The role of physical activity and bilingualism in the development of neurodegenerative disorders: cross-sectional and neuroimaging evidence.

Stefano Brini
Bachelor of Psychology (Hons)

This thesis is presented for the degree of

Doctor of Philosophy

from

Murdoch University

2020

Discipline of Psychology, Exercise Science, Chiropractic and Counselling

Murdoch University
Author’s Declaration

I declare that:

a) The thesis is my own account of my research, except where other sources are acknowledged.

b) The extent to which the work of others has been used is clearly stated in each chapter and certified by my supervisors.

c) The thesis contains as its main content, work that has not been previously submitted for a degree at any other university.

Stefano Brini
Statement of Contribution of Others

I declare that this PhD thesis includes my own research. I developed the i) research question in each chapter ii) study protocol for each systematic review, iii) first draft of each chapter as well as iv) analyzed the data when necessary. In the published/under review papers, I was responsible for the submission of the article and addressing reviewers’ comments. To the best of my knowledge, this PhD thesis does not contain any material published previously by any other person except where due acknowledgement has been made.

I also declare the greater part of the workload involved in preparing these articles as the first author. Also, the co-authors presented below contributed to the relevant articles by providing their own respective expertise in analyzing and interpreting the data as well as providing critical commentary during the development of each article. Each author provided final approval before submission of each chapter. Authors contributed to the following chapters:

**Chapter 2:** Dr. Sohrabi, Dr. Peiffer, Dr. Karrasch, Prof. Hämäläinen, Prof. Martins, and Dr. Fairchild

**Chapter 3:** Dr. Sohrabi, Prof. Hebert, Mr. Forrest, Prof. Laine, Prof Hämäläinen, Dr. Karrasch, Dr. Peiffer, Prof. Martins, Dr. Fairchild

**Chapter 4:** Dr. Shahtahmassebi, Dr. Sohrabi, Dr. Karrasch, Prof. Martins, and Dr. Fairchild

**Chapter 5:** Prof. Veliz, Dr. Fairchild, Dr. Sohrabi, Prof. Martins, Dr. Peiffer, Prof. Hebert

**Chapter 6:** Dr. Shahtahmassebi, Prof. Hebert, and Dr. Fairchild

________________________________________________________________________

Timothy Fairchild
Heikki Hämäläinen

Hamid Sohrabi

Jeffrey Hebert

Jeremiah Peiffer

Philip Veliz

Matti Laine

Mira Karrasch

Behnaz Shahtahmassebi

Mitchell Forrest

Ralph Martins
Notes on formatting and referencing style

Chapter two (*Sports Medicine*), three (*Neuropsychology Review*), five (*Journal of Alzheimer’s Disease*), and six (*Trials*) were formatted according to their respective journals’ guidelines. Therefore, the formatting and referencing styles are not consistent throughout the thesis but reflects the journals’ formatting and referencing guidelines. Chapter four, which has not yet been drafted for submission to a particular journal, follows the referencing guidelines described in the *Publication Manual of the American Psychological Association* (APA) 6th Edition.
Thesis Abstract

Delaying Alzheimer’s disease (AD), the most common form of dementia, by five years could decrease the global prevalence of AD by 57% and halve the annual economic impact which is currently estimated to surpass US$2 trillion by 2030. Since no treatment or cure for dementia exist, identifying modifiable factors to reduce the incidence of dementia has become a public health priority. Increased physical activity (PA) has been associated with a lower risk of developing dementia in observational studies. Observational studies have also linked bilingualism (the ability to speak two languages) with a delayed onset of dementia but no risk-reduction in dementia in bilinguals relative to monolinguals. Differences in study outcomes in the fields of PA- and bilingualism- related research to methodological limitations including poor measurement of the exposure (PA and language profiles), small sample sizes, and recruitment of participants with different dementia etiologies. The purpose of this thesis was twofold: i) to explore the roles of PA and bilingualism in dementia risk and ii) to inform the development of a randomized controlled trial (RCT) to test whether studying a foreign language combined with increasing PA can improve cognitive performance in seniors who are at a higher risk of developing AD.

The aim of Chapter two was to review the available evidence linking PA with the risk of developing dementia as well as to explore the effects of increasing PA on cognition in individuals with dementia. Results showed that aerobic, and high-intensity, habitual PA was associated with improved cognition-related biomarkers and lower dementia risk in epidemiological studies. Experimental evidence showed increasing PA improved cognition-related biomarkers and cognition in preclinical phases of dementia, but not in clinical phases. The findings showed that PA is linked with a lower risk of dementia in epidemiological studies, but experimental studies showed little to no improvements in cognition in
participants with dementia following a structured PA program. There was evidence indicating that increasing PA levels in the preclinical phase of AD may result in greater translation impact than in participants at the more advanced clinical stage of AD. Most studies assessed PA with self-report measures questioning the accuracy and precision of exposure and recruited participants with dementia irrespective of aetiology, which makes it problematic to discern whether PA is differentially related to varying dementia aetiologies.

The aim of Chapter three was to systematically review the association between bilingualism and the delay in the diagnosis of dementia and AD. Here, we retrieved a total of 20 studies, 15 of which were meta-analysed. Results showed that bilinguals were on average 3.2 (95% CI: 1.5, 4.9) years older than monolinguals at the time of dementia. Moreover, at the time of dementia diagnosis, bilinguals and monolinguals demonstrated a similar level of global cognitive impairment (Hedges’ \( g = 0.05 \) 95% CI: -0.10, 0.21). Prediction intervals however showed a large dispersion of effect sizes in the meta-analysis comparing monolinguals to bilinguals on the age of dementia diagnosis. To explore possible reasons for the observed dispersion in effect sizes, we conducted subgroup analyses. In one subgroup meta-analysis comparing studies that had recruited participants with dementia to studies that had recruited participants with AD, bilinguals were 4.2 (95% CI: 2.0, 6.2) and 1.7 (95% CI: -1.4, 4.7) years older than monolinguals at dementia and AD diagnosis, respectively. Meta-analytic results combining prospective longitudinal studies showed no risk reduction in dementia among bilinguals compared to monolinguals (Odds Ratio: 0.85; 95% CI: 0.69-1.05). Risk of bias assessment revealed that most studies carried several methodological limitations including poor measurement of participants’ language profiles and small sample sizes.
The aim of Chapter four was to explore the underlying mechanisms in the brain that may be responsible for the observed findings in the first systematic review (Chapter three). In this study, we observed that bilinguals compared to monolinguals had greater brain volume in the frontostriatal and frontoparietal circuits. Also, functional neuroimaging studies showed that bilinguals made use of relevant brain areas more efficiently than monolinguals when completing interference cognitive tasks. Results from the cross-sectional studies showed that higher levels of language acculturation were associated with significantly greater verbal and psychomotor speed performance than lower levels of language acculturation.

The aim of the Chapter five was to explore the link between language acculturation and cognition in older individuals from ethnic minorities (Hispanic and Asian) living in the United States of America using an epidemiological dataset. In this cross-sectional epidemiological study, we analyzed data from the National Health and Nutrition Examination Survey using a larger sample size than previous studies. We found that higher levels of language acculturation (i.e. speaking the native language and that of the recipient’s country at home) was associated with greater psychomotor speed processing than lower levels of language acculturation (mostly speaking the native language at home) and some, but not all, measures of verbal fluency. Overall, the findings suggest that higher levels of language acculturation are associated with greater cognitive performance in older individuals from ethnic minorities.

Overall, the evidence gathered in the previous chapters indicate that i) increasing PA in individuals who are at a higher risk of developing AD might be more useful in improving cognitive performance than in individuals who already have developed AD and ii) bilingualism might render the brain areas typically affected by AD such as the frontostriatal
and frontoparietal brain circuits more resilient against neurodegeneration and in turn, delay the onset of AD symptoms and diagnosis. Therefore, because no randomized-controlled trial (RCT) testing the combined effects of increased PA with studying a foreign language currently exist, Chapter five presents a detailed study protocol for an RCT addressing this gap in the literature while addressing the limitations of previous studies in the fields of PA- and bilingualism-based research.

The purpose of this thesis was to explore the roles of PA and bilingualism in dementia onset and risk and to inform the development of a randomized controlled trial (RCT) testing the effects of studying a foreign language with increased PA in individuals at a higher risk of dementia. Increasing PA levels are associated with greater cognition in individuals at the preclinical phase of AD rather than in participants with a diagnosis of dementia or AD. Bilingualism was also associated with later age of AD diagnosis on average by 4.7 years. This finding is clinically relevant because a five-year delay in the onset of AD could lower the number of individuals with AD worldwide by 57% and as a consequence, halving the associated economic costs. Moreover, we also showed that bilingualism may be responsible for rendering brain areas typically affected by AD more resilient against neuropathology. Moreover, this thesis revealed that studies in the field of exercise science and bilingualism research in dementia carry important methodological limitations that question the internal validity of these two lines of research. Consequently, the evidence gathered within this thesis led us to propose a novel RCT exploring the effects of increasing PA levels and studying a foreign language in monolingual individuals at a higher risk of AD while addressing the most important limitations of previous research.
Acknowledgements

This PhD was the product of an international collaboration among difference universities and institutions in Perth, Western Australia, and Turku, in Finland including Murdoch University, University of Turku, Åbo Akademi University, the Australian Alzheimer's Research Foundation, and the Turku Brain and Mind centre. I would like to thank the staff in each institution that contributed and helped me in completing this PhD.

More specifically, I would like to thank my primary supervisor Tim who helped me develop as a researcher starting from my undergraduate degree through to my PhD. Thanks to him, I learnt the application of critical and analytical thinking in appraising scientific evidence and many practical aspects of research. It is also thanks to him that this PhD was made possible and completed with the successful publication of two manuscripts.

I would also like to thank Hamid Sohrabi, Jeremiah Peiffer and the Finnish research group: Heikki Hämäläinen, Matti Laine, Mira Karrasch, and Juha Salmitaival who helped me developed and publish manuscripts from this PhD. In particular, I would like to thank Hamid and Jeremiah for teaching me important aspects of Alzheimer’s disease research – I have gained knowledge in this field of research that will continue to help me in future research projects. Many thanks also go to the Finnish research group who have also made this PhD project possible by providing critical input into the development of all research projects herein. Their expertise has helped me gain knowledge in bilingualism-based research and allowed me to publish in this field of research.

