Alzheimer's Research & Therapy*, *11*(34), 1–15. doi:10.1186/s13195-019-0485-0

Hensley, K., Carney, J. M., Mattson, M. P., Aksenova, M., Harris, M., Wu, J. F., Floyd, R. A., & Butterfield, D. A. (1994). A model for beta-amyloid aggregation and neurotoxicity based on free radical generation by the peptide: Relevance to Alzheimer disease. *PNAS*, *91*(8), 3270–3274. doi:10.1073/pnas.91.8.3270

Hernández, M., Martin, C. D., Barceló, F., & Costa, A. (2013). Where is the bilingual advantage in task-switching? *Journal of Memory and Language*, *69*, 257–276. doi:10.1016/j.jml.2013.06.004

Ho, D., Imai, K., King, G., & Stuart, E. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*(3), 199–236. http://gking.harvard.edu/files/abs/matchp-abs.shtml

Horner, J., Alberts, M. J., Dawson, D. V., & Cook, G. M. (1994). Swallowing in Alzheimer's disease. *Alzheimer Disease and Associated Disorders*, *8*(3), 177–189. doi:10.1097/00002093-199408030-00004

Hou, X.-H., Feng, L., Zhang, C., Cao, X.-P., Tan, L., & Yu, J.-T. (2019). Models for predicting risk of dementia: A systematic review. *Journal of Neurology, Neurosurgery & Psychiatry*, *90*(4), 373–379. doi:10.1136/jnnp-2018-318212

Hua, X., Lee, S., Hibar, D. P., Yanovsky, I., Leow, A. D., Toga, A. W., Jack, C. R., Jr, Bernstein, M. A., Reiman, E. M., Harvey, D. J., Kornak, J., Schuff, N., Alexander, G. E., Weiner, M. W., Thompson, P. M., & Alzheimer's Disease Neuroimaging Initiative (2010). Mapping Alzheimer's disease progression in 1309 MRI scans: Power estimates for

different inter-scan intervals. *NeuroImage*, *51*(1), 63–75. doi:10.1016/j.neuroimage.2010.01.104

Huang, K., Lin, Y., Yang, L., Wang, Y., Cai, S., Pang, L., Wu, X., & Huang, L. (2020). A multipredictor model to predict the conversion of mild cognitive impairment to Alzheimer's disease by using a predictive nomogram. *Neuropsychopharmacology*, *45*(2), 358–366. doi:10.1038/s41386-019-0551-0

Izaks, G. J., Gansevoort, R. T., van der Knaap, A. M., Navis, G., Dullaart, R. P., & Slaets, J. P. (2011). The association of APOE genotype with cognitive function in persons aged 35 years or older. *PloS One*, *6*(11), e27415. doi:10.1371/journal.pone.0027415

Kametani, F., & Hasegawa, M. (2018). Reconsideration of amyloid hypothesis and tau hypothesis in Alzheimer's Disease. *Frontiers in Neuroscience*, *12*, 25. doi:10.3389/fnins.2018.00025

Kaufman, D. M., Geyer, H. L., & Milstein, M. J. (2017). Dementia. In D. M. Kaufman, H. L. Geyer, & M. J. Milstein (Eds.), *Kaufman's Clinical Neurology for Psychiatrists* (8th ed., pp. 105–149). Elsevier. doi:10.1016/B978-0-323-41559-0.00007-1

Kharkhurin, A.V. (2009). The role of bilingualism in creative performance on divergent thinking and invented alien creatures tests. *The Journal of Creative Behaviour*, *43*, 59–71. doi:10.1002/j.2162-6057.2009.tb01306.x

Kok, E., Haikonen, S., Luoto, T., Huhtala, H., Goebeler, S., Haapasalo, H., & Karhunen, P. J. (2009). Apolipoprotein E-dependent accumulation of Alzheimer disease-related lesions begins in middle age. *Annals of Neurology*, *65*(6), 650–657. doi:10.1002/ana.21696

Kraft, S. A., Cho, M. K., Gillespie, K., Halley, M., Varsava, N., Ormond, K. E., Luft, H. S., Wilfond, B. S., & Soo-Jin Lee, S. (2018). Beyond consent: Building trusting relationships

with diverse populations in precision medicine research. *The American Journal of Bioethics*, *18*(4), 3–20. doi:10.1080/15265161.2018.1431322

Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. *Current Directions in Psychological Science*, *23*(3), 159–163. doi:10.1177/0963721414528511

Kroll, J. F., Bobb, S., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, *9*(2), 119–135. doi:10.1017/S1366728906002483

Kroll, J. (2008). Juggling two languages in one mind. *Psychological Science Agenda*, *22*(1), 1–6.

Laird, F. M., Cai, H., Savonenko, A. V., Farah, M. H., He, K., Melnikova, T., …Wong, P. C. (2005). BACE1, a major determinant of selective vulnerability of the brain to amyloid-β amyloidogenesis, is essential for cognitive, emotional, and synaptic functions. *The Journal of Neuroscience*, *25*(50), 11693–11709. doi: 10.1523/JNEUROSCI.2766-05.2005

Lawton, D. M., Gasquoine, P. G., & Weimer, A. A. (2015). Age of dementia diagnosis in community dwelling bilingual and monolingual Hispanic Americans. *Cortex*, *66*, 141–145. doi:10.1016/j.cortex.2014.11.017

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin*, *144*(4), 394–425. doi:10.1037/bul0000142

Leivada, E., Westergaard, M., Duñabeitia, J. A., & Rothman, J. (2020). On the phantom-like appearance of bilingualism effects on neurocognition: (How) should we proceed? *Bilingualism: Language and Cognition*, 1–14. doi:10.1017/S1366728920000358

Lever, J., Krzywinski, M. & Altman, N. (2017). Principal component analysis. *Nature Methods*, *14*, 641–642. doi:10.1038/nmeth.4346

Lezak, M. D., Howieson, D. B., Loring, D. W., & Fischer, J. S. (2004). *Neuropsychological assessment*. Oxford University Press, USA.

Luk, G., Anderson, J. A. E., Craik, F. I. M., Grady, C., & Bialystok, E. (2010). Distinct neural correlates for two types of inhibition in bilinguals: Response inhibition versus interference suppression. *Brain and Cognition*, *74*(3), 347–357. doi:10.1016/j.bandc.2010.09.004

Luk, G., Bialystok, E., Craik, F. I. M., & Grady, C. L. (2011). Lifelong bilingualism maintains white matter integrity in older adults. *Journal of Neuroscience, 31*(46), 16808–16813. doi:10.1523/jneurosci.4563-11.2011

Luk, G., Green, D. W., Abutalebi, J., & Grady, C. (2011). Cognitive control for language switching in bilinguals: A quantitative meta-analysis of functional neuroimaging studies. *Language and Cognitive Processes*, *27*(10), 1479–1488. doi:10.1080/01690965.2011.613209

Madden, D. J., Bennett, I. J., & Song, A. W. (2009). Cerebral white matter integrity and cognitive aging: Contributions from diffusion tensor imaging. *Neuropsychological Review*, *19*(4), 415–435. doi:10.1007/s11065-009-9113-2

Marian, V., & Spivey, M. J. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*(2), 97–115. doi:10.1017/S1366728903001068

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods*, *9*(4), 403–425. doi:10.1037/1082-989X.9.4.403

Mehta, D., Jackson, R., Paul, G., Shi, J., & Sabbagh, M. (2017). Why do trials for Alzheimer's disease drugs keep failing? A discontinued drug perspective for 2010–2015. *Expert Opinion on Investigational Drugs*, *26*(6), 735–739. doi:10.1080/13543784.2017.1323868

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. doi:10.1371/journal.pone.0038268

Michielse, S., Coupland, N., Camicioli, R., Carter, R., Seres, P., Sabino, J., & Malykhin, N. (2010). Selective effects of aging on brain white matter microstructure: A diffusion tensor imaging tractography study. *NeuroImage*, *52*, 1190–1201. doi:10.1016/j.neuroimage.2010.05.019

Mitchell A. (2009). A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. *Journal of Psychiatric Research*, *43*(4), 411–431. doi:10.1016/j.jpsychires.2008.04.014

Morris, J. C., Roe, C. M., Xiong, C., Fagan, A. M., Goate, A. M., Holtzman, D. M., & Mintun, M. A. (2010). APOE predicts amyloid-beta but not tau Alzheimer pathology in cognitively normal aging. *Annals of Neurology*, *67*(1), 122–131. doi:10.1002/ana.21843

Mukadam, N., Sommerlad, A., & Livingston, G. (2017). The relationship of bilingualism compared to monolingualism to the risk of cognitive decline or dementia: A systematic review and meta-analysis. *Journal of Alzheimer's Disease*, *58*(1), 45–54. doi:10.3233/JAD-170131

Murphy, M. P., & LeVine, H. (2010). Alzheimer's disease and the amyloid-β peptide. *Journal of Alzheimer's Disease*, *19*(1), 311–323. doi:10.3233/JAD-2010-1221

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*(21). doi:10.3389/fnbot.2013.00021

Ossenkoppele, R., Schonhaut, D. R., Schöll, M., Lockhart, S. N., Ayakta, N., Baker, S. L., O'Neil, J. P., Janabi, M., Lazaris, A., Cantwell, A., Vogel, J., Santos, M., Miller, Z. A., Bettcher, B. M., Vossel, K. A., Kramer, J. H., Gorno-Tempini, M. L., Miller, B. L., Jagust, W. J., & Rabinovici, G. D. (2016). Tau PET patterns mirror clinical and neuroanatomical variability in Alzheimer's disease. *Brain*, *139*(5), 1551–1567. doi:10.1093/brain/aww027

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*(2), 232–258. doi:10.1016/j.cogpsych.2012.12.002

Paap, K. R., Johnson, H. A., & Sawi, O. (2016). Should the search for bilingual advantages in executive functioning continue? *Cortex*, *74*, 305–314. doi:10.1016/j.cortex.2015.09.010

Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: Problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology*, *5*, 962. doi:10.3389/fpsyg.2014.00962

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*(2), 299–320.