Finally, I thank my parents Tuula and Marino for having supported me in many different ways since the beginning of my undergraduate degree until the end of my PhD. Without their support I would not have been able to complete my academic journey and begin my academic career. It is thanks to them that I will be able to purse a profession that I truly enjoy doing.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author’s Declaration</td>
<td>2</td>
</tr>
<tr>
<td>Statement of Contribution of Others</td>
<td>3</td>
</tr>
<tr>
<td>Notes on formatting and referencing style</td>
<td>5</td>
</tr>
<tr>
<td>Thesis Abstract</td>
<td>6</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>10</td>
</tr>
<tr>
<td>Contents</td>
<td>11</td>
</tr>
<tr>
<td>List of published manuscripts</td>
<td>16</td>
</tr>
<tr>
<td>List of abbreviations</td>
<td>17</td>
</tr>
<tr>
<td><strong>Chapter 1</strong> Introduction</td>
<td>19</td>
</tr>
<tr>
<td><strong>Background</strong></td>
<td>20</td>
</tr>
<tr>
<td><strong>Putative underlying mechanisms</strong></td>
<td>25</td>
</tr>
<tr>
<td><strong>Purpose of this thesis</strong></td>
<td>26</td>
</tr>
<tr>
<td><strong>Chapter 2</strong> Physical activity in preventing Alzheimer’s disease and cognitive decline: a narrative review.</td>
<td>35</td>
</tr>
<tr>
<td><strong>Abstract</strong></td>
<td>37</td>
</tr>
<tr>
<td><strong>Introduction</strong></td>
<td>38</td>
</tr>
<tr>
<td><strong>The pathophysiology of AD</strong></td>
<td>39</td>
</tr>
<tr>
<td><strong>Risk Factors for AD</strong></td>
<td>40</td>
</tr>
<tr>
<td>The relationship between subjective cognitive decline and AD</td>
<td>40</td>
</tr>
<tr>
<td>Cardiovascular diseases and low PA as risk factors for AD</td>
<td>41</td>
</tr>
<tr>
<td>PA, cognition and risk of AD: Epidemiological findings</td>
<td>46</td>
</tr>
<tr>
<td>PA and cognition in dementia: Experimental findings</td>
<td>50</td>
</tr>
<tr>
<td>Genetic markers and risk of cognitive decline and AD</td>
<td>54</td>
</tr>
<tr>
<td>Apolipoprotein ε4 allele and PA</td>
<td>55</td>
</tr>
<tr>
<td>Brain-derived neurotrophic factors (BDNF) and PA</td>
<td>56</td>
</tr>
<tr>
<td><strong>Brain biomarkers in AD pathophysiology and their interaction with PA</strong></td>
<td>59</td>
</tr>
<tr>
<td>PA findings on Pittsburgh compound B-PET</td>
<td>59</td>
</tr>
<tr>
<td>PA findings on [18F] fluorodeoxyglucose</td>
<td>61</td>
</tr>
<tr>
<td><strong>Limitations of studies on PA and AD with suggestions for future research</strong></td>
<td>62</td>
</tr>
<tr>
<td>Epidemiological data</td>
<td>62</td>
</tr>
<tr>
<td>Experimental data</td>
<td>62</td>
</tr>
<tr>
<td>Suggestions for future research</td>
<td>63</td>
</tr>
<tr>
<td><strong>Conclusions</strong></td>
<td>64</td>
</tr>
<tr>
<td><strong>Compliance with Ethical Standards</strong></td>
<td>66</td>
</tr>
<tr>
<td>Funding</td>
<td>66</td>
</tr>
<tr>
<td>Conflict of interest</td>
<td>66</td>
</tr>
<tr>
<td><strong>References</strong></td>
<td>67</td>
</tr>
<tr>
<td><strong>Between chapter (2-3) rationale</strong></td>
<td>85</td>
</tr>
</tbody>
</table>
Chapter 3 Bilingualism is associated with a delayed onset of dementia but not with a lower risk of developing it: a systematic review with meta-analyses.__________________________86

Abstract ........................................................................................................................................ 88
Key points ......................................................................................................................................... 89
Introduction ..................................................................................................................................... 90
  Rationale ......................................................................................................................................... 90
  Objectives ...................................................................................................................................... 91
Methods .......................................................................................................................................... 92
  Search strategy and selection criteria .......................................................................................... 92
  Study selection ............................................................................................................................... 93
  Description of mild cognitive impairment and dementia ............................................................ 95
  Data extraction and risk of bias ..................................................................................................... 95
  Data analysis ................................................................................................................................. 96
Results ............................................................................................................................................ 99
  Data collection process ................................................................................................................ 99
  Study characteristics .................................................................................................................... 99
  Risk of bias for cross-sectional studies ....................................................................................... 108
  Risk of bias for longitudinal studies .......................................................................................... 109
  Age at Alzheimer’s disease symptom onset ............................................................................... 110
  Age at MCI, dementia, and AD diagnosis .................................................................................... 111
  Subgroup Analysis: Type of Diagnosis (Dementia vs. AD) ......................................................... 112
  Subgroup analysis: immigration status (adjusted vs. did not adjust for immigration) .............. 113
  Disease severity ............................................................................................................................. 114
  Meta-analysis of prospective longitudinal studies: Risk of dementia .......................................... 115
  Small-study effects ....................................................................................................................... 116
Discussion .................................................................................................................................... 118
  Mild cognitive impairment ........................................................................................................... 118
  Age of AD Symptom Onset ......................................................................................................... 119
  Age of Dementia and AD Diagnosis ............................................................................................ 119
  Risk of dementia .......................................................................................................................... 121
  Possible mechanisms and disease severity at dementia diagnosis ............................................ 121
  Sources of uncertainty and risk of bias in cross-sectional and prospective longitudinal studies .......................................................................................................................................................................................................................................................... 124
  Small-study effects ....................................................................................................................... 126
  Limitations ................................................................................................................................... 127
  Strengths ....................................................................................................................................... 129
  Suggestions for future research .................................................................................................... 129
  Implications and conclusion ......................................................................................................... 131
Conflict of Interest .......................................................................................................................... 132
References ...................................................................................................................................... 133

Between-chapter (3-4) rationale ___________________________________________________________145

Chapter 4 The bilingual brain in healthy and neuropathological aging: a systematic review and bran mapping of neuroimaging studies. _____________________________________________146

Highlights ..................................................................................................................................... 147
Abstract ....................................................................................................................................... 148
Introduction ................................................................................................................................... 149
  Objectives .................................................................................................................................... 152
Methods ........................................................................................................................................ 154
  Search strategy ............................................................................................................................ 154
Abstract

Background

Method

Conclusion

Keywords

Administrative information

Introduction

Background and rationale

Bilingualism and foreigner-language learning

Mechanisms of action in bilingualism

Physical activity

Objective

Trial design

Methods: participants, interventions, and outcomes

Study settings

Participants

Inclusion criteria

Exclusion criteria

Who will take informed consent?

Interventions

Explanation for the choice of comparators

Intervention description

Language

Physical activity

Active control

Criteria for discontinuing or modifying allocated interventions

Strategies to improve adherence to interventions

Relevant concomitant care permitted or prohibited during the trial

Provisions for post-trial care

Outcomes

Primary outcome

Secondary outcome

Demographic factors

Neuropsychiatric symptoms

Participant timeline

A description of the timeline content is presented in Table 2.

Assessment of eligibility

Baseline

Post-intervention testing

Sample size

Recruitment

Assignment of interventions

Sequence generation

Concealment mechanism

Implementation
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection, management, and analysis</td>
<td>234</td>
</tr>
<tr>
<td>Plans for assessment and collection of outcomes [18a]</td>
<td>234</td>
</tr>
<tr>
<td>Plans to promote participant retention and complete follow-up [18b]</td>
<td>234</td>
</tr>
<tr>
<td>Data management [19]</td>
<td>235</td>
</tr>
<tr>
<td>Statistical methods</td>
<td>235</td>
</tr>
<tr>
<td>Statistical methods for primary and secondary outcomes [20a]</td>
<td>235</td>
</tr>
<tr>
<td>Methods in analysis to handle protocol non-adherence and any statistical methods to handle missing data [20c]</td>
<td>236</td>
</tr>
<tr>
<td>Discussion</td>
<td>237</td>
</tr>
<tr>
<td>References</td>
<td>239</td>
</tr>
<tr>
<td>Chapter 7 General Discussion</td>
<td>247</td>
</tr>
<tr>
<td>Limitations of this thesis</td>
<td>254</td>
</tr>
<tr>
<td>Suggestions for future research</td>
<td>255</td>
</tr>
<tr>
<td>Clinical significance</td>
<td>256</td>
</tr>
<tr>
<td>Conclusion</td>
<td>256</td>
</tr>
<tr>
<td>References</td>
<td>259</td>
</tr>
</tbody>
</table>
List of published manuscripts


# List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aβ</td>
<td>Beta-amyloid</td>
</tr>
<tr>
<td>B (PiB)</td>
<td>Pittsburgh compound B</td>
</tr>
<tr>
<td>AD</td>
<td>Alzheimer’s disease</td>
</tr>
<tr>
<td>ADAS-cog</td>
<td>Alzheimer's Disease Assessment Scale-Cognitive Subscale</td>
</tr>
<tr>
<td>APOE</td>
<td>Apolipoprotein E</td>
</tr>
<tr>
<td>BDNF</td>
<td>Brain-derived neurotrophic factor</td>
</tr>
<tr>
<td>BL</td>
<td>Bilingual(s)</td>
</tr>
<tr>
<td>CAMCOG</td>
<td>Cambridge Cognitive Examination</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval(s)</td>
</tr>
<tr>
<td>CRF</td>
<td>Cardiorespiratory fitness</td>
</tr>
<tr>
<td>CSF</td>
<td>Cerebrospinal fluid</td>
</tr>
<tr>
<td>CT</td>
<td>Computed tomography</td>
</tr>
<tr>
<td>CVD</td>
<td>Cardiovascular disease</td>
</tr>
<tr>
<td>MCI</td>
<td>Mild cognitive impairment</td>
</tr>
<tr>
<td>MD</td>
<td>Mean difference</td>
</tr>
<tr>
<td>ML</td>
<td>Monolingual(s)</td>
</tr>
<tr>
<td>MMSE</td>
<td>Mini-Mental State Examination</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>NHANES</td>
<td>National Health and Nutrition Examination Survey</td>
</tr>
<tr>
<td>NINCDS-ADRDA</td>
<td>National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>PA</td>
<td>Physical activity</td>
</tr>
<tr>
<td>PET</td>
<td>Positron emission tomography</td>
</tr>
<tr>
<td>PI</td>
<td>Prediction interval(s)</td>
</tr>
<tr>
<td>PiB-PET</td>
<td>$^{11}$C-labeled Pittsburgh compound B-positron emission tomography</td>
</tr>
<tr>
<td>SCD</td>
<td>Subjective cognitive decline</td>
</tr>
<tr>
<td>SMC</td>
<td>Subjective memory complaints</td>
</tr>
<tr>
<td>SPECT</td>
<td>Single-photon emission computed tomography</td>
</tr>
<tr>
<td>T2DM</td>
<td>Type 2 diabetes mellitus</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
Chapter 1  Introduction
Background

An estimated 43.8 million individuals had dementia in 2016 (Nichols et al., 2019) and projections indicate that by 2050 there will be more than 115 million individuals living with this disease (Prince et al., 2013). The global economic cost of dementia is estimated to surpass US$2 trillion by the year 2030 (Wimo et al., 2017). Dementia is an umbrella term that describes a range of neurodegenerative diseases affecting cognition; the most common cause is Alzheimer’s disease (AD; Boyle et al., 2018; Schneider et al., 2007), which accounts for approximately 70% of all dementia cases (Graham, Bonito-Oliva, & Sakmar, 2017). The clinical phenotype of AD is marked by a progressive and irreversible decline in thinking abilities such as episodic memory (the ability to learn new information) in the first instance and executive functions, language, and praxis in the latter stages of the disease (Herrup, 2015). Neuropsychiatrist symptoms including depression and anxiety are also common in AD (Echávarri et al., 2013; Jacobs, Riphagen, Ramakers, & Verhey, 2019). Ultimately, AD will lead to total dependence on a caregiver and lastly, death due to secondary causes such as pneumonia (McKhann et al., 2011). Despite decades of research, however, there is still no cure or treatment for AD (Wortmann, 2012).

Since AD is not part of the normal aging process (McKhann et al., 2011) and its aetiology and pathogenic course are largely driven by lifestyle behaviours, identifying modifiable risk factors that are associated with developing AD has become a global public health priority (Wortmann, 2012). While there is a clear link between certain modifiable risk factors including low levels of education, smoking, and social isolation among others with the risk of developing AD (Livingston et al., 2017), other lifestyle behaviours such as physical inactivity [not meeting physical activity (PA) recommendations for public health] or monolingualism...
(speaking only one language) have received considerable scientific attention but their link with AD is still equivocal (Gold, 2015; Gold, Johnson, & Powell, 2013; Kivimäki et al., 2019).

Prolonged physical inactivity is a serious public health concern contributing to increased risk of obesity, type II diabetes mellitus, and metabolic syndrome, each of which can independently elevate the risk of developing AD (Bellou, Belbasis, Tzoulaki, & Evangelou, 2018; Kivimäki et al., 2019). By contrast, increasing PA is associated with improved health outcomes from preventing the developing of chronic cardiometabolic conditions such as cardiovascular disease (Wahid et al., 2016), and cancer (Rezende et al., 2018) to decreasing the risk of mortality (Ekelund et al., 2016). Individuals with low-moderate or high-intensity PA relative to those who were sedentary (e.g. sitting for a prolonged time) had a 35% decreased risk of cognitive impairment later in life (Sofi et al., 2011) and PA was associated with a 45% decreased AD risk (Hamer & Chida, 2009). However, epidemiological studies tended to use self-report measures of PA and did not generally differentiate between the different type of PA such as walking, gardening, etc. (Hamer & Chida, 2009; Sofi et al., 2011). Also, questionnaires are susceptible to recall bias which might have influenced the magnitude of the relationship between and PA the risk of AD (Sallis & Saelens, 2000). Evidence gathered with accelerometers, which measure PA objectively and do not rely on participants' need to assess and recall their past PA levels, nonetheless support the epidemiological evidence suggesting that high PA is associated with a lower risk of AD (Buchman et al., 2012). However, a recent large individual-level meta-analysis including 400,000 European and North American participants concluded that targeting physical inactivity alone may have limited effects in preventing AD (Kivimäki et al., 2019).
Nevertheless, even modest departures from physical inactivity can have notable implications for improving a wide range of health-related outcomes (Sofi et al., 2011; Wahid et al., 2016).

What is more, randomized-controlled trials (RCTs) testing the impact of increasing PA on cognition in individuals with AD have shown little translation impact. Typically, increasing the intensity and frequency of PA can improve cognition in individuals with mild cognitive impairment a prodromal phase of AD (Öhman, Savikko, Strandberg, & Pitkälä, 2014), but not in those who had already reached a diagnosis of dementia (Öhman et al., 2014). Although others showed that increasing PA might improve cognition in people with dementia (Forbes, Forbes, Blake, Thiessen, & Forbes, 2015). However, the magnitude of the observed improvement was only modest and therefore likely carried limited translational impact (Forbes et al., 2015). The majority of studies were also fairly short (no more than three months), did not report dropout rate, and several studies were statistically underpowered (Öhman et al., 2014). Low statistical power can increase the risk of Type I error and inflate the effect size (Ioannidis, 2005, 2008). Therefore, while increasing PA levels may be to some extent beneficial in some individuals (Forbes et al., 2015), it may carry limited clinical relevance in those with dementia who have accumulated a greater degree of neuropathology (Öhman et al., 2014; Sperling, Jack, & Aisen, 2011). Increasing PA in individuals who have not yet reached such an advanced degree of neuropathology to warrant a diagnosis of AD, but are nonetheless at a higher risk of developing it, may carry greater clinical utility (Frank Jessen et al., 2014; Sperling, Jack, et al., 2011).

Approximately 50% of individuals aged 60 years and above experience subjective cognitive decline (SCD), which is a subjective feeling of diminishing cognitive abilities whose decline cannot be detected with objective measures of cognition (F. Jessen et al., 2014). The
relationship between SCD and memory performance in objective measures is weak (Crumley, Stetler, & Horhota, 2014). While several possible explanations may contribute to SCD including the presence of depressive and anxiety-related symptoms (Balash et al., 2013; Molinuevo et al., 2017), the experience of SCD may signal the earliest phase of AD (F. Jessen et al., 2014; Slot et al., 2019). This is because, individuals with SCD are more likely to convert to AD than those without SCD, which places them at a greater risk of developing AD (Slot et al., 2019). As such, being at the earliest phase of AD, individuals with SCD may not have yet accumulated substantial neuropathology to exclude the possibility of slowing down the future expected cognitive decline or delay the clinical manifestation of AD. Interestingly, one RCT recruiting participants with memory complaints (a feature of SCD) showed that while cognitive performance decline by 1.04 points in the Alzheimer Disease Assessment Scale–Cognitive Subscale in the control group, those who increased their PA (walking more) across six months experienced an improvement of 0.26 points at the end of the trial (Lautenschlager et al., 2008). Another study recruiting cognitively intact individuals who were nonetheless at a higher risk of developing AD showed that a multicomponent behavioral strategy improved cognitive performance in the experimental relative to the control group after two years (Ngandu et al., 2015). These studies suggest targeting behavioral risk factors such as PA in individuals at a higher risk of AD such as those with SCD phase, may not only slow down the expected cognitive decline but also improve cognitive performance in this target group.

Another potentially modifiable risk factor for dementia is bilingualism, which refers to the ability to speak two or more languages (Luk & Bialystok, 2013). Several studies have shown that bilinguals, as opposed to monolinguals, are diagnosed with dementia and AD four to five years later (Alladi et al., 2013; Bialystok, Craik, Binns, Ossher, & Freedman, 2014;
Bialystok, Craik, & Freedman, 2007). Moreover, these studies have shown that while bilinguals were older at the time of dementia diagnosis, they demonstrated a similar degree of global cognitive impairment as monolinguals (Alladi et al., 2013; Bialystok et al., 2014; Bialystok et al., 2007). This might mean that bilingualism may help maintain a working level of cognition despite the progression of AD-related neurodegenerative processes (Gold et al., 2013). Bilingualism may render the brain more resilient against neurodegeneration by strengthening frontostriatal and frontoparietal circuits which are severely affected in AD and, in turn, delay its clinical manifestation (Gold, 2015; Gold et al., 2013). While these findings are promising, results are based on studies with several methodological limitations. For example, participants’ language profiles were generally poorly measured and not all studies controlled for other important lifestyle factors such as PA that might have explained the observed delay in dementia diagnosis among bilinguals. Furthermore, prospective longitudinal studies, however, have not shown that bilingualism can decrease the risk of dementia (Mukadam, Sommerlad, & Livingston, 2017).

While studies in exercise science and bilingualism research show that increasing PA may lower the risk of AD and bilingualism may delay AD diagnosis, there remain several unanswered questions. The analytic samples of previous studies assessing the impact of increasing PA on cognition included participants with SCD and mild cognitive impairment together and did not use a validated measure of SCD (Lautenschlager et al., 2008) or did not measure SCD altogether (Ngandu et al., 2015). To our knowledge, there is limited evidence applying specific criteria for recruiting individuals with SCD testing the effect of increasing PA on cognition (Andrieu et al., 2017; Molinuevo et al., 2017; Sperling, Aisen, et al., 2011). Therefore, it is not known whether modifying risk factors in people with SCD may be useful in improving cognition and as such, trials targeting risk factors in this cohort are urgently
needed. What is critical, however, is that increasing PA alone may not be sufficient in protecting against AD (Kivimäki et al., 2019) and since bilingualism may protect against AD, increasing PA in combination with studying a foreign language (Antoniou, Gunasekera, & Wong, 2013; Bialystok, Abutalebi, Bak, Burke, & Kroll, 2016) in individuals with SCD may generate more clinically relevant outcomes than modifying each behaviour in isolation (Kivimäki et al., 2019; National Academies of Sciences & Medicine, 2017).