Perani, D., Farsad, M., Ballarini, T., Lubian, F., Malpetti, M., Fracchetti, A., Magnani, G., March, A., & Abutalebi, J. (2017). The impact of bilingualism on brain reserve and metabolic connectivity in Alzheimer's dementia. *Proceedings of The National Academy of*

Sciences, *114*(7), 1690–1695. doi:10.1073/pnas.1610909114

Pettigrew, C., Soldan, A., Li, S., Lu, Y., Wang, M. C., Selnes, O. A., Moghekar, A., O'Brien, R., Albert, M., & The Biocard Research Team (2013). Relationship of cognitive reserve and APOE status to the emergence of clinical symptoms in preclinical Alzheimer's disease. *Cognitive Neuroscience*, *4*(3-4), 136–142. doi:10.1080/17588928.2013.831820

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

Pliatsikas, C. (2020). Understanding structural plasticity in the bilingual brain: The Dynamic Restructuring Model. *Bilingualism: Language and Cognition*, *23*, 459–471. doi:10.1017/S1366728919000130

Polvikoski, T., Sulkava, R., Haltia, M., Kainulainen, K., Vuorio, A., Verkkoniemi, A., Niinistö, L., Halonen, P., & Kontula, K. (1995). Apolipoprotein E, dementia, and cortical deposition of beta-amyloid protein. *The New England Journal of Medicine*, *333*(19), 1242–1247. doi:10.1056/NEJM199511093331902

Prior, A., & Macwhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, *13*, 253–262. doi:10.1017/S1366728909990526

Qiu, C., Bäckman, L., Winblad, B., Agüero-Torres, H., & Fratiglioni, L. (2001). The influence of education on clinically diagnosed dementia incidence and mortality data from the Kungsholmen Project. *Archives of Neurology*, *58*(12), 2034–2039. doi:10.1001/archneur.58.12.2034

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in

observational studies for causal effects. *Biometrika*, *70*(1), 41–55. doi:10.1093/biomet/70.1.41

Rumbaut, R. G., & Massey, D. S. (2013). Immigration and language diversity in the United States. *Daedalus*, *142*(3), 141–154. doi:10.1162/daed_a_00224

Sala, G., & Gobet, F. (2020). Cognitive and academic benefits of music training with children: A multilevel meta-analysis. *Memory & Cognition*, *48*(8), 1429–1441. doi:10.3758/s13421-020-01060-2

Salat, D. H., Tuch, D. S., van der Kouwe, A. J. W., Greva, D. N., Pappu, V., Lee, S. Y., Hevelone, N. D., Zaleta, A. K., Growdon, J. H., Corkin, S., Fischl, B., & Rosas, H. D. (2010). White matter pathology isolates the hippocampal formation in Alzheimer's disease. *Neurobiology of Aging*, *31*(2), 244–256. doi:10.1016/j.neurobiolaging.2008.03.013

Salvatierra, J. L., & Rosselli, M. (2011). The effect of bilingualism and age on inhibitory control. *The International Journal of Bilingualism*, *15*, 26–37. doi:10.1177/1367006910371021

Sanders, A. E., Hall, C. B., Katz, M. J., & Lipton, R. B. (2012). Non-native language use and risk of incident dementia in the elderly. *Journal of Alzheimer's Disease*, *29*(1), 99–108. doi:10.3233/JAD-2011-111631

Sateren, W. B., Trimble, E. L., Abrams, J., Brawley, O., Breen, N., Ford, L., McCabe, M., Kaplan, R., Smith, M., Ungerleider, R., & Christian, M. C. (2002). How sociodemographics, presence of oncology specialists, and hospital cancer programs affect accrual to cancer treatment trials. *Journal of Clinical Oncology*, *20*(8), 2109–2117. doi:10.1200/JCO.2002.08.056

Scarmeas, N., Albert, S. M., Manly, J. J., & Stern, Y. (2006). Education and rates of cognitive decline in incident Alzheimer's disease. *Journal of Neurology, Neurosurgery, and*

Psychiatry, *77*(3), 308–316. doi:10.1136/jnnp.2005.072306

Scarmeas, N., & Stern, Y. (2003). Cognitive reserve and lifestyle. *Journal of Clinical and Experimental Neuropsychology*, *25*(5), 625–633. doi:10.1076/jcen.25.5.625.14576

Schilling, K. G., Blaber, J., Huo, Y., Newton, A., Hansen, C., Nath, V., … Landman, B. A. (2019). Synthesized b0 for diffusion distortion correction (Synb0-DisCo). *Magnetic Resonance Imaging*. doi:10.1016/j.mri.2019.05.008

Schmand, B., Huizenga, H., & Van Gool, W. (2010). Meta-analysis of CSF and MRI biomarkers for detecting preclinical Alzheimer's disease. *Psychological Medicine*, *40*(1), 135–145. doi:10.1017/S0033291709991516

Schmechel, D. E., Saunders, A. M., Strittmatter, W. J., Crain, B. J., Hulette, C. M., Joo, S. H., Pericak-Vance, M. A., Goldgaber, D., & Roses, A. D. (1993). Increased amyloid beta-peptide deposition in cerebral cortex as a consequence of apolipoprotein E genotype in late-onset Alzheimer disease. *PNAS*, *90*(20), 9649–9653. doi:10.1073/pnas.90.20.9649

Schroeder, S.R. (2018). Do bilinguals have an advantage in Theory of Mind? A meta-analysis. *Frontiers in Communication*, *3*, 36. doi:10.3389/fcomm.2018.00036

Schweizer, T. A., Ware, J., Fischer, C. E., Craik, F. I. M., & Bialystok, E. (2012). Bilingualism as a contributor to cognitive reserve: Evidence from brain atrophy in Alzheimer's disease. *Cortex*, *48*, 991–996. doi:10.1016/j.cortex.2011.04.009

Selkoe, D. J. (1994). Normal and abnormal biology of the β-amyloid precursor protein. *Annual Review of Neuroscience*, *17*, 489–517. doi: 10.1146/annurev.ne.17.030194.002421

Seshadri, S., Drachman, D. A., & Lippa, C. F. (1995). Apolipoprotein E e4 allele and the lifetime risk of Alzheimer's Disease. *Archives of Neurology*, *52*(11), 1074–1079. doi:10.1001/archneur.1995.00540350068018

Sexton, C. E., Walhovd, K. B., Storsve, A. B., Tamnes, C. K., Westlye, L. T., Johansen-Berg, H., & Fjell, A. M. (2014). Accelerated changes in white matter microstructure during aging: A longitudinal diffusion tensor imaging study. *Journal of Neuroscience*, *34*(46), 15425–15436. doi:10.1523/JNEUROSCI.0203-14.2014

Shoji, M., Matsubara, E., Kanai, M., Watanabe, M., Nakamura, T., Tomidokoro, Y., Shizuka, M., Wakabayashi, K., Igeta, Y., Ikeda, Y., Mizushima, K., Amari, M., Ishiguro, K., Kawarabayashi, T., Harigaya, Y., Okamoto, K., & Hirai, S. (1998). Combination assay of CSF Tau, Aβ1-40 and Aβ1-42(43) as a biochemical marker of Alzheimer's disease. *Journal of the Neurological Sciences*, *158*(2), 134–140. doi:10.1016/S0022-510X(98)00122-1

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology, 51*(3), 300–304. doi:10.1037/h0020586

Singh, N., & Mishra, R. K. (2012). Does language proficiency modulate oculomotor control? Evidence from Hindi–English bilinguals. *Bilingualism: Language and Cognition*, *15*, 771–781. doi:10.1017/S1366728912000065

Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Mathews, P. M., & Behrens, T. E. (2006). Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage*, *31*(4), 1487–1505. doi:10.1016/j.neuroimage.2006.02.024

Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D. R., & Markesbery, W. R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study. *JAMA*, *275*(7), 528–532.

Soldan, A., Pettigrew, C., Li, S., Wang, M.-C., Moghekar, A., Selnes, O. A., Albert, M., & O'Brien, R. (2013). Relationship of cognitive reserve and CSF biomarkers to emergence of clinical symptoms in preclinical Alzheimer's Disease. *Neurobiology of Aging*, *34*(12). doi:10.1016/j.neurobiolaging.2013.06.017

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*, *8*, 448–460. doi:10.1017.S1355617701020240

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia*, *47*(10), 2015–2028. doi:10.1016/j.neuropsychologia.2009.03.004

Stern, Y. (2012). Cognitive reserve in ageing and Alzheimer's disease. *The Lancet Neurology*, *11*(11), 1006–1012. doi: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. doi:10.1212/WNL.53.9.1942

Stern, Y., Arenaza-Urquijo, E. M., Bartrés-Faz, D., Belleville, S., Cantilon, M., Chetelat, G., Ewers, M., Franzmeier, N., Kempermann, G., Kremen, W. S., Okonkwo, O., Scarmeas, N., Soldan, A., Udeh-Momoh, C., Valenzuela, M., Vemuri, P., Vuoksimaa, E., Arenaza Urquiljo, E. M., Bartrés-Faz, D., … Vuoksimaa, E. (2020). Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimer's & Dementia*, S1552526018334915. doi:10.1016/j.jalz.2018.07.219

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662. doi:10.1037/h0054651

Tapiola, T., Alafuzoff, I., Herukka, S.-K., Parkkinen, L., Hartikainen, P., Soininen, H., & Pirttilä, T. (2009). Cerebrospinal fluid β-amyloid 42 and tau proteins as biomarkers of Alzheimer-type pathologic changes in the brain. *Archives of Neurology*, *66*(3), 382–389. doi:10.1001/archneurol.2008.596

U.S. Food and Drug Administration (2020, November). *Enhancing the diversity of clinical trial populations — Eligibility criteria, enrollment practices, and trial designs guidance for industry*. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/enhancing-diversity-clinical-trial-populations-eligibility-criteria-enrollment-practices-and-trial

Valenzuela, M. J., & Sachdev, P. (2006). Brain reserve and dementia: A systematic review. *Psychological Medicine*, *36*(4), 441–454. doi:10.1017/S0033291705006264

van den Noort, M., Struys, E., Bosch, P., Jaswetz, L., Perriard, B., Yeo, S., Barisch, P., Vermeire, K., Lee, S. H., & Lim, S. (2019). Does the bilingual advantage in cognitive control exist and if so, what are its modulating factors? A systematic review. *Behavioral Sciences (Basel, Switzerland)*, *9*(3), 27. doi:10.3390/bs9030027

von Bastian, C. C., Souza, A. S., & Gade, M. (2016). No evidence for bilingual cognitive advantages: A test of four hypotheses. *Journal of Experimental Psychology. General*, *145*(2), 246–258. doi:10.1037/xge0000120

Wang, H. X., Gustafson, D. R., Kivipelto, M., Pedersen, N. L., Skoog, I., Windblad, B., & Fratiglioni, L. (2012). Education halves the risk of dementia due to apolipoprotein ε4 allele: A collaborative study from the Swedish brain power initiative. *Neurobiology of Aging*, *33*(5), 1007.e1–7. doi:10.1016/j.neurobiolaging.2011.10.003

Ward, A., Crean, S., Mercaldi, C. J., Collins, J. M., Boyd, D., Cook, M. N., & Arrighi, H. M. (2012). Prevalence of apolipoprotein e4 genotype and homozygotes (APOE e4/4) among patients diagnosed with Alzheimer's Disease: A systematic review and meta-analysis. *Neuroepidemiology*, *38*(1), 1–17. doi:10.1159/000334607

Ware, A. T., Kirkovski, M., & Lum, J. A. G. (2020) Meta-analysis reveals a bilingual advantage that is dependent on task and age. *Frontiers in Psychology*, *11*, 1458. doi:10.3389/fpsyg.2020.01458