**Putative underlying mechanisms**

There is evidence showing that increasing PA levels and learning a second language may promote neural reserve, neural compensation, and cognitive reserve. Broadly defined, these terms refer to the brain’s ability to maintain functioning cognition despite presence of an underlying neuropathological disorder such as AD. While brain reserve refers to gross brain tissue including neurons and synapses and how much brain damage an individual may receive before experiencing cognitive impairment, cognitive reserve refers to how flexible and adaptable a person’s thinking ability is in response to age- or neuropathological related decline in cognition (Stern, Barnes, Grady, Jones, & Raz, 2019). Increasing PA may promote the supramolecular mechanisms including the formation of new neurons and synapses (Lista & Sorrentino, 2010), which could result in greater brain reserve. Cognitively stimulating activities such as studying a foreign language may also promote these two processes (Perani & Abutalebi, 2015). These underlying processes may explain why some individuals who report high levels of PA or the ability to speak more than one language, also develop AD at a later age and carry a lower risk of developing AD than those with low PA and monolinguals (Perani & Abutalebi, 2015).
Neurodegenerative disorders including AD severely affect brain regions such as striatal and frontostriatal areas; atrophy of these areas of the brain is common in AD (Bertoux, O’Callaghan, Flanagan, Hodges, & Hornberger, 2015). Engaging in mentally stimulating activities may be useful to counteract the neurodegenerative effects on these brain regions (Gold, 2015). For example, observational studies show that bilinguals relative to monolinguals of similar ages demonstrate greater striatal and frontostriatal areas and that studying a foreign language may increase volume in the same regions (Li, Legault, & Litcofsky, 2014). This evidence suggests that PA and bilingualism could strengthen the same brain areas that are typically affected by AD. The underlying mechanisms responsible for the observed delays in the onset and in lowering the risk of developing AD, particularly when combining PA with bilingualism, have not been systematically nor empirically explored.

**Purpose of this thesis**

One of the aims of this thesis was to determine whether increasing PA might be an effective intervention in improving cognitive performance among individuals at a higher risk of developing AD such as those with SCD. Another aim was to determine whether bilingualism relative to monolingualism can delay the onset and lower the risk of developing AD. As such, the purpose of this thesis was to i) review the available evidence exploring the link between PA and dementia by conducting a narrative review of the literature (Chapter 2), ii) conduct a systematic review of the literature exploring the link between bilingualism and dementia onset and risk (Chapter 3), iii) conduct a systematic review of literature exploring the underlying link between bilingualism and the brain (Chapter 4), iv) conduct a cross-sectional epidemiological study investigating the link between language acculturation and cognition in older individuals from ethnic minorities living in the United States (Chapter 5), and v) the evidence gathered from the previous chapters resulted in the development of a proposal for
an RCT investigating the effects of increasing PA levels and studying a foreign language among English senior monolinguals at a higher risk of AD. We chose to conduct a narrative review of the literature (Chapter 2) because several systematic reviews in this field answering specific questions had already been published and due to the multiple components of exercise (e.g., the intensity, duration, frequency, and type) and difficulty in measuring different types of habitual physical activity. We chose to conduct systematic reviews in Chapters 3 and 4 because no systematic reviews existed addressing our research questions assessing the role of bilingualism in dementia risk and the brain at the start of this project.

Therefore, we hypothesized that i) increasing the intensity and frequency of PA particularly at the preclinical phase of AD is associated with improvements in cognitive performance (Chapter 2), ii) bilingualism relative to monolingualism will be associated with a delayed onset of dementia (Chapter 3), iii) bilingualism will be associated with stronger frontostriatal and frontoparietal circuits in the brain of bilinguals relative to monolinguals (Chapter 4), and iv) higher levels of language acculturation are associated with greater cognitive performance in older individuals from ethnic minorities living in the United States (Chapter 5).
References


Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), e124. doi:10.1371/journal.pmed.0020124


Jessen, F., Amariglio, R. E., Van Boxtel, M., Breteler, M., Ceccaldi, M., Chételat, G., . . .


doi:10.1016/j.jalz.2016.09.012


Authors: Stefano Brini1,2, Hamid R. Sohrabi3,4,6, Jeremiah J. Peiffer1, Mira Karrasch5, Heikki Hämäläinen2, Ralph N. Martins3,4,6, Timothy J. Fairchild1

1. School of Psychology and Exercise Science, Murdoch University, Perth, Western Australia, Australia
2. Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland
3. School of Medical and Health Sciences, Edith Cowan University, Perth, Western Australia, Australia
4. Department of Biomedical Sciences, Macquarie University, New South Wales, Australia
5. Department of Psychology, Åbo Akademi University, Turku, Finland
6. Australia Alzheimer’s Research Foundation, Perth, Western Australia, Australia
Abstract

A large body of epidemiological and experimental data exploring the relationship between physical activity (PA) and Alzheimer’s disease (AD) are now available. Despite observational evidence supporting a role for PA in delaying the onset of AD, randomized-controlled trials have reported mixed findings; this is likely due to the heterogeneity in study cohorts, outcome measures and the adopted PA intervention. The primary objective of this narrative review is to evaluate the extant evidence on the relationship between PA, cognitive decline and AD in older populations. The interaction between PA and the putative mechanisms underlying AD progression, including genetic factors and amyloid beta levels will be explored. In this context, particular attention will be given to studies assessing PA in the early clinical and preclinical, asymptomatic stages of AD. Based on current evidence, clinical considerations for implementation of exercise-based interventions are discussed, along with limitations of previous research and directions for future studies.
Introduction

Dementia is an umbrella term encompassing many neurodegenerative disorders such as Alzheimer’s Disease (AD). The global estimated costs of dementia in 2015 were US$818 billion annually and these are expected to extend to US$2 trillion by the year 2030 [1]. Without a treatment or cure, the worldwide estimated prevalence of dementia will be 115.5 million people by the year 2050 [2]. Consequently, the World Health Organization (WHO) has labelled dementia a public health priority [3].

The most common form of dementia is AD, which accounts for 60-70% of all dementia cases and for which no treatment or cure currently exist [4, 5]. While most individuals experience subtle decreases in cognitive performance with advancing age [6], the pronounced cognitive deterioration resulting from AD is not part of the normal ageing process [7]. Cognitive decline associated with AD can be divided into three stages: a preclinical, asymptomatic stage referred to as subjective cognitive decline (SCD; [8, 9]), a symptomatic, preclinical stage of mild cognitive impairment (MCI; [10]), and finally AD, which is characterised by cognitive impairment and significant dysfunction in daily living activities [7]. Importantly, SCD or MCI are not always considered as part of the preclinical phase of AD, since many individuals with SCD or MCI do not always progress to AD [11].

Research has traditionally focused on tertiary preventions aiming to alleviate or slow the pathogenic processes in the symptomatic preclinical and clinical stages of AD. However interventions in these stages have, to a large degree, been unsuccessful in achieving these aims [4, 7, 12]. This is likely due to the presence of advanced neuronal damage in individuals within symptomatic stages, and this neuronal damage has proven difficult to reverse with current interventions [12, 13]. Administering disease-modifying interventions such as increasing levels of physical activity (PA) in the asymptomatic stage (i.e. SCD) may be an important strategy in delaying the onset or slowing the progression of AD pathogenesis.
before the neurodegeneration becomes irreversible [9, 13, 14]. Specifically, a 5-year delay in the onset of AD, could reduce the number of AD patients by 57% and the associated economic costs by half [15].

The pathophysiology of AD

AD is a complex neurodegenerative disorder with a clinical phenotype characterized by insidious and progressive decline in episodic memory, attention, executive functions, language, and praxis, followed by loss of motor control, resulting in complete dependence and ultimately death [16]. A definitive diagnosis for AD can only be achieved histopathologically at post-mortem. For many years the amyloid cascade hypothesis was the prevailing explanation for the pathogenesis of AD [17]. This hypothesis arose from histopathological observations showing accumulation of intra- and extracellular misfolded proteins called amyloids, which contain phosphorylated tau and amyloid plaques, comprising of beta amyloid (Aβ) peptides. The accumulation of these Aβ peptides was observed to exacerbate synaptic dysfunction resulting in tau hyper-phosphorylation which aggregate and deposit intracellurally, ultimately leading to synaptic loss and neuronal death [18]. This process can be confirmed during post-mortem examination of AD brain tissues which reveals microscopic lesions such as senile plaques and neurofibrillary tangles across the central nervous system, particularly in the cerebral cortex. Gross inspection of the AD brain post-mortem also reveals normally distributed and hemi-symmetrical atrophy of the neocortex suggesting neurodegeneration [17].

This cascade of neuropathological events arising from the abnormal accumulation of the Aβ peptide is thought to be a central event in AD pathophysiology, indicating that selectively targeting Aβ with pharmacotherapy would successfully treat AD [18]. Consequently, the amyloid cascade hypothesis was used as a benchmark from which
therapeutic interventions were developed [17], many of which however, have not been successful [4, 19]. Indeed, it now appears the AD pathophysiology cannot be reduced to a single aetiology as had been previously hypothesized; rather, it likely arises from several toxic pathogenic processes [5, 16].

**Risk Factors for AD**

*The relationship between subjective cognitive decline and AD*

Subjective cognitive decline (SCD) and subjective memory complaints (SMC) are used interchangeably in the literature. While SMC remains the prominent feature during the asymptomatic, preclinical stage of AD, researchers have preferentially adopted SCD when referring to this stage. The prevalence of SCD among individuals aged 60 and over ranges between 25% and 50% [20], with the prevalence increasing concurrently with age [21]. A person with SCD experiences subjective impairment in memory and cognitive functions, but these subjective impairments cannot be detected by objective measures [22]. The conversion rate from SCD to MCI in studies conducted over four years is 24.4%, while the conversion rate to dementia over this time-period is 10.9% compared to 4.6% in individuals without SCD [23]. Overall, individuals with SCD (but without objectively measurable complaints) have twice the risk of developing dementia than those without SCD [23]. Structural Magnetic Resonance Imaging has also indicated that participants with SCD have a significantly smaller mean left-hippocampal volume \( [n = 20; 2.0 \ (0.4) \ cm^3] \) compared to control participants \( [n = 28; 2.3 \ (0.4) \ cm^3] \) without memory complaints [24]. Consequently, SCD is an important risk indicator in the natural history across the AD spectrum [25].

There is experimental evidence indicating that PA interventions may be more successful when delivered in the early preclinical (Table 1 [26-28]) and early clinical
populations [29], as opposed to the latter clinical phase of AD (Table 2; [30]). For example, in a randomized-controlled trial, Lautenschlager et al. [26] tested the effects of increased PA levels on cognition among individuals with SCD and MCI and found a significant improvement in cognition as measured by the Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-cog). This difference was still detectable at the 18-month follow-up. Another trial comparing an active control (regular health advice) and a multi-domain intervention comprising diet, exercise, and cognitive training among AD at-risk individuals across two years, revealed that the multidomain intervention improved memory and processing speed tasks [28]. Similarly, Shah et al. [27] demonstrated that in older individuals at risk of AD but without SCD or MCI, a combination of PA and computer-based cognitive training improved cognition and cerebral glucose metabolism, which is a marker of cognitive performance, more than each intervention alone. Finally, that early intervention is likely associated with improved cognitive outcomes is supported by systematic reviews that PA interventions seem to be more successful when delivered during the MCI stage rather than at the AD stage (Table 2 and 3; i.e., [29]).

**Cardiovascular diseases and low PA as risk factors for AD**

Several modifiable, cardiovascular risk factors have been associated with increased risk of AD including obesity and obesity-related diseases such as Type 2 Diabetes Mellitus (T2DM) and Cardiovascular Disease (CVD). While individuals with obesity have a higher risk of AD [31-33], the magnitude of the association, independently of other risk factors, remains a matter of debate. Nevertheless, two systematic reviews with meta-analyses have since confirmed overweight and obesity in mid-life, as independent risk factors for AD [34, 35], while obesity later in life has been associated with a lower risk of AD [36]. Although, the strength of these relationships were found to be less than the relationship with
Apolipoprotein E (APOE), which is a major genetic risk factor for AD [35]. T2DM has also been associated with poorer cognitive performance in working memory, executive functions, and attention in older adults [37, 38], as well as immediate and delayed verbal recall, and verbal fluency among elderly women [39, 40] which are cognitive abilities that are strongly affected by AD. Moreover, improving glycaemic control has been shown to improve cognitive performance [41]. That T2DM has been found to accelerate cognitive decline and increase AD risk independently of other comorbid factors (e.g. obesity) [35, 42] is therefore unsurprising. Similarly, individuals with CVD have an elevated risk for AD [43-45]. The independent and direct relationship between CVD and AD is likely related to hypoperfusion and microemboli which may present in CVD and can accelerate the pathogenesis of AD [45, 46]. Also, reduced PA and sedentary behaviour are strongly associated with obesity [47], T2DM and CVD [44, 47, 48], both of which increase the risk of developing AD [49-51]. Indeed, a sedentary lifestyle is associated with increased AD risk [52] with physical inactivity being the single largest modifiable risk factor for AD, while increased PA is considered protective against AD [53].
### Table 0-1 Experimental trials investigating the effects of PA programs on persons with increased risk of dementia or SCD/MCI

<table>
<thead>
<tr>
<th>Study</th>
<th>N; age</th>
<th>Type of PA program</th>
<th>Neurological condition</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lautenschlager</td>
<td>170; PA n = 68.6 (8.7), control n = 68.7 (8.5)</td>
<td>Three 50 minute home-based sessions/week of moderate intensity PA</td>
<td>SCD; MCI</td>
<td>i) $\text{Tx} \uparrow 0.26 \text{ADAS-Cog}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ii) $\text{Cx} \downarrow 1.04 \text{ADAS-Cog}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>iii) Absolute difference (Tx and Cx): $-1.3$ points</td>
</tr>
<tr>
<td>Shah</td>
<td>224; 67.6 (5.42)</td>
<td>PA: 48 walking sessions 60 min/day, 3 days/week and 32 resistance training sessions 40/day, 2 days/week; CS: 40 sessions at 60min/day for 5 days/week for the auditory-based Brain Fitness Program and the visual-based Insight Program; PA+CS: both PA and CS sessions</td>
<td>Elderly individuals at a higher risk of AD</td>
<td>i) NS $\uparrow$ in cognition in the PA or CS group</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ii) Significant $\uparrow$ in the RAVLT: LTDR in the PA+CS group relative to controls</td>
</tr>
<tr>
<td>Ngandu</td>
<td>1260; 69.4 (4.7) years</td>
<td>Progressive muscle strength training (1–3 times/week), aerobic exercise (2–5 times/week), and exercises to enhance postural balance</td>
<td>Elderly individuals at a higher risk of AD</td>
<td>i) $\text{Tx Z scores} \uparrow 0.20 \text{NTB}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ii) $\text{Cx Z scores} \uparrow 0.16 \text{NTB}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>iii) Between-group difference (Tx and Cx): 0.022</td>
</tr>
</tbody>
</table>

SCD: subjective cognitive decline; MCI: mild cognitive impairment; CS: Cognitive stimulation; RAVLT: LTDR: Rey Auditory Verbal Learning Test; Long-Term Delayed Recall; NS: no significant

### Table 0-2 Systematic reviews of intervention studies investigating the effects of increased PA on cognitive decline/dementia

<table>
<thead>
<tr>
<th>Study</th>
<th>Design; N</th>
<th>Assessment of PA</th>
<th>Neurological condition</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Study</th>
<th>Number of RCTs</th>
<th>Number of Participants</th>
<th>Intervention</th>
<th>Condition</th>
<th>Findings</th>
</tr>
</thead>
</table>
| Öhman (2014) [29]             | 22             | 1699                   | Walking, Tai Chi, ergocycling, and strength training lasting from 6 to 12 months | MCI; dementia | i) PA ↑ global cognition, executive function, attention, and delayed recall in MCI subjects  
                          |                |                        |                                                                               |           | ii) NS effects on cognition in subjects with dementia                     |
| Forbes (2015) [30]            | 17             | 1067                   | Tx: any combination of aerobic, strength, or balance training vs. Cx: usual care, or social contact/activities; PA programs; frequency (range: 2 to 5 times/week, to daily, from 20 to 75 mins/session, from 2 weeks to 18 months | Dementia | i) NS effects on cognitive performance                                      |
| Groot (2016) [74]             | 18             | 802                    | i) Aerobic only, ii) non-aerobic, and iii) combined aerobic with non-aerobic exercise; high and low frequency | Dementia; AD | i) PA ↑ cognitive function in AD and non-AD related dementia                
                          |                |                        |                                                                               |           | ii) Greater ↑ for combined PA than aerobic-only                            
                          |                |                        |                                                                               |           | iii) NS effect for non-aerobic PA;                                      
                          |                |                        |                                                                               |           | iv) Both high and low frequency ↑ cognition                               |
| van Uffelen (2008) [154]      | 8              | 543                    | Aerobic exercise only, strength exercise, strength and balance exercise, all-round exercise including aerobic, strength, balance, and flexibility training, from 6 to 52 | CD; dementia | i) Significant ↑ in general cognitive function, executive functions, and memory |
weeks of 20 to 65 min sessions at a
frequency of 1 to 3 times/week

RCT: randomized-controlled trial; PD: Parkinson’s Disease

Table 0-3 Systematic reviews on epidemiological studies investigating the relationship between PA and cognitive decline/impairment/dementia

<table>
<thead>
<tr>
<th>Study</th>
<th>Design; N</th>
<th>Assessment of PA</th>
<th>Neurological condition</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sofi (2010)</td>
<td>15 prospective studies (12 cohorts); 33816 without CI at baseline, followed for 1–12 years</td>
<td>Self-report PA</td>
<td>CD or CI</td>
<td>i) PA ↓ CI risk by 35% versus sedentary individuals</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ii) High-intensity PA by 38% versus sedentary individuals</td>
</tr>
<tr>
<td>Hamer (2009)</td>
<td>16 prospective studies, 163797 without dementia at baseline</td>
<td>Self-report PA</td>
<td>Dementia; AD; PD</td>
<td>i) PA associated 28% ↓ dementia risk and 45% AD risk</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ii) NS association between PA and PD</td>
</tr>
</tbody>
</table>

NS: no significant
PA, cognition and risk of AD: Epidemiological findings

The epidemiological evidence (Table 4) suggests a strong association between moderate or high levels of PA and improved cognitive performance later in life [54], even after adjusting for factors such as sex, age, baseline cognitive status, and depression. Women who reported being physically active at different ages (30, 50, late life) had a reduced likelihood of developing cognitive impairment later in life compared to physically inactive women, particularly when they engaged in higher PA in their earlier years [55]. A meta-analysis of prospective studies confirmed a consistent protection of PA against cognitive impairment later in life and this protection occurred at all levels of PA (Table 3; [56]). The beneficial effects of PA have also been extended to a reduced risk of all-cause dementia and AD [57]; a finding confirmed in a systematic review of prospective studies demonstrating a link between increased PA and a lower risk of dementia and AD [58]. Of particular note is one study adopting actigraphy to objectively measure PA, as opposed to self-reported PA, which found that high total daily PA was associated with lower risk of AD [59]. Compared to self-report, actigraphy is a valid and objective measure of PA in aging [60, 61] and does not rely on participants’ subjective recall of previous PA levels, which is often challenging and inaccurate [62].