Weingarten, M. D., Lockwood, A. H., Hwo, S. Y., & Kirschner, M. W. (1975). A protein factor essential for microtubule assembly. *PNAS, 72*, 1858–1862. doi:10.1073/pnas.72.5.1858

Wilson, R. S., Boyle, P. A., Yang, J., James, B. D., & Bennett, D. A. (2015). Early life instruction in foreign language and music and incidence of mild cognitive impairment. *Neuropsychology*, *29*(2), 292–302. doi:10.1037/neu0000129

World Health Organization (2020, September 21). *Dementia*. http://www.who.int/en/news-room/fact-sheets/detail/dementia

Woumans, E., Santens, P., Sieben, A., Versijpt, J., Stevens, M., & Duyck, W. (2015). Bilingualism delays clinical manifestation of Alzheimer's disease. *Bilingualism: Language and Cognition, 18*(3), 568–574. doi:10.1017/S136672891400087X

Yeung, C. M., St John, P. D., Menec, V., & Tyas, S. L. (2014). Is bilingualism associated with a lower risk of dementia in community-living older adults? Cross-sectional and prospective analyses. *Alzheimer Disease and Associated Disorders*, *28*(4), 326–332. doi:10.1097/WAD.0000000000000019

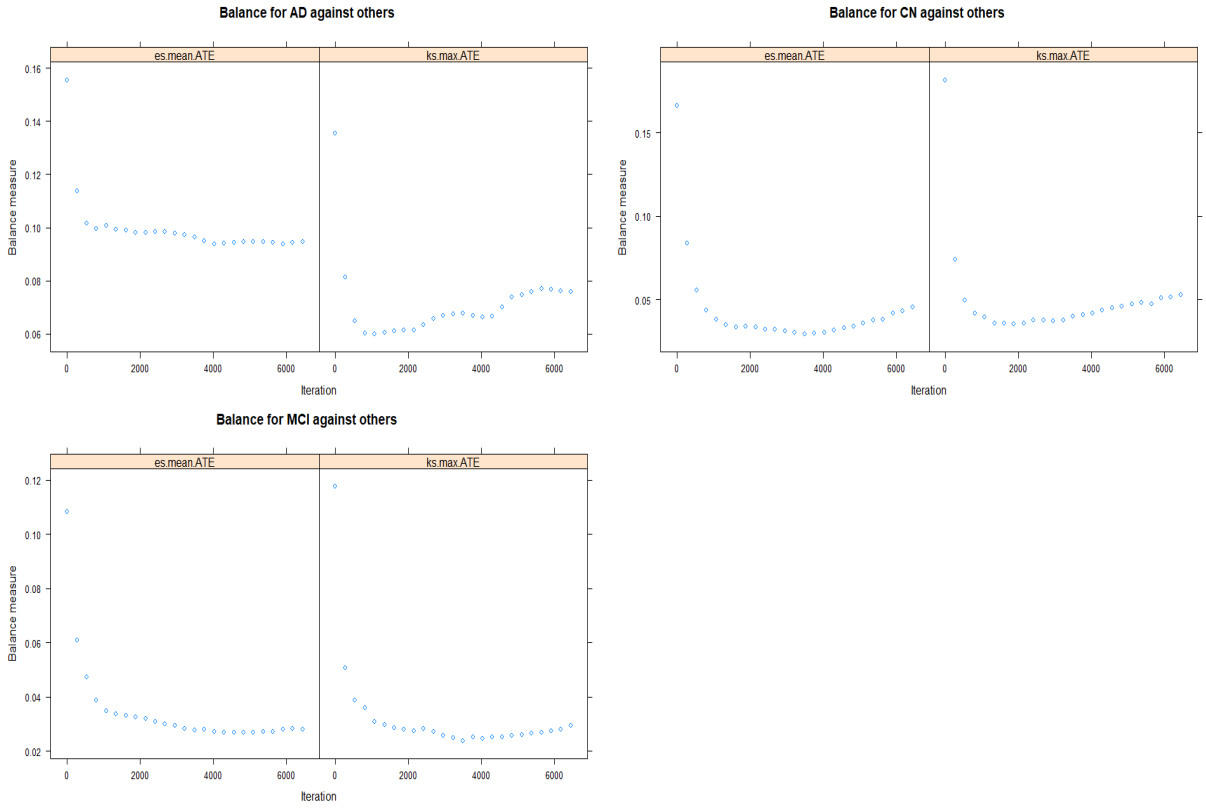
Zahodne, L. B., Schofield, P. W., Farrell, M. T., Stern, Y., & Manly, J. J. (2014). Bilingualism does not alter cognitive decline or dementia risk among Spanish-speaking immigrants. *Neuropsychology*, *28*(2), 238–246. doi:10.1037/neu0000014

Zhang, H., Wu, Y. J., & Thierry, G. (2020). Bilingualism and aging: A focused neuroscientific review. *Journal of Neurolinguistics*, *54*, 100890. doi:10.1016/j.jneuroling.2020.100890

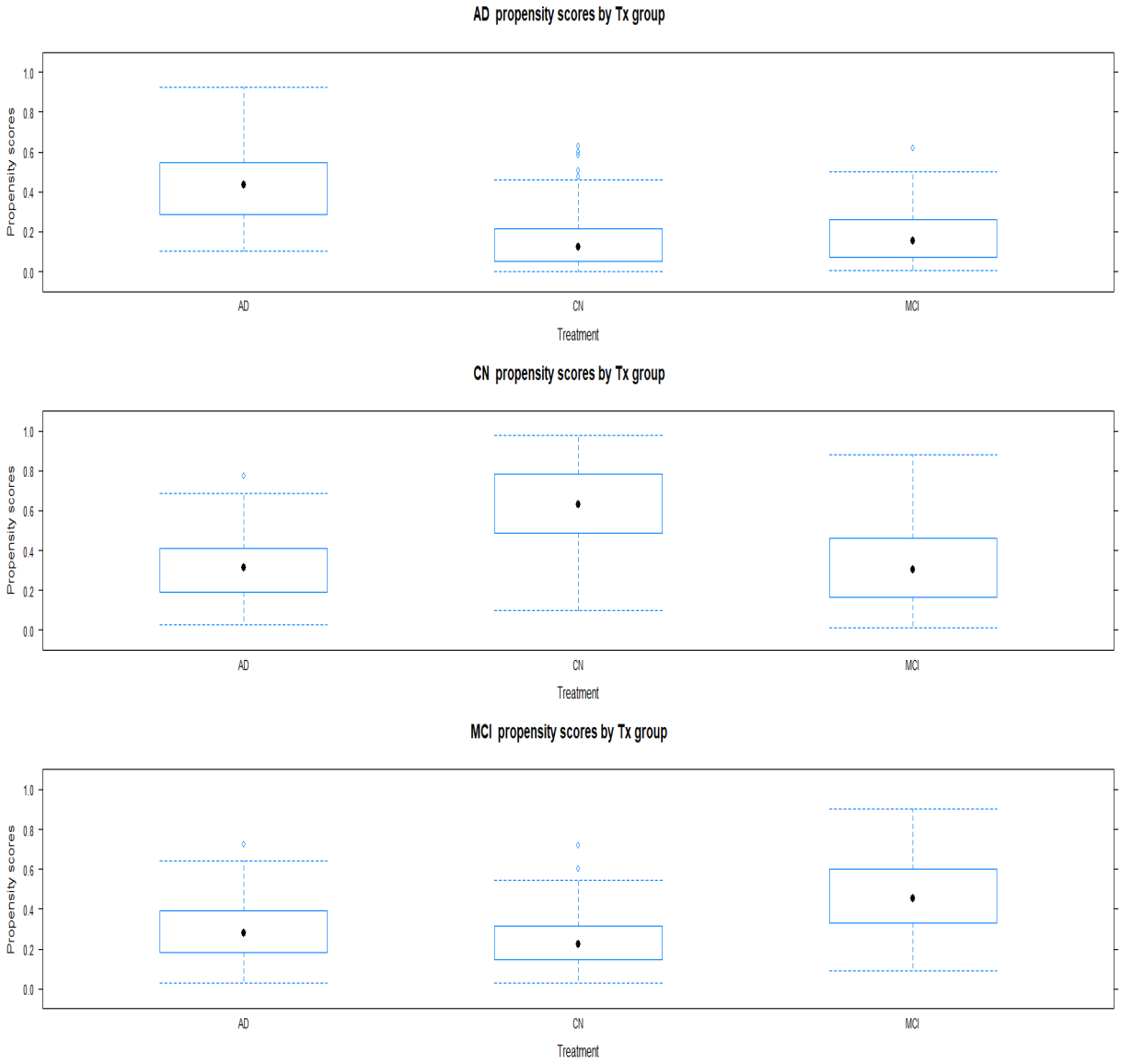
Zheng, Y., Wu, Q., Su, F., Fang, Y., Zeng, J., & Pei, Z. (2018). The protective effect of Cantonese/Mandarin bilingualism on the onset of Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, *45*(3–4), 210–219. doi:10.1159/000488485

APPENDIX A: GBM BALANCE PLOTS AND TABLES

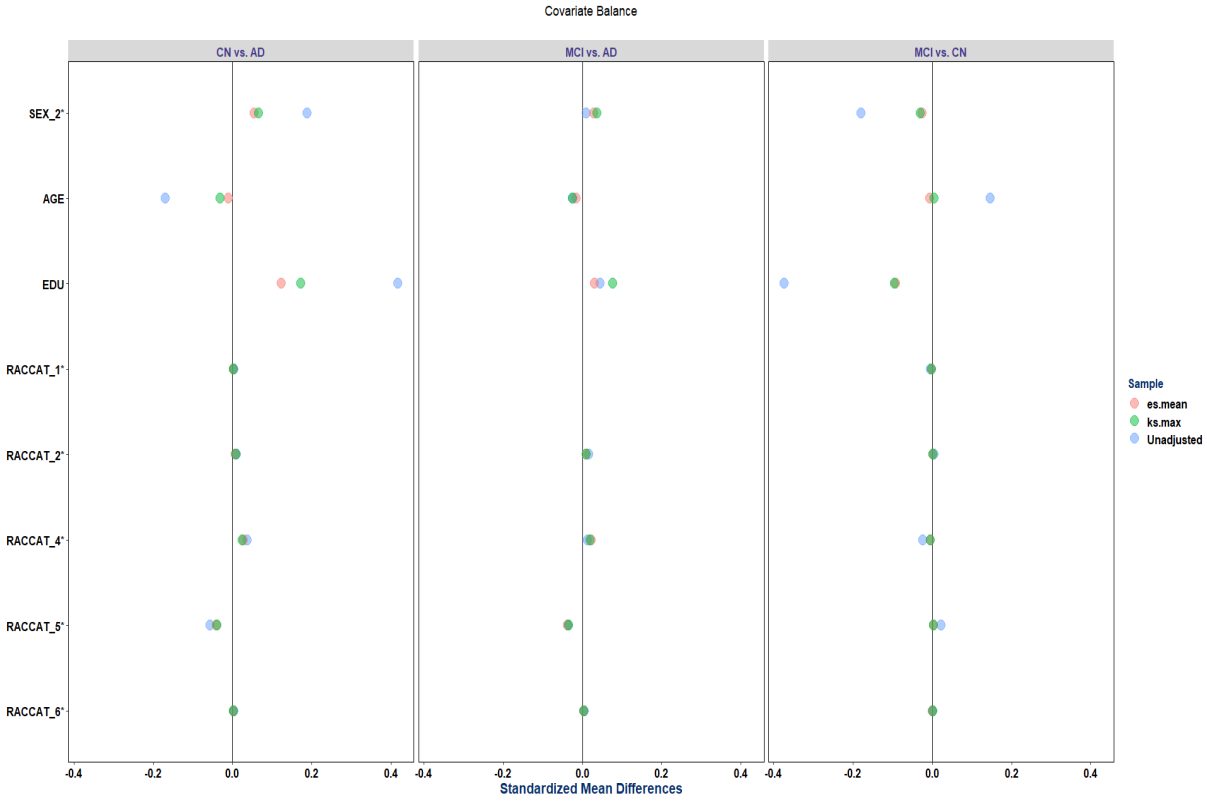
Model 1, Plot 1



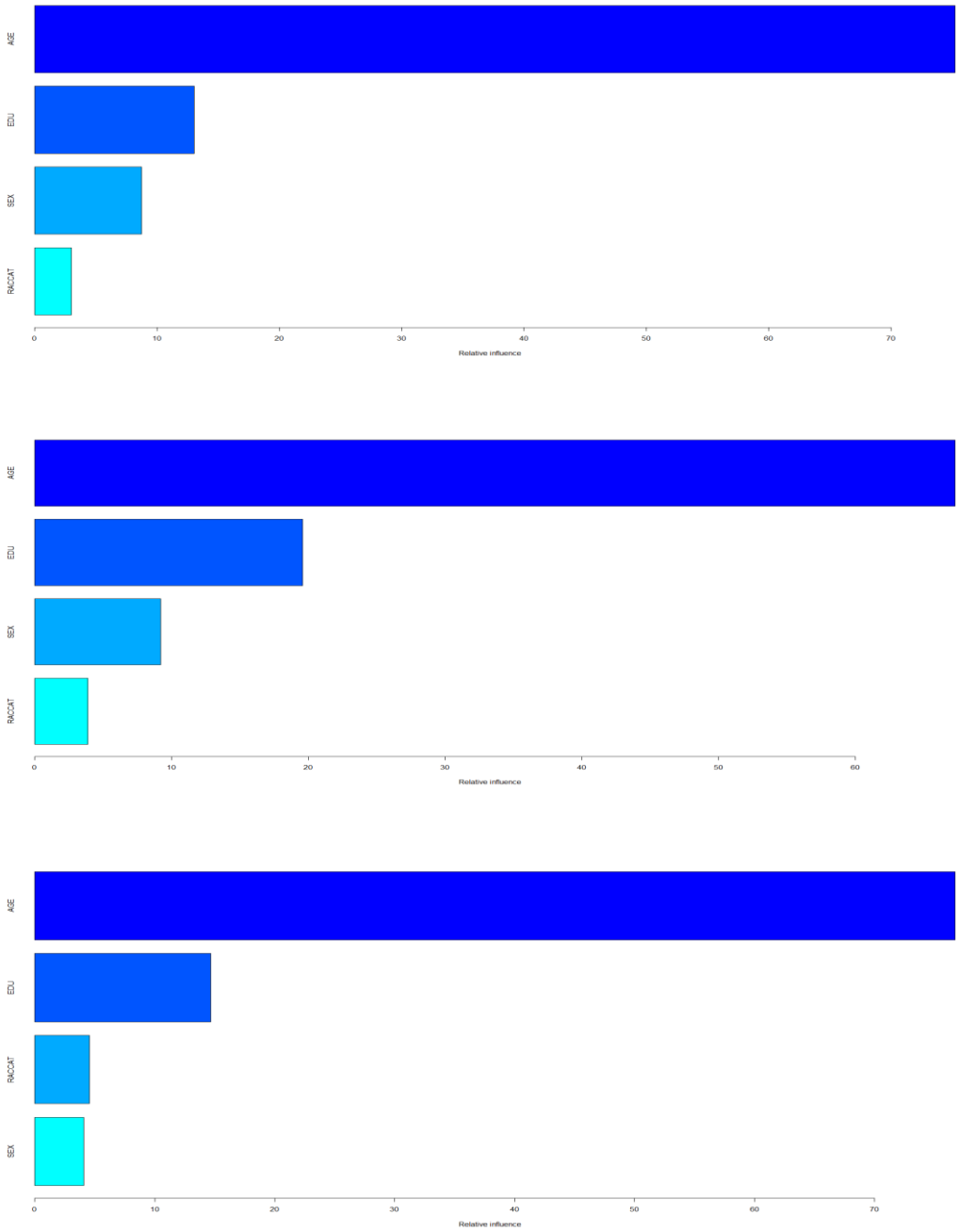
Model 1, Plot 2



Model 1, Plot 3



Model 1, Plot 4. *Relative influence of predictors on AD, CN, and MCI, respectively*.



Model 1 Balance Tables

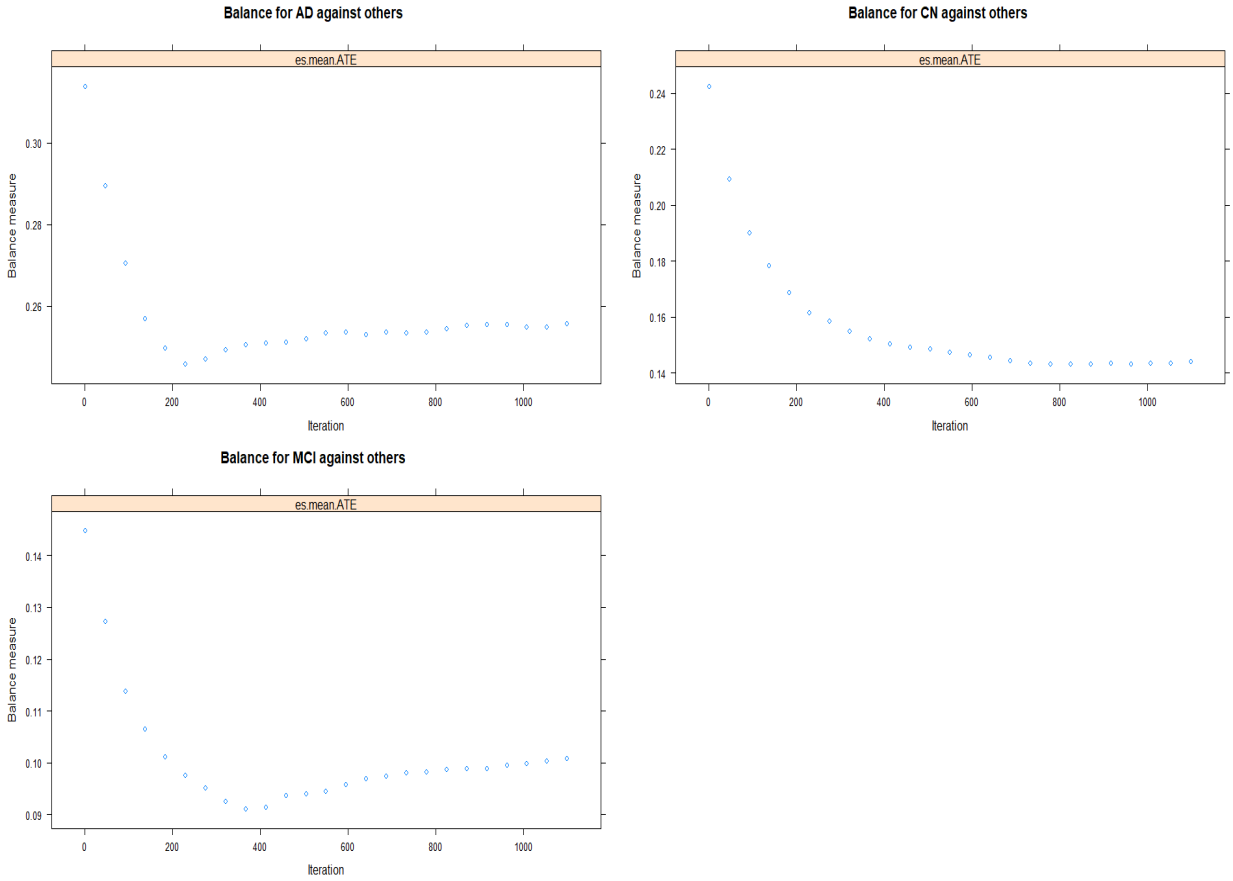
Unweighted and weighted comparisons for AD, CN, and MCI, respectively.

Variables	unw.tx.mn	unw.tx.sd	unw.ct.mn	unw.ct.sd	unw.std.eff.sz	unw.stat	unw.p	es.mean.ATE.tx.mn	es.mean.ATE.tx.sd	es.mean.ATE.ct.mn	es.mean.ATE.ct.sd	es.mean.ATE.std.eff.sz	es.mean.ATE.stat	es.mean.ATE.p
SEX:1	0.611	0.487	0.497	0.5	0.229	5.16	0.023	0.566	0.496	0.518	0.5	0.096	0.696	0.404
SEX:2	0.389	0.487	0.503	0.5	-0.229	NA	NA	0.434	0.496	0.482	0.5	-0.096	NA	NA
AGE	74.688	8.158	74.032	7.611	0.086	1.043	0.297	74.145	7.463	74.032	7.611	0.015	0.198	0.843
EDU	15.484	2.889	16.043	2.715	-0.206	-2.508	0.012	15.862	2.791	16.043	2.715	-0.067	-0.761	0.447
RACCAT:1	0	0	0.002	0.047	-0.053	NA	NA	0	0	0.002	0.043	-0.044	NA	NA
RACCAT:2	0	0	0.013	0.115	-0.131	NA	NA	0	0	0.011	0.105	-0.11	NA	NA
RACCAT:4	0.008	0.089	0.035	0.185	-0.163	NA	NA	0.005	0.068	0.03	0.171	-0.15	NA	NA
RACCAT:5	0.992	0.089	0.945	0.229	0.229	NA	NA	0.995	0.068	0.953	0.211	0.203	NA	NA
RACCAT:6	0	0	0.004	0.066	-0.075	NA	NA	0	0	0.004	0.06	-0.062	NA	NA
RACCAT:7	0	0	0	0	NA	NA	NA	0	0	0	0	NA	NA	NA