While the association between PA and lower risk of AD is well established, this effect might be mediated or moderated by additional factors not easily accounted for such as social engagement, educational level, depression, cognitive activity, and the number of different types of activities (as opposed to total duration alone) performed [56, 63]. Indeed, some studies have found that a previously significant relationship between PA and cognitive impairment became non-significant after adjusting for such factors [64-67]. For example, Podewils et al. [67] found that individuals engaging in a greater number of activities (≥4) had a significant lower relative risk of dementia and AD than those engaging in 0-1 activities;
individuals with the highest quartile of energy expenditure were not protected from relative risk of dementia after multivariate adjustment. Overall, the epidemiological data suggest a strong association between higher PA levels and a lower risk of cognitive impairment and AD later in life [55, 56] even after multivariate adjustment. However, due to the clustering of multiple risk factors with lower levels of PA, the magnitude of the independent association of PA with cognitive impairment and AD remains equivocal.
<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Assessment of PA</th>
<th>Neurological condition</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etgen (2010) [54]</td>
<td>N = 3903</td>
<td>No activity, moderate activity (&lt;3 times/week), and high activity (≥3 times/week)</td>
<td>Incident CI at two-year follow-up</td>
<td>Baseline moderate or high PA was associated with ↓ incident CI risk, compared to no PA</td>
</tr>
<tr>
<td>Middleton (2010) [55]</td>
<td>N = 9344</td>
<td>Current (late life) yearly frequencies of low (e.g. walking or gardening), moderate (e.g. dancing or tennis), or high (jogging or skiing) intensity PA; modified Paffenbarger questionnaire</td>
<td>CI</td>
<td>PA during youth, age 30 and 50, and late life was associated with a ↓ likelihood of CI later in life compared to physical inactivity</td>
</tr>
<tr>
<td>Laurin (2001) [57]</td>
<td>N = 6434</td>
<td>Combination of two questions from a risk factor questionnaire: frequency (low, moderate, or high: ≥3 times per week, weekly, or less than weekly) and intensity (more vigorous, equal to, or less vigorous than walking)</td>
<td>Incident CI and dementia</td>
<td>High PA was associated with ↓ risk of CI, AD, and dementia of any type</td>
</tr>
<tr>
<td>Buchman (2012) [59]</td>
<td>N = 716</td>
<td>Actigraphy for 10 days (total daily PA)</td>
<td>AD at four-year follow-up</td>
<td>Lowest 10th PA percentile compared to highest 90th was associated with ↑ AD risk</td>
</tr>
<tr>
<td>Sturman (2005) [64]</td>
<td>N = 4055</td>
<td>US Health Interview Survey (walking for exercise, jogging, yard</td>
<td>CD</td>
<td>Each additional hour of PA/week was associated with ↓ rate of CD</td>
</tr>
</tbody>
</table>
work, etc.); PA was measured as hours/week

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Activities/Activities Measured</th>
<th>Disease</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang (2002)</td>
<td>776</td>
<td>Mental, physical, social, productive, and recreational activities; social and leisure activity data were gathered during personal interview with trained nurses</td>
<td>Dementia</td>
<td>i) Mental, social, or productive activity associated with ↓ dementia risk</td>
</tr>
<tr>
<td>Niti (2008)</td>
<td>1635</td>
<td>Social activities (attending church/temple/mosque), productive activities (shopping, hobbies), PA activities (walking, jogging, sports)</td>
<td>CD</td>
<td>i) Higher LA was associated with ↓ CD risk more than PA or SA</td>
</tr>
<tr>
<td>Podewils (2005)</td>
<td>5888</td>
<td>MLTAQ (walking, household chores, mowing, raking, gardening, etc.)</td>
<td>Dementia</td>
<td>i) Highest PA quartile was associated with ↓ dementia risk versus lowest PA quartile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ii) Engaging in ≥4 activities was associated with ↓ dementia risk than 0–1 activity</td>
</tr>
</tbody>
</table>

CI: cognitive impairment; CD: cognitive decline; LA: leisure activity; SA: social activity; MLTAQ: Minnesota Leisure Time Activity Questionnaire


**PA and cognition in dementia: Experimental findings**

Whether PA interventions improve cognition in clinical populations with existing AD is also contentious [29, 30]. Additional details regarding studies’ methodologies and findings can be found in Table 5. A two-point increase in Mini-Mental State Examination (MMSE) scores has been observed following a three-month music-based exercise intervention in a clinical population with moderate dementia, which ranges from 10 to 18 MMSE points [68]. However, when considering a two-point increase from baseline ($M = 12.87; SD = 5.01$) in the experimental group, at the end of the trial participants were still in the moderate phase of dementia, therefore the clinical implications of these findings are likely limited. Using treadmill walking, Arcoverde et al. [69] found that MMSE scores remained similar from baseline to post-intervention, while the control group experienced a decrease in MMSE scores suggesting exercise slowed the rate of cognitive decline in this clinical AD group. Similarly, Venturelli et al. [70] found that aerobic walking significantly slowed cognitive decline when global cognition was measured with the Cambridge Cognitive Examination (CAMCOG) but not with the MMSE in individuals with borderline moderate-severe AD. This is likely explained on the basis that the MMSE has low sensitivity in detecting changes across periods of fewer than six months [71], which could explain why Venturelli et al. [70] observed significant changes in the CAMCOG but not in the MMSE scores. However, aerobic walking was found not to improve global cognition in nursing home residents with moderate dementia, but the intervention in this study lasted only six weeks [72].

There is experimental evidence which also shows that combining aerobic with non-aerobic exercise (i.e. resistance or strength exercises) can improve cognition more than aerobic-exercise alone [73, 74]. For example, global cognition, executive functions, verbal and visual memory were improved more in individuals with dementia combining nine weeks of aerobic (walking) and non-aerobic exercise (strength exercises) than individuals
completing aerobic based exercises only [73]. However, despite similar methodology and good adherence Steinberg et al. [75] showed only modest (non-significant) improvements in global cognition following a multi-modal exercise program by the end of the study [75]. Similarly, Miu et al. [76] did not detect improvements in global cognition following a combination of treadmill, bicycle, and arm ergometry relative to controls, among individuals with dementia. Recently, a meta-analysis revealed that although aerobic-based PA interventions were beneficial in AD patients, the combination of aerobic with non-aerobic exercise produced greater effects on cognition [74]. However, a systematic review by Forbes et al. [30] reported a similar effect size for exercise on cognition as in Groot et al. [74], but concluded that a meta-analysis could not be conducted due to substantial unexplained statistical heterogeneity ($I^2 = 80$).

To date, there is promising evidence supporting a role for exercise in improving cognition in individuals with dementia, but the field does have several studies yielding conflicting results [30]. There appear to be several key elements required for a PA intervention to demonstrate significant and clinically meaningful improvements in cognition among individuals with AD. These include multimodal PA interventions with at least one aerobic component [74] consisting of at least 150 minutes during the week, and which is delivered for longer than 10 weeks [29, 77]. The exercise intervention should be supported by additional PA (incidental) throughout the day to improve global cognition [78]. Moreover, since physical exercise likely affects particular dementia subtypes in different ways, the specific programming (duration, intensity, type, frequency) of exercise is likely to be an important consideration for studies in individuals with existing AD [29, 30, 79].
Table 0-5 Experimental trials investigating the effects of single-intervention PA programs on clinical dementia

<table>
<thead>
<tr>
<th>Study</th>
<th>N; age</th>
<th>Type of PA program</th>
<th>Neurological condition</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Van de Winckel (2004) [68]</td>
<td>PA: $n = 15, 81.33 (4.24);$ control: $n = 10, 81.90 (4.18)$</td>
<td>Daily face-to-face 30min PA sessions focusing on upper and lower body strengthening, balance, trunk movements, and flexibility training, which was supported by music such as folkloric accordion songs including polka, folk, country, and western music</td>
<td>AD; MID</td>
<td>i) $\text{Tx} \uparrow$ from 12.87 to 15.53 MMSE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ii) $\text{Cx} \uparrow$ from 10.80 to 11.00 MMSE</td>
</tr>
<tr>
<td>Arcoverde (2014) [69]</td>
<td>PA: $n = 10, 79 (74.7-82.2);$ control $n = 10, 78.5 (64-81.2)$</td>
<td>PA on a treadmill for 30 mins, twice a week at moderate intensity ($60% \text{VO2max}$) for 3 months</td>
<td>AD</td>
<td>i) $\text{Tx} \uparrow 6.10 (6.7) \text{CAMCOG}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ii) $\text{Cx} \downarrow 6.10 (4.3) \text{CAMCOG}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>iii) NS effects on the MMSE</td>
</tr>
<tr>
<td>Venturelli (2011) [70]</td>
<td>PA: $n = 12, 83 (6);$ control: $n = 12, 85 (5)$</td>
<td>At least 30 mins of moderate aerobic exercise (walking) 4 times/week</td>
<td>AD</td>
<td>i) $\text{Tx}$ showed slower $\downarrow$ (-13%) compared to $\text{Cx}$ (-47%) in MMSE*</td>
</tr>
<tr>
<td>Eggermont (2009) [72]</td>
<td>PA: $n = 51, NA;$ $n = 46, NA$</td>
<td>Walking for 30 mins, 5 days/week, for 6 weeks</td>
<td>Dementia</td>
<td>i) NS effects on cognitive performance</td>
</tr>
</tbody>
</table>

**Experimental trials investigating the effects of multimodal PA programs on clinical dementia**

<table>
<thead>
<tr>
<th>Study</th>
<th>N; age</th>
<th>Type of PA program</th>
<th>Neurological condition</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bossers (2015) [73]</td>
<td>Combined group: $n = 37,$ $85.7 (5.1);$ aerobic group: $n = 36, 85.4 (5.4);$</td>
<td>Combined group: strength exercises (lower-limb strengthening) and aerobic exercise (moderate-high intensity walking); aerobic group received only the aerobic exercise;</td>
<td>Dementia</td>
<td>i) Combined $\text{Tx} \uparrow 2.3 \text{MMSE}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ii) Aerobic $\text{Tx} \uparrow 1 \text{MMSE}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>iii) $\text{Cx} \uparrow 0.72 \text{MMSE}$</td>
</tr>
<tr>
<td>Study</td>
<td>N</td>
<td>Safety</td>
<td>Exercise</td>
<td>Training Protocol</td>
</tr>
<tr>
<td>-------</td>
<td>---</td>
<td>--------</td>
<td>----------</td>
<td>------------------</td>
</tr>
<tr>
<td>Steinberg (2009) [75]</td>
<td>37</td>
<td>74.0 (8.1)</td>
<td>76.5 (3.9)</td>
<td>Combined group = aerobic fitness (brisk walking), strength training (major muscle groups), balance and flexibility training (shifting center of gravity, tandem walks, forward and backward walks, and chair sit to stands); daily exercises for 12 weeks; control group = home safety assessments</td>
</tr>
<tr>
<td>Miu (2008) [76]</td>
<td>Combined group: n = 36, 75 (7); control group: n = 49, 78 (6),</td>
<td>Combined group = treadmill, bicycle and arm ergometry, and ten-minute flexibility training (at the start of each session) for three months, twice/week for 45-60mins</td>
<td>Dementia</td>
<td>i) NS effects on cognitive performance</td>
</tr>
</tbody>
</table>

MID: Multi-infarct dementia; NS: no significant; JTT: Jebsen Total Time; NS: no significant; MMSE: Mini-Mental State Examination
Genetic markers and risk of cognitive decline and AD

Several genetic markers have been associated with increased AD risk including APOE, klotho, WW domain-containing protein 1 (WWC1), and brain derived neurotrophic factor (BDNF). Although klotho (encoding the transmembrane protein klotho; [80]) and WWC1 (encoding the KIBRA protein; [81]) have both been implicated in cognitive impairment, their involvement in cognitive decline and AD development are equivocal [82, 83]. As such, BDNF and APOE have garnered the greatest attention in this field.