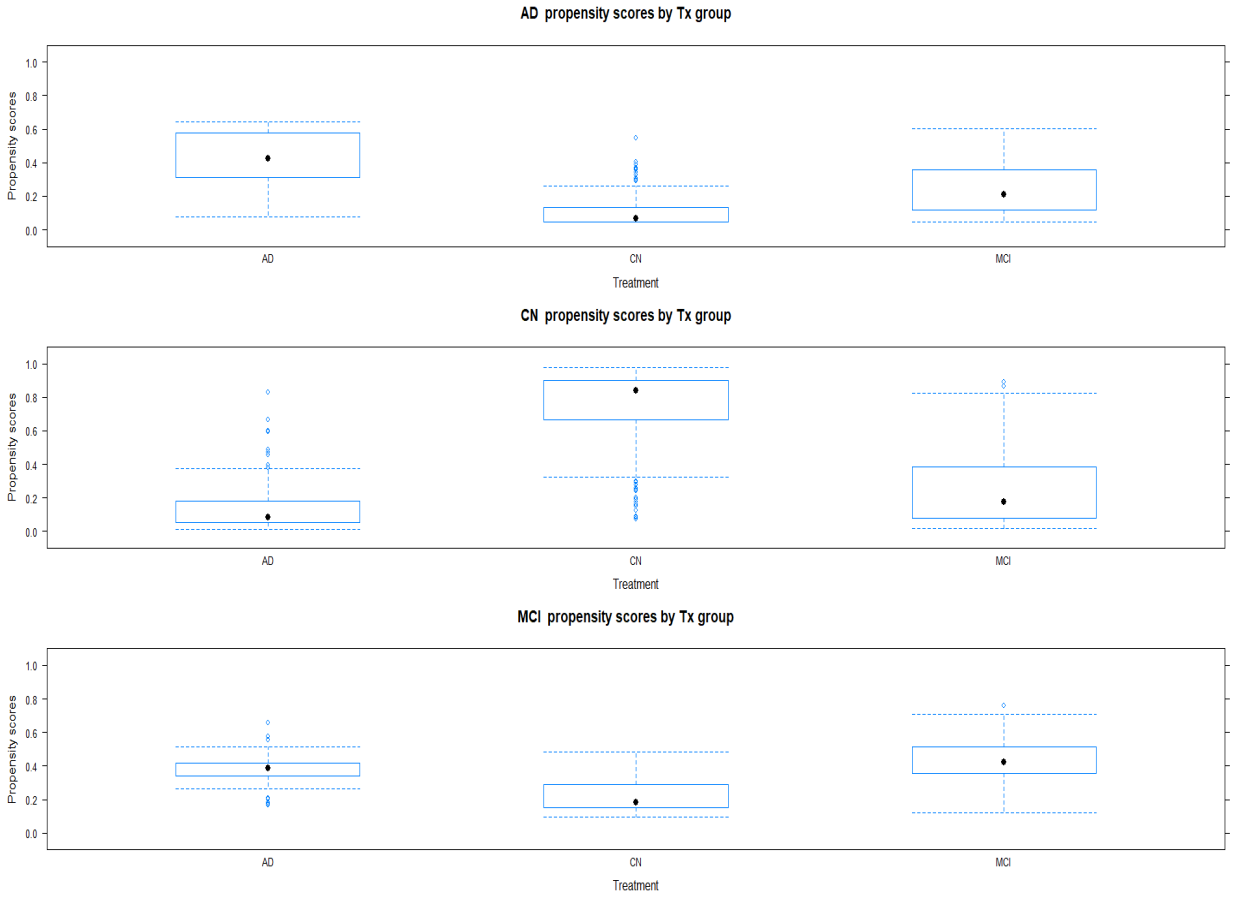
Variables	unw.tx.mn	unw.tx.sd	unw.ct.mn	unw.ct.sd	unw.std.eff.sz	unw.stat	unw.p	es.mean.ATE.tx.mn	es.mean.ATE.tx.sd	es.mean.ATE.ct.mn	es.mean.ATE.ct.sd	es.mean.ATE.std.eff.sz	es.mean.ATE.stat	es.mean.ATE.p
SEX:1	0.422	0.494	0.605	0.489	-0.366	19.184	0	0.511	0.5	0.542	0.498	-0.061	0.404	0.525
SEX:2	0.578	0.494	0.395	0.489	0.366	NA	NA	0.489	0.5	0.458	0.498	0.061	NA	NA
AGE	73.376	7.058	74.032	7.611	-0.086	-1.922	0.055	74.07	7.232	74.032	7.611	0.005	-0.21	0.834
EDU	16.624	2.434	16.043	2.715	0.214	4.857	0	16.2	2.581	16.043	2.715	0.058	1.133	0.258
RACCAT:1	0.004	0.062	0	0	0.091	NA	NA	0.002	0.046	0	0	0.05	NA	NA
RACCAT:2	0.011	0.106	0.01	0.097	0.018	NA	NA	0.008	0.09	0.008	0.089	0.002	NA	NA
RACCAT:4	0.046	0.209	0.016	0.125	0.176	NA	NA	0.032	0.177	0.034	0.18	-0.007	NA	NA
RACCAT:5	0.935	0.246	0.971	0.167	-0.173	NA	NA	0.955	0.208	0.955	0.208	0	NA	NA
RACCAT:6	0.004	0.062	0.003	0.056	0.011	NA	NA	0.003	0.05	0.004	0.06	-0.019	NA	NA
RACCAT:7	0	0	0	0	NA	NA	NA	0	0	0	0	NA	NA	NA

Variables	unw.tx.mn2	unw.tx.sd	unw.ct.mn	unw.ct.sd	unw.std.eff.sz	unw.stat	unw.p	es.mean.ATE.tx.mn	es.mean.ATE.tx.sd	es.mean.ATE.ct.mn	es.mean.ATE.ct.sd	es.mean.ATE.std.eff.sz	es.mean.ATE.stat	es.mean.ATE.p
SEX:1	0.601	0.49	0.483	0.5	0.236	7.033	0.008	0.536	0.499	0.513	0.5	0.047	0.228	0.633
SEX:2	0.399	0.49	0.517	0.5	-0.236	NA	NA	0.464	0.499	0.487	0.5	-0.047	NA	NA
AGE	74.51	7.94	74.032	7.611	0.063	1.027	0.305	74.026	7.348	74.032	7.611	-0.001	-0.066	0.947
EDU	15.606	2.82	16.043	2.715	-0.161	-2.645	0.008	15.948	2.685	16.043	2.715	-0.035	-0.491	0.624
RACCAT:1	0	0	0.003	0.051	-0.062	NA	NA	0	0	0.002	0.044	-0.047	NA	NA
RACCAT:2	0.016	0.125	0.008	0.087	0.081	NA	NA	0.01	0.101	0.008	0.091	0.021	NA	NA
RACCAT:4	0.021	0.144	0.033	0.18	-0.072	NA	NA	0.028	0.166	0.031	0.174	-0.018	NA	NA
RACCAT:5	0.957	0.202	0.954	0.21	0.018	NA	NA	0.958	0.201	0.956	0.206	0.01	NA	NA
RACCAT:6	0.005	0.073	0.003	0.051	0.047	NA	NA	0.004	0.06	0.003	0.053	0.013	NA	NA
RACCAT:7	0	0	0	0	NA	NA	NA	0	0	0	0	NA	NA	NA