The APOE gene is polymorphic, with three major alleles, namely ε2, ε3, and ε4, with a global prevalence of 8.4%, 77.9%, and 13.7%, respectively; the ε4 is ~40% prevalent among individuals with AD [84]. Presence of at least one copy of the ε4 allele (ε2/ε4; ε3/ε4) increases the risk of AD by 2.6 and 3.2 odds ratio respectively, while two copies (ε4/ε4) by 14.9 odds ratio among Caucasian people [84] and presence of one or two APOE ε4 alleles can trigger the onset of AD five years or 10 years earlier, respectively [85]. Therefore, the APOE ε4 is a major non-modifiable risk factor for AD and is associated with a younger age of onset [86]. While the APOE ε4 is a risk factor for CVD [87], coronary heart disease [88], and higher amyloid aggregation which is in itself a risk factor for AD [89], the exact mechanisms explaining the link between APOE ε4 and increased AD risk are still unclear. Of particular relevance to this review, individuals with SCD who are also carriers of the APOE ε4 allele are twice as likely to progress to AD than those without SCD and the APOE ε4 allele, suggesting that both factors produce an additive effect in cognitive decline [90]. Additionally, SCDs who test positive for APOE ε4 allele demonstrate glucose hypo-metabolism in brain areas typically affected by AD such as the parieto-temporal lobe when compared to non-carriers [91].

The BDNF gene encodes the brain-derived neurotrophic factor (BDNF), which is a growth factor in the central nervous system that has been shown to regulate neuronal growth,
promote neuronal survival, and regulate synaptic plasticity [92]. In addition, BDNFs may regulate neuroplastic processes such as long-term potentiation in the hippocampus [93], and as such, BDNFs may play an important role in the pathogenesis of AD [94]. Phillips et al. [95] were the first to show reduced post-mortem in situ expression of BDNF mRNA in the hippocampal formation of nine AD compared to six controls donors. Since then, several studies have shown decreased BDNF serum levels among individuals in the preclinical [96] and clinical phase of AD [97, 98] and that the BDNF levels correlate with the degree of cognitive impairment [99]. A large cross-sectional study including 4463 community-living elderly participants, found that after adjusting for covariates, lower serum BDNF levels were associated with a decline in memory performance and with an elevated risk for MCI [100]. Importantly, circulating BDNF is generally accepted as a suitable surrogate marker of total (central and peripheral) BDNF concentration, however, the short half-life and low blood brain barrier perfusion capacity of BDNF [101] require further investigation.

Apolipoprotein ε4 allele and PA

Given that presence of at least one of the APOE ε4 alleles is associated with poorer performance on cognitive tasks [102] and increased risk of AD [85], researchers have assessed whether cognitive benefits arising from PA are moderated by APOE ε4 allele status. Indeed, evidence indicates that PA levels have a greater effect on cognitive function and future incidence of cognitive decline among individuals who carry at least one copy of the APOE ε4 allele [103-105]; although some researchers have not observed the same relationship [67]. For example, less than an hour/day of PA was associated with an increased risk of cognitive decline (N = 347; odds ratio 2.0, 95% CI: 0.9-4.8) and this effect was stronger in APOE ε4 allele carriers (adjusted odds ratio 3.7, 95% CI: 1.1-12.6) even after adjusting for additional confounders [103]. Similarly, a longitudinal population-based survey
(mean follow-up period: 21 years) found that the beneficial effects of PA on dementia risk were more robust between APOE ε4 allele carriers (odds ratio = 0.23) compared to non-carriers (odds ratio = 0.59) among 1449 older adults (age range: 65–79) [106]. In contrast however, Podewils et al. [67] found an inverse association between PA and dementia risk for APOE ε4 non-carriers, but found no association for APOE ε4 carriers in their prospective study including 3075 men and women (age: ≥65; mean follow-up: 5.4 years). In this study, PA levels were obtained via self-report and the kcal/week and number of activities (range: 0-14) performed in the previous two weeks reported. It was interesting to note that participation in multiple (≥4) different types of activities appeared to be as important if not more important than the self-reported PA levels in this cohort [67].

*Brain-derived neurotrophic factors (BDNF) and PA*

There is evidence [107, 108] suggesting a possible association between habitual PA or cardiorespiratory fitness (CRF) and BDNF concentration, although this is not consistently observed [109-112]. For example, Zoladz et al. [108] found that basal BDNF concentration was significantly higher in trained athletes (n = 16) compared to untrained individuals (n = 13) which agrees with findings by Correia et al. [107] in international and domestic level sprinters (versus sedentary individuals). In contrast however, Winker et al. [112] found no difference in BDNF concentration between active elderly marathon runners and cyclists (n = 56) matched for age, sex, and years of education to sedentary individuals (n = 58). While Nofuji et al. [111] found basal BDNF concentrations were lower in a group of trained males (n = 12) compared to a group of sedentary males (n = 14). In line with this view, Chan et al. [109] and Jung et al. [110] observed an inverse relationship between BDNF concentrations and PA level (n = 85) and CRF (n = 995) respectively. Therefore, based on the current evidence, a clear relationship between PA or CRF and BDNF cannot be made [113]. While
these mixed findings may be attributed to several potential confounders, the circadian variation in circulating BDNF concentrations (typically reflecting cortisol concentration) in both men [114] and women [115] may partly explain some of the variance in study outcomes.

Findings from prospective research assessing the effects of aerobic exercise on serum BDNF concentration have also been mixed, with resting BDNF concentration remaining largely unchanged in response to chronic exercise training despite acute exercise yielding substantial transient increases in BDNF concentration [113]. Aerobic exercise training (three or five weeks) did not alter basal BDNF concentrations in 47 sedentary adult males [116]. Likewise, a six-month longitudinal intervention testing the effects of low and moderate PA in 62 cognitively healthy elderly subjects found only a non-significant positive trend between BDNF and PA [117]. More specific to the current review is the finding that in individuals with amnestic MCI ($N = 33$), chronic (six months, four days/week, supervised) aerobic exercise did not significantly alter BDNF levels, although a potential sex difference has been proposed [118]. In contrast, a robust albeit transient increase in BDNF concentration has been observed in response to acute aerobic exercise [113] and this has occurred concomitant with improved cognitive performance. For example, acute cycling [119-123], stepping [124], and rowing [125] have been shown to increase serum BDNF levels in healthy and clinical populations with major depression [121] and individuals with spinal cord injury [123]. There is evidence that the magnitude of change in BDNF concentration is intensity-dependent, with acute high-intensity exercise having a greater effect on peripheral BDNF concentration compared to acute low-intensity exercise [113, 119, 122]. For example, Ferris et al. [119] showed a pre-post significant increase in serum BDNF concentration following 30 minutes of cycling at 10% above, but not at 20% below ventilatory threshold. Nevertheless, increases in peripheral BDNF following acute exercise seem to be transient, since concentrations in BDNF return to baseline during passive recovery [126]. Moreover, Ferris et al. [119] and
Griffin et al. [116] also observed improvements in cognitive performance (as assessed using the Stroop colour/word test and face-name matching task respectively); however, Ferris et al. [119] lacked a control group and cognitive scores did not correlate with circulating BDNF levels. It is worth noting that some of these studies had small sample sizes including eight [125], 11 [120], and 15 [119] participants. Nevertheless, a recent review concluded that acute \((n = 14\) studies) and chronic \((n = 6\) studies) aerobic exercise increased peripheral BDNF concentrations [113], albeit this effect was more robust for the transient increase in response to acute exercise.

There is currently little evidence to suggest that resistance training performed in isolation will increase BDNF concentration [113]. However, in a convenience sample of 48 elderly women Coelho et al. [127] found increases in plasma BDNF levels following resistance training three times/week for 10 weeks, while Yarrow et al. [128] found transient increases in serum BDNF after an acute bout of resistance training in 20 males but did not observe changes in resting BDNF levels after five weeks of training. The lack of effect on serum BDNF following acute resistance training has been replicated in healthy males \((N = 16)\) [129], sports students \((N = 19; 12\) weeks) [130], and untrained older individuals \([\text{mean age: } 50.9 (6.2)\) who trained three times/week for 10 weeks [131].

In sum, observational and experimental evidence in healthy and cognitively impaired populations suggest that increased PA is associated with increases in basal BDNF levels. These increases may be moderated by the type of exercise, with aerobic exercise generating more robust effects on BDNFs and by a dose-response relationship which may be intensity-dependent [113, 119, 122], but the observed increases return to baseline \([15-60\) minutes\) during passive recovery [126]. In some cases, increases in BDNF levels correlate with improved cognitive performance [116], while in other they do not [119]. It is worth mentioning that blood processing (serum: clotting time and temperature; plasma: platelet
stores) [132], circadian rhythms [114, 115], and phases of the menstrual cycle [115] may affect BDNF concentrations.

**Brain biomarkers in AD pathophysiology and their interaction with PA**

Pathophysiological abnormalities can occur years before clinical symptoms manifest which makes neuroimaging techniques particularly useful in the preclinical and early clinical stages of AD [15, 133]. Because a definitive AD diagnosis can only be achieved with histopathological confirmation, the inclusion of neuroimaging methods can increase specificity and diagnostic value in clinical and research settings [134]. Neuroimaging methods used in diagnosing neurodegenerative disorders include structural imaging such as magnetic resonance imaging (MRI), computer-assisted tomography (CT), functional imaging such as single photon emission tomography (SPECT), and positron emission tomography (PET) [135]. For diagnosing AD and compared to other neuroimaging techniques, PET has one of the highest rates of sensitivity (86%) and specificity (86%; [136]). Therefore, the following sections will only discuss PET imaging.