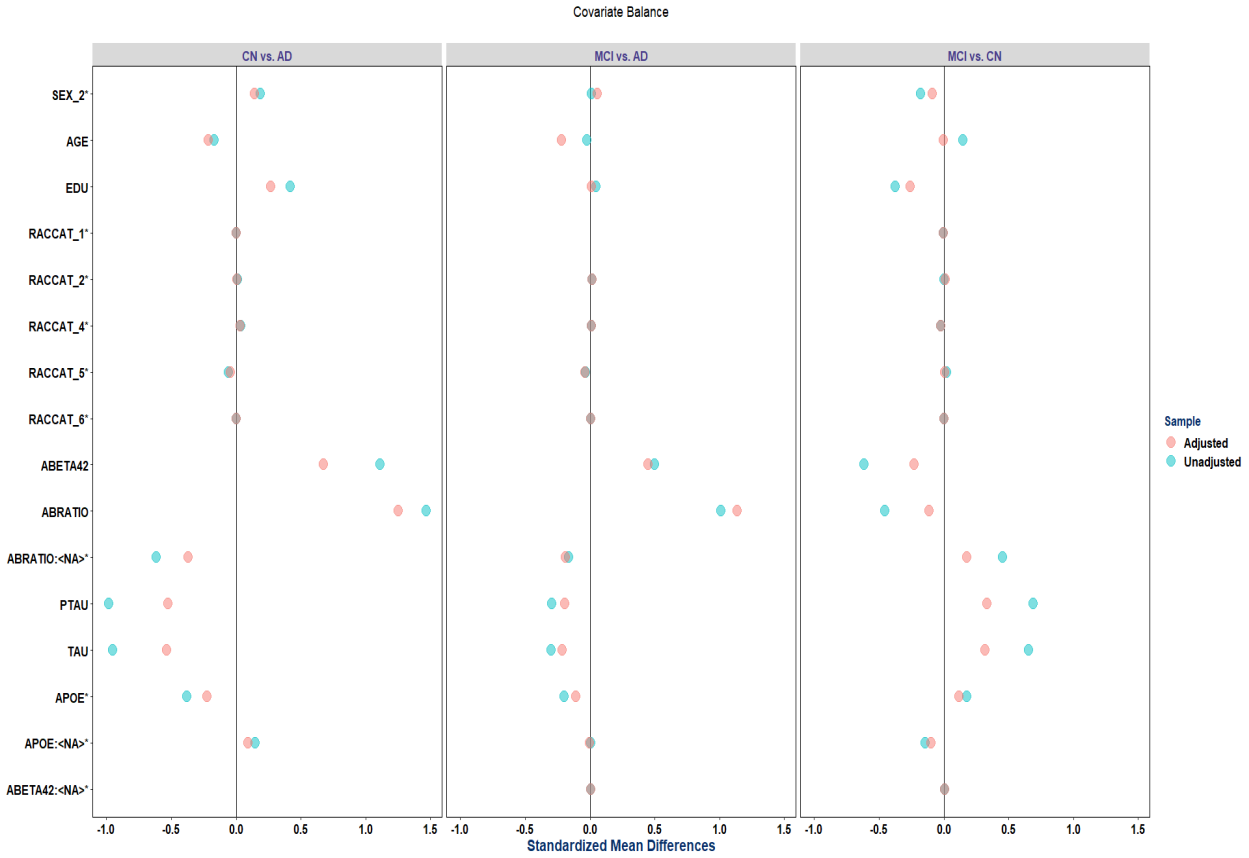
Model 2. Plot 1



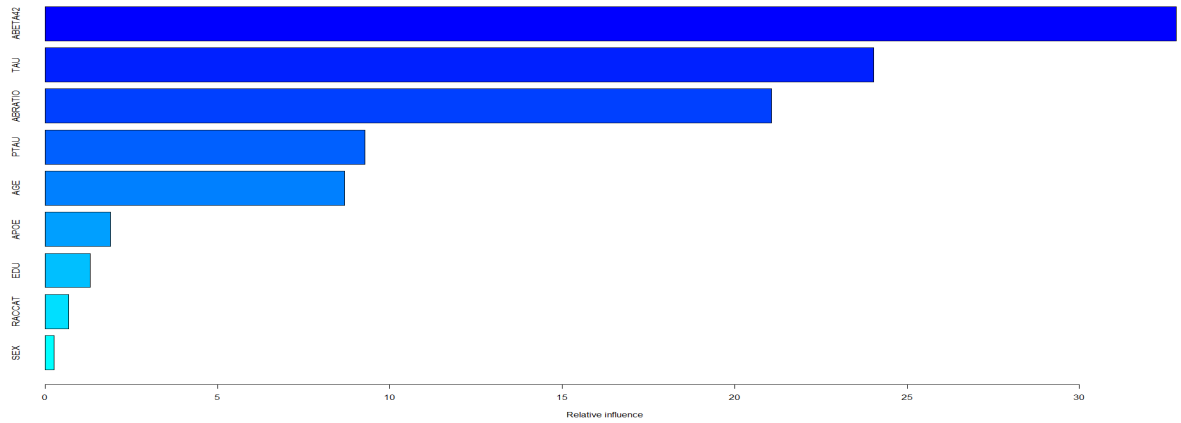
Model 2, Plot 2

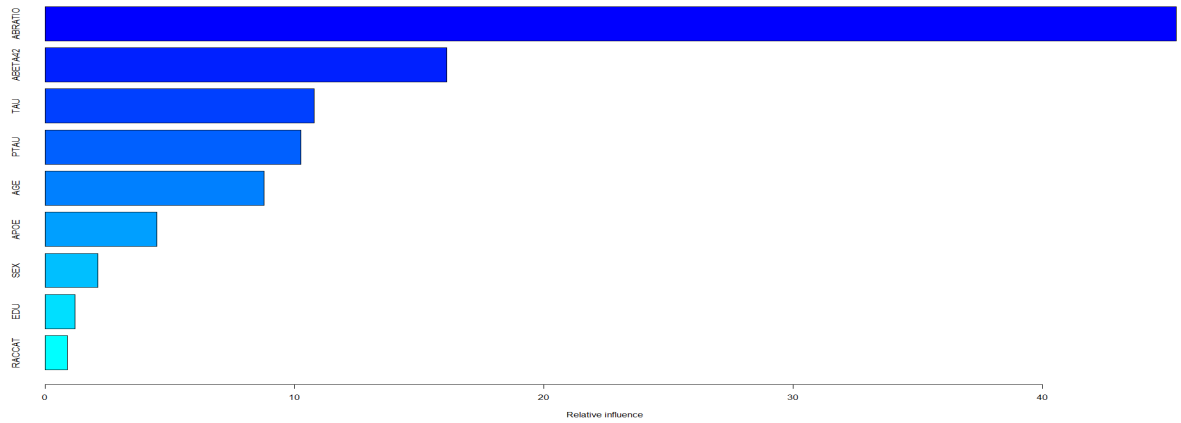


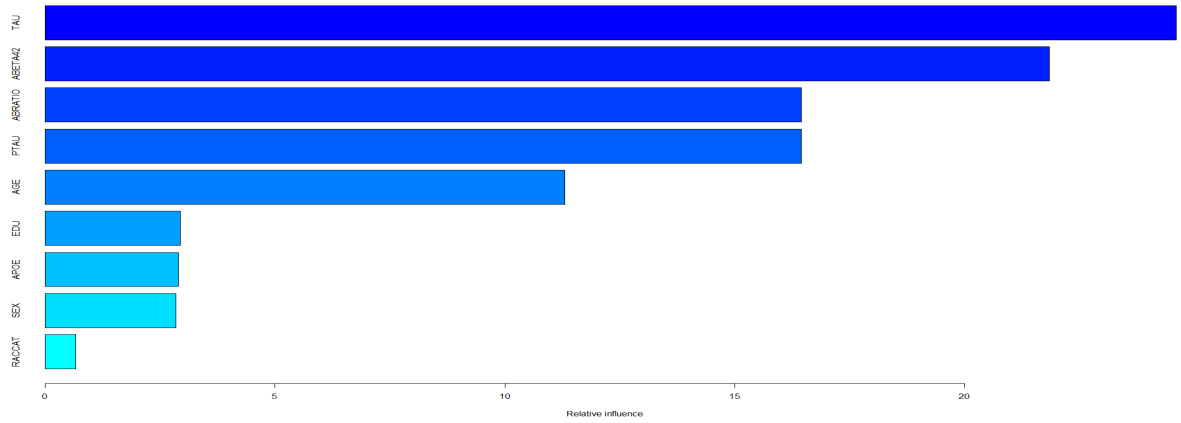
Model 2, Plot 3



Model 2, Plot 4. *Relative influence of predictors on AD, CN, and MCI, respectively.*







Model 2 Balance Tables

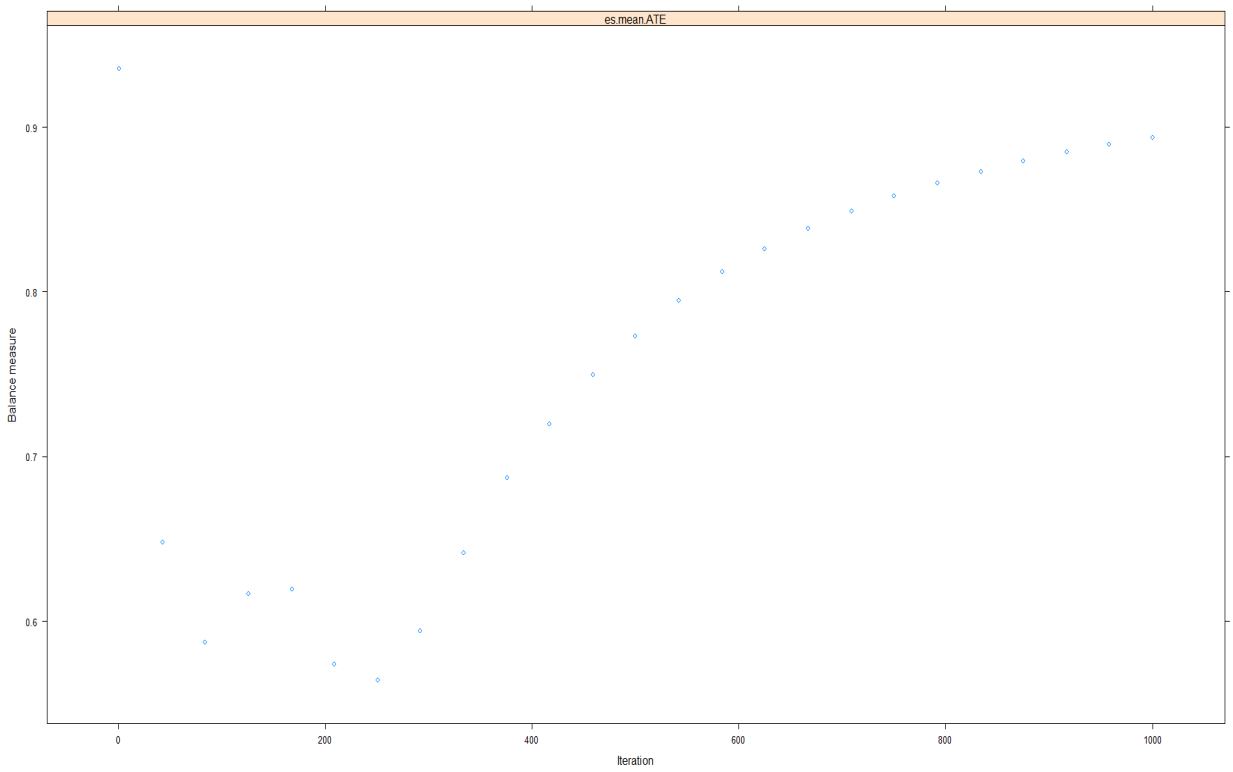
Unweighted and weighted comparisons for AD, CN, and MCI, respectively.

Variables	unw.tx.mn	unw.tx.sd	unw.ct.mn	unw.ct.sd	unw.std.eff.sz	unw.stat	unw.p	es.mean.ATE.tx.mn	es.mean.ATE.tx.sd	es.mean.ATE.ct.mn	es.mean.ATE.ct.sd	es.mean.ATE.std.eff.sz	es.mean.ATE.stat	es.mean.ATE.p
SEX:1	0.611	0.487	0.497	0.5	0.229	5.16	0.023	0.637	0.481	0.507	0.5	0.259	5.262	0.022
SEX:2	0.389	0.487	0.503	0.5	-0.229	NA	NA	0.363	0.481	0.493	0.5	-0.259	NA	NA
AGE	74.688	8.158	74.032	7.611	0.086	1.043	0.297	76.488	8.191	74.032	7.611	0.323	2.579	0.01
EDU	15.484	2.889	16.043	2.715	-0.206	-2.508	0.012	15.714	2.94	16.043	2.715	-0.121	-1.171	0.242
RACCAT:1	0	0	0.002	0.047	-0.053	NA	NA	0	0	0.002	0.043	-0.045	NA	NA
RACCAT:2	0	0	0.013	0.115	-0.131	NA	NA	0	0	0.013	0.112	-0.126	NA	NA
RACCAT:4	0.008	0.089	0.035	0.185	-0.163	NA	NA	0.005	0.07	0.035	0.183	-0.176	NA	NA
RACCAT:5	0.992	0.089	0.945	0.229	0.229	NA	NA	0.995	0.07	0.947	0.225	0.233	NA	NA
RACCAT:6	0	0	0.004	0.066	-0.075	NA	NA	0	0	0.004	0.063	-0.067	NA	NA
RACCAT:7	0	0	0	0	NA	NA	NA	0	0	0	0	NA	NA	NA
ABETA42	657.89	365.102	1028.082	628.122	-0.589	-10.644	0	762.969	427.276	1028.082	628.122	-0.422	-5.326	0
ABETA42:<NA>	0	0	0.002	0.042	-0.042	-16.374	0	0	0	0.002	0.042	-0.042	-16.697	0
ABRATIO	0.037	0.01	0.067	0.027	-1.117	-10.483	0	0.04	0.013	0.067	0.027	-1.006	-5.332	0
ABRATIO:<NA>	0.881	0.324	0.548	0.498	0.67	7.501	0	0.823	0.382	0.548	0.498	0.553	4.335	0
PTAU	35.43	14.983	27.676	14.842	0.522	6.675	0	31.535	15.147	27.676	14.842	0.26	2.475	0.014
PTAU:<NA>	0	0	0.002	0.042	-0.042	-16.374	0	0	0	0.002	0.042	-0.042	-16.697	0
TAU	353.518	127.893	286.942	130.527	0.51	6.658	0	322.449	130.239	286.942	130.527	0.272	2.652	0.008
TAU:<NA>	0	0	0.002	0.042	-0.042	-16.374	0	0	0	0.002	0.042	-0.042	-16.697	0
APOE:0	0.278	0.448	0.503	0.5	-0.453	20.287	0	0.378	0.485	0.475	0.499	-0.194	3.185	0.042
APOE:1	0.659	0.474	0.344	0.475	0.64	NA	NA	0.537	0.499	0.385	0.487	0.307	NA	NA
APOE:<NA>	0.063	0.244	0.153	0.36	-0.263	NA	NA	0.085	0.279	0.14	0.347	-0.161	NA	NA

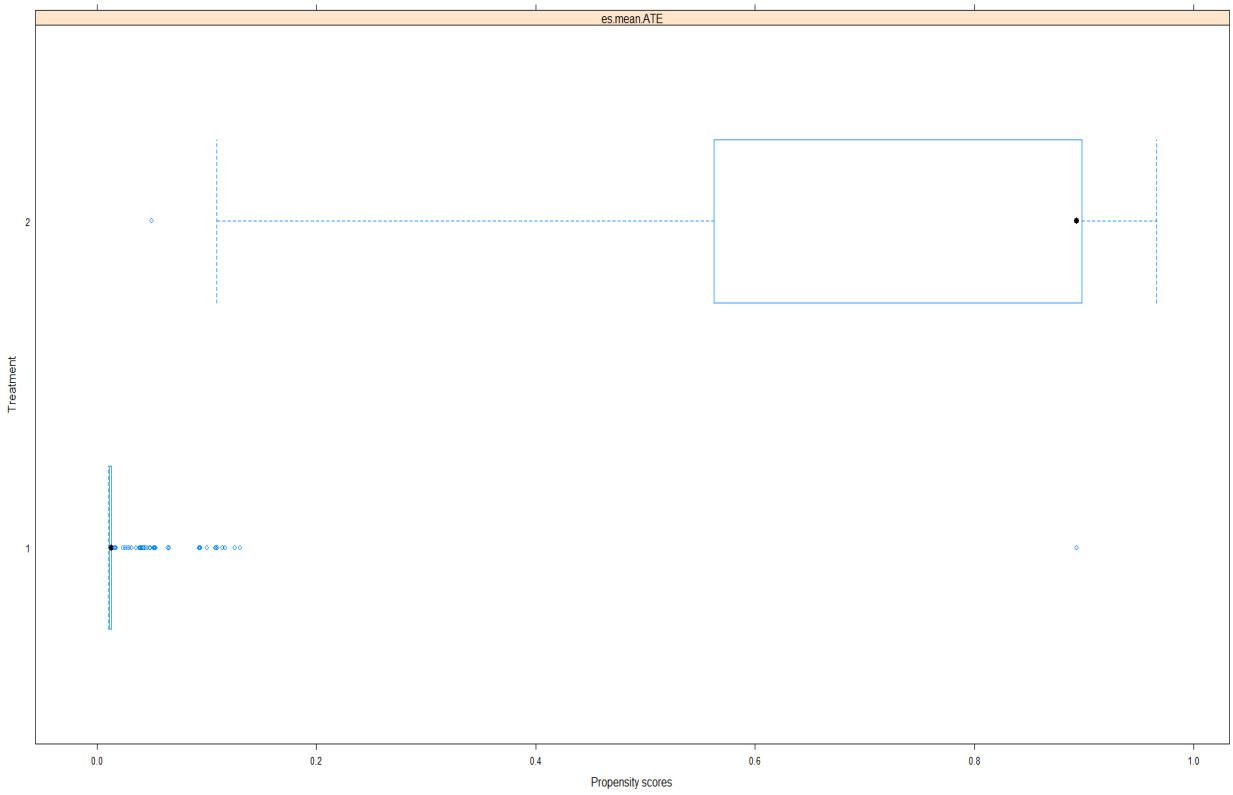
Variables	unw.tx.mn	unw.tx.sd	unw.ct.mn	unw.ct.sd	unw.std.eff.sz	unw.stat	unw.p	es.mean.ATE.tx.mn	es.mean.ATE.tx.sd	es.mean.ATE.ct.mn	es.mean.ATE.ct.sd	es.mean.ATE.std.eff.sz	es.mean.ATE.stat	es.mean.ATE.p
SEX:1	0.422	0.494	0.605	0.489	-0.366	19.184	0	0.493	0.5	0.584	0.493	-0.183	2.916	0.088
SEX:2	0.578	0.494	0.395	0.489	0.366	NA	NA	0.507	0.5	0.416	0.493	0.183	NA	NA
AGE	73.376	7.058	74.032	7.611	-0.086	-1.922	0.055	74.818	7.099	74.032	7.611	0.103	-0.119	0.906
EDU	16.624	2.434	16.043	2.715	0.214	4.857	0	16.447	2.42	16.043	2.715	0.149	2.537	0.011
RACCAT:1	0.004	0.062	0	0	0.091	NA	NA	0.003	0.05	0	0	0.06	NA	NA
RACCAT:2	0.011	0.106	0.01	0.097	0.018	NA	NA	0.008	0.089	0.012	0.111	-0.044	NA	NA
RACCAT:4	0.046	0.209	0.016	0.125	0.176	NA	NA	0.037	0.189	0.013	0.115	0.139	NA	NA
RACCAT:5	0.935	0.246	0.971	0.167	-0.173	NA	NA	0.95	0.218	0.97	0.171	-0.095	NA	NA
RACCAT:6	0.004	0.062	0.003	0.056	0.011	NA	NA	0.003	0.052	0.005	0.067	-0.031	NA	NA
RACCAT:7	0	0	0	0	NA	NA	NA	0	0	0	0	NA	NA	NA
ABETA42	1272.544	631.131	1028.082	628.122	0.389	9.055	0	1136.705	638.076	1028.082	628.122	0.173	3.054	0.002
ABETA42:<NA>	0	0	0.002	0.042	-0.042	-17.758	0	0	0	0.002	0.042	-0.042	-17.808	0
ABRATIO	0.071	0.025	0.067	0.027	0.153	4.066	0	0.069	0.026	0.067	0.027	0.076	1.483	0.139
ABRATIO:<NA>	0.266	0.442	0.548	0.498	-0.566	-11.752	0	0.453	0.498	0.548	0.498	-0.191	-3.668	0
PTAU	21.435	9.714	27.676	14.842	-0.421	-10.451	0	24.034	11.038	27.676	14.842	-0.245	-4.095	0
PTAU:<NA>	0	0	0.002	0.042	-0.042	-17.758	0	0	0	0.002	0.042	-0.042	-17.808	0
TAU	234.27	87.957	286.942	130.527	-0.404	-9.941	0	255.585	97.79	286.942	130.527	-0.24	-4.053	0
TAU:<NA>	0	0	0.002	0.042	-0.042	-17.758	0	0	0	0.002	0.042	-0.042	-17.808	0
APOE:0	0.532	0.499	0.389	0.487	0.289	29.22	0	0.523	0.499	0.44	0.496	0.167	7.621	0.001
APOE:1	0.255	0.436	0.545	0.498	-0.589	NA	NA	0.297	0.457	0.478	0.5	-0.367	NA	NA
APOE:<NA>	0.213	0.409	0.067	0.25	0.429	NA	NA	0.179	0.384	0.082	0.274	0.286	NA	NA

Variables	unw.tx.mn	unw.tx.sd	unw.ct.mn	unw.ct.sd	unw.std.eff.sz	unw.stat	unw.p	es.mean.ATE.tx.mn	es.mean.ATE.tx.sd	es.mean.ATE.ct.mn	es.mean.ATE.ct.sd	es.mean.ATE.std.eff.sz	es.mean.ATE.stat	es.mean.ATE.p
SEX:1	0.601	0.49	0.483	0.5	0.236	7.033	0.008	0.581	0.493	0.507	0.5	0.149	2.475	0.116
SEX:2	0.399	0.49	0.517	0.5	-0.236	NA	NA	0.419	0.493	0.493	0.5	-0.149	NA	NA
AGE	74.51	7.94	74.032	7.611	0.063	1.027	0.305	74.807	7.721	74.032	7.611	0.102	1.107	0.269
EDU	15.606	2.82	16.043	2.715	-0.161	-2.645	0.008	15.741	2.809	16.043	2.715	-0.111	-1.57	0.117
RACCAT:1	0	0	0.003	0.051	-0.062	NA	NA	0	0	0.002	0.047	-0.053	NA	NA
RACCAT:2	0.016	0.125	0.008	0.087	0.081	NA	NA	0.019	0.137	0.007	0.081	0.124	NA	NA
RACCAT:4	0.021	0.144	0.033	0.18	-0.072	NA	NA	0.016	0.127	0.033	0.178	-0.098	NA	NA
RACCAT:5	0.957	0.202	0.954	0.21	0.018	NA	NA	0.959	0.199	0.956	0.205	0.013	NA	NA
RACCAT:6	0.005	0.073	0.003	0.051	0.047	NA	NA	0.006	0.076	0.002	0.048	0.06	NA	NA
RACCAT:7	0	0	0	0	NA	NA	NA	0	0	0	0	NA	NA	NA
ABETA42	933.701	618.63	1028.082	628.122	-0.15	-2.529	0.012	1011.37	634.71	1028.082	628.122	-0.027	-0.332	0.74
ABETA42:<NA>	0.005	0.073	0.002	0.042	0.086	18.263	0	0.006	0.075	0.002	0.042	0.094	17.244	0
ABRATIO	0.061	0.029	0.067	0.027	-0.24	-1.822	0.07	0.066	0.027	0.067	0.027	-0.022	-0.257	0.797
ABRATIO:<NA>	0.718	0.45	0.548	0.498	0.342	5.609	0	0.633	0.482	0.548	0.498	0.172	2.076	0.038
PTAU	31.23	16.991	27.676	14.842	0.239	3.722	0	28.775	15.358	27.676	14.842	0.074	1.166	0.244
PTAU:<NA>	0.005	0.073	0.002	0.042	0.086	18.263	0	0.006	0.075	0.002	0.042	0.094	17.244	0
TAU	316.161	151.642	286.942	130.527	0.224	3.447	0.001	295.48	135.535	286.942	130.527	0.065	1.043	0.298
TAU:<NA>	0.005	0.073	0.002	0.042	0.086	18.263	0	0.006	0.075	0.002	0.042	0.094	17.244	0
APOE:0	0.463	0.499	0.45	0.497	0.026	5.379	0.005	0.477	0.499	0.441	0.497	0.071	2.001	0.136
APOE:1	0.468	0.499	0.386	0.487	0.168	NA	NA	0.441	0.497	0.414	0.493	0.056	NA	NA
APOE:<NA>	0.069	0.254	0.165	0.371	-0.28	NA	NA	0.082	0.274	0.145	0.352	-0.185	NA	NA

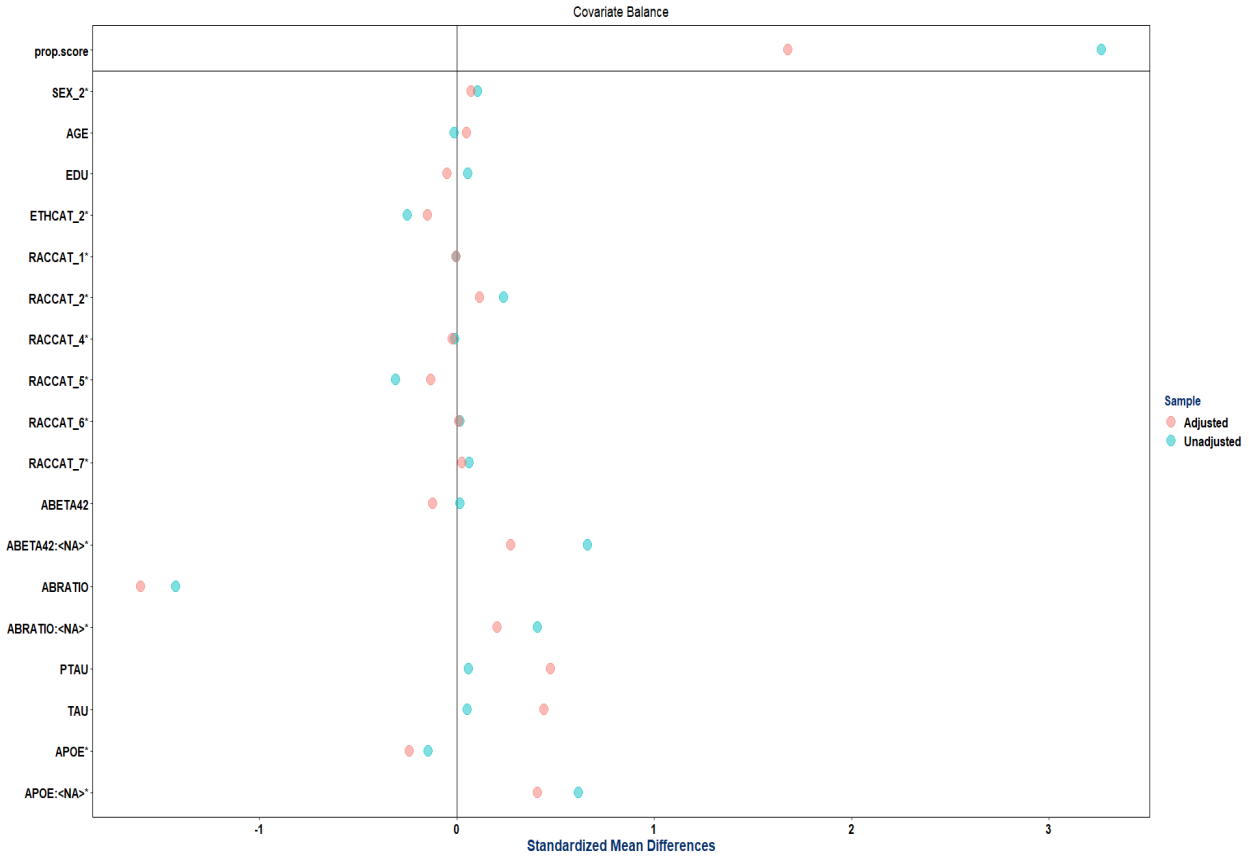
Model 3, Plot 1



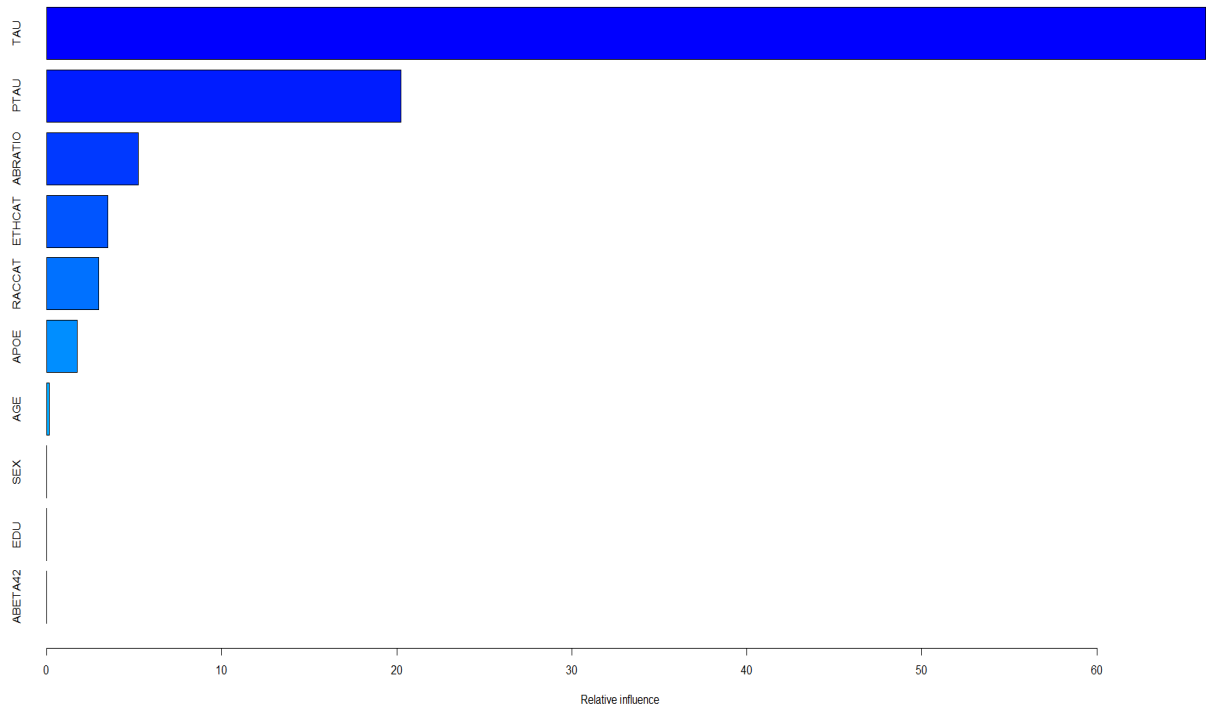
Model 3, Plot 2



Model 3, Plot 3



Model 3, Plot 4



Model 3 Balance table

Variables	unw.tx.mn	unw.tx.sd	unw.ct.mn	unw.ct.sd	unw.std.eff.sz	unw.stat	unw.p	es.mean.ATE.tx.mn	es.mean.ATE.tx.sd	es.mean.ATE.ct.mn	es.mean.ATE.ct.sd	es.mean.ATE.std.eff.sz	es.mean.ATE.stat	es.mean.ATE.p
SEX:1	0.417	0.493	0.522	0.5	-0.21	1.952	0.163	0.441	0.497	0.515	0.5	-0.147	0.365	0.546
SEX:2	0.583	0.493	0.478	0.5	0.21	NA	NA	0.559	0.497	0.485	0.5	0.147	NA	NA
AGE	73.941	8.128	74.032	7.611	-0.012	-0.076	0.94	74.642	6.409	74.246	7.727	0.052	0.314	0.753
EDU	16.208	3.222	16.043	2.715	0.06	0.348	0.728	15.862	2.842	16.003	2.719	-0.051	-0.241	0.809
ETHCAT:1	0.25	0.433	0	0	1.822	NA	NA	0.148	0.355	0	0	1.078	NA	NA
ETHCAT:2	0.75	0.433	1	0	-1.822	NA	NA	0.852	0.355	1	0	-1.078	NA	NA
ETHCAT:3	0	0	0	0	NA	NA	NA	0	0	0	0	NA	NA	NA
RACCAT:1	0	0	0.002	0.042	-0.043	26.376	0	0	0	0.002	0.041	-0.042	10.025	0
RACCAT:2	0.25	0.433	0.01	0.101	1.433	NA	NA	0.128	0.334	0.011	0.106	0.695	NA	NA
RACCAT:4	0.021	0.143	0.029	0.169	-0.052	NA	NA	0.009	0.092	0.029	0.168	-0.124	NA	NA
RACCAT:5	0.646	0.478	0.955	0.207	-1.221	NA	NA	0.824	0.381	0.954	0.209	-0.515	NA	NA
RACCAT:6	0.021	0.143	0.003	0.059	0.251	NA	NA	0.015	0.12	0.004	0.061	0.158	NA	NA
RACCAT:7	0.062	0.242	0	0	0.904	NA	NA	0.026	0.158	0	0	0.37	NA	NA
ABETA42	1037.819	483.03	1028.082	628.122	0.016	0.08	0.937	959.602	468.731	1028.274	628.086	-0.11	-0.467	0.64
ABETA42:<NA>	0.667	0.471	0.002	0.042	2.973	6.73	0	0.29	0.454	0.016	0.124	1.228	3.035	0.003
ABRATIO	0.036	0.011	0.067	0.027	-1.15	-4.028	0	0.032	0.01	0.067	0.027	-1.295	-5.214	0
ABRATIO:<NA>	0.958	0.2	0.548	0.498	0.832	4.046	0	0.76	0.427	0.555	0.497	0.415	1.18	0.238
PTAU	28.731	20.072	27.676	14.842	0.07	0.211	0.833	36.259	18.882	27.749	14.95	0.567	1.51	0.132
PTAU:<NA>	0.667	0.471	0.002	0.042	2.973	6.73	0	0.29	0.454	0.016	0.124	1.228	3.035	0.003
TAU	295.488	174.464	286.942	130.527	0.065	0.196	0.844	356.72	157.673	287.554	131.423	0.525	1.558	0.12
TAU:<NA>	0.667	0.471	0.002	0.042	2.973	6.73	0	0.29	0.454	0.016	0.124	1.228	3.035	0.003
APOE:0	0.167	0.373	0.454	0.498	-0.58	56.93	0	0.339	0.473	0.447	0.497	-0.217	7.54	0.001
APOE:1	0.083	0.276	0.412	0.492	-0.676	NA	NA	0.104	0.305	0.404	0.491	-0.617	NA	NA
APOE:<NA>	0.75	0.433	0.133	0.34	1.602	NA	NA	0.557	0.497	0.149	0.356	1.06	NA	NA