*PA findings on Pittsburgh compound B-PET*

Currently, PET imaging using radioactive tracers are available to detect Aβ deposits in the brain. The Pittsburgh compound B (PiB), which is the analogue of thioflavin T, is the most commonly used tracer for amyloid imaging [137]. In PiB-PET, the compound binds to fibrillar Aβ, while it does not bind to diffuse plaques and soluble Aβ, and can be used to differentiate AD and healthy controls [138, 139]. Moreover, individuals with subtle decline in episodic memory- the first cognitive faculty affected by AD- and a PiB-PET positive scan,
have a 50% increased risk of progressing to MCI and AD within three years [140]. Nordberg et al. [139] found that individuals with MCI with PiB-PET positive scans, not only had greater memory impairment compared to those with PiB-PET negative scans at baseline, but also progressed to AD at a rate of ~25% per year compared to a 0% conversion rate in those with PiB-PET negative scans. It is worth noting however, that while some individuals who are PiB positive progress to AD, others do not; indeed, PiB retention alone, does not appear to correlate well with cognitive impairment [141].

There is evidence that PA levels might affect Aβ levels in the brain, cerebrospinal fluid (CSF), and blood. For example, Liang et al. [142] found that individuals (N = 69) meeting the American Heart Association guidelines of 7.5 metabolic equivalent (MET)-hours/week of exercise showed significantly lower PiB binding and higher levels of Aβ42 (considered the more fibrillogenic form of Aβ and more closely associated with disease states) levels in the CSF. Importantly, Aβ42 levels in the CSF are inversely associated with Aβ aggregation in the brain and as such, higher levels in the CSF are generally accepted as being indicative of a healthier amyloid profile. In a larger sample (N = 546), Brown et al. [143] found that higher PA was negatively associated with lower plasma Aβ1−42/1−40 ratio, and, after stratifying participants by APOE ε4 allele status, this association was present in APOE ε4 allele non-carriers but absent in APOE ε4 allele carriers. Conversely, there was an association between higher PA levels and lower amyloid brain load as measured by PiB PET, in APOE ε4 allele carriers only [143]. Furthermore, in a sample of 201 cognitively intact adults, Head et al. [144] found a sedentary lifestyle was associated with higher PiB binding and lower CSF Aβ42 levels. Moreover, a significant interaction between APOE ε4 allele status and exercise engagement for PiB binding was revealed, with a sedentary lifestyle being associated with higher PiB binding in APOE ε4 allele carriers (p = .013) but not for APOE ε4 non-carriers.
PA findings on [18F] fluorodeoxyglucose

At rest, brain activity almost exclusively depends on glucose metabolism; therefore, glucose hypo-metabolism is considered a more robust surrogate marker of cognitive decline compared to presence of excessive Aβ [145]. The [F18] fluorodeoxyglucose (FDG-PET) can be used to generate images of glucose uptake into neural cells in different brain regions. Topographical patterns of reduced uptake, shown by reduced [18F] signal intensity, indicate neurodegeneration even after correcting for cortical atrophy in AD patients [146]. Therefore, FDG-PET can be used to investigate the relationship between PA and glucose metabolism. Deeny et al. [147] found that during a working memory task among individuals with the APOE ε4 allele, highly fit (indicative of high PA) compared to low fit (indicative of low PA) elderly females, showed greater glucose uptake in the temporal lobe, which is a brain region affected by AD. In contrast, low fit APOE ε4 allele carriers, showed greater cerebral glucose uptake in the frontal and parietal regions; this relationship however, was not observed during resting glucose metabolism. This indicates that CRF levels likely affect glucose metabolism in the brain in APOE ε4 allele carriers compared to non-carriers, which is in line with previous studies having found that the effects of exercise on cognition tend to be more pronounced in APOE ε4 allele carriers compared to non-carriers [103, 104]. This conclusion is supported by experimental data showing that a combination of PA and computerized brain training improved glucose metabolism in the left sensorimotor cortex even after adjusting for age, sex, premorbid IQ, APOE ε4 allele status, and history of head injury among non-clinical older individuals [27]. Together, the findings from Shah et al. [27] and Deeny et al. [147] suggest that increased PA may improve cerebral glucose metabolism and counteract the decline in brain glucose uptake that is often present in the preclinical phases of AD.
Limitations of studies on PA and AD with suggestions for future research

Epidemiological data

Findings drawn from the epidemiological and experimental literature are limited by several methodological shortcomings. For example, in many observational studies, researchers assessed PA through a single self-report questionnaire [56, 58, 67, 148]. These are susceptible to information bias [149], which threatens internal validity [150], and self-rated fitness has been shown to be inversely related to perceived physical exertion in men and women [151]. Additionally, some epidemiological studies administered PA questionnaires that were not designed for elderly populations [67] or lacked psychometric properties [103, 106, 152]. Several studies have also used different methods to classify duration and intensity of PA or combined frequency, duration, and intensity [56] making it difficult to establish a clear dose-response relationship [153]. From the extant epidemiological literature it is unclear whether lifelong engagement in high PA is necessary to alter disease-risk profiles, due to most studies having focused in elderly populations [56]. Finally, the effects of PA might at least in part be mediated by other factors including social engagement [63] and number of different activity-types [67], which could explain some of the variance in the incidence rate of AD; but this has not been considered in all epidemiological literature [148].

Experimental data

A limitation of studies in the field relates to the different types of dementia and the different severity-levels being assessed [29, 30] with PA likely affecting different dementia subtypes to a varying extent [79]. Also, studies included a wide range of different types of exercise programs which included both supervised and unsupervised training prescribed with different
intensities, frequencies, and durations. Similarly, to the epidemiological studies, the assessment of cognition (e.g. MMSE versus executive function tasks) and chosen risk-markers of AD or cognitive impairment (blood samples versus CSF samples; neuroimaging techniques versus protein concentrations) vary widely in the extant literature, making comparisons problematic [30]. In addition to these methodological limitations, some trials did not report intention-to-treat analyses [68, 77], adherence or attrition rates [77], and were likely statistically underpowered [29, 69, 154] particularly when multivariate adjustments were conducted.

Suggestions for future research

Drawing from the limitations identified above, to improve methodology and decrease risk of bias, several suggestions for future epidemiological and experimental research are warranted. Since the extant epidemiological evidence has mostly focused on elderly individuals [56], from a public health perspective, it is not possible to generate recommendations for different age groups for mitigating the age-related decline in cognition [6] or alter risk profiles for dementia. While not without complications, future epidemiological research should focus on exploring the relationship between lifelong exposure to PA and dementia [153] using a validated tool in combination with a CRF test or by linking with existing longitudinal studies which have collected PA levels or data-linkage programs which have access to individuals PA levels. The level of PA needs to be assessed in a more precise way using objective measures such as accelerometry [59] over at least a 2-week period, although these devices are not without short-comings [155]. While in experimental trials with smaller numbers, direct supervision of exercise participation is critical to generate more precise estimates of PA. Since AD populations find it difficult to adhere to PA interventions [148, 156], it is also important to investigate what type of PA intervention is appropriate for cognitively impaired
individuals [148] in prospective studies, particularly since comorbid conditions (e.g., osteoarthritis) often present in elderly population [157].

Additionally, most experimental studies have focused on cognition as the primary outcome measure excluding other important variables [30]. The European consensus on endpoints for trials on AD advises to include functional and proxy endpoints to capture a more comprehensive functioning of the patient [158]. Thus, to render the findings more clinically relevant, the implementation of an exhaustive battery of global functioning (with a priori adjustment for Type I error) including cognitive and functional tests is important for establishing whether the intervention has helped the patient to meaningfully regain quality of life [158, 159]. In studies focussed on individuals at greater risk of progressing to AD (e.g., SCD), reports of SCDs by proxy (e.g., informant) correlate better with objective performance [160] and is a robust predictor of progression to AD [161] and should therefore be considered. Finally, to assess the role of exercise in altering AD-risk and progression in the early stages (e.g., SCD and MCI), adoption of the multiple biomarkers which can better predict AD-risk and progression are recommended.

**Conclusions**

There is compelling evidence that exercise improves biomarkers of AD and cognitive performance and that greater engagement in chronic PA is associated with improved cognition, a later-onset AD, and slowed disease progression. However, conflicting results are prevalent in the extant literature. Epidemiological evidence suggests a clear benefit of PA on improving cognition and reducing AD risk, particularly when the PA is performed at high-intensity, however the magnitude of these associations are typically diminished after analyses are adjusted for the multiple confounding variables (i.e. age, sex, BMI, APOE ε4 status,
educational level, smoking, alcohol intake, social support, difficulty in activities of daily living, and instrumental activities of daily living) which often cluster with PA levels.

The experimental studies assessing the impact of acute exercise suggest (i) a transient (minutes to hours) increase in BDNF following aerobic exercise, which is not observed following resistance exercise; (ii) greater BDNF responses following higher intensity aerobic exercise; (iii) improvement in executive function tasks. While the experimental studies assessing exercise training suggest (i) improvements in cognition following exercise training in pre-clinical and early clinical stages of dementia/AD (i.e., SMC, MCI), but mixed results in individuals with dementia/AD; (ii) greater improvements in cognition across the dementia/AD spectrum when training incorporated a greater frequency of both aerobic and non-aerobic activities. The findings from studies adopting neuroimaging techniques, which are considered more sensitive than other outcome measures, suggest that exercise and PA plays an important protective role against AD pathophysiology, particularly in individuals with an APOE ε4 allele, which is consistent with the findings from studies using cognitive outcome measures in this cohort. Despite the equivocal nature of the available evidence, it remains noteworthy that a single modifiable risk factor such as PA has largely been shown to affect both the onset and progression of a disease as complex as AD. From a public health perspective, it is important that future research addresses the limitations of previous research and establish the direct association between PA and AD.
Compliance with Ethical Standards

Funding

The first author was on a PhD scholarship from Murdoch University during the completion of this review.

Conflict of interest

Ralph N. Martins is the founder and chief scientific officer of the biotech company, Alzhyme. Hamid R. Sohrabi has, and continues, to receive remuneration from activities with Takeda Pharmaceuticals.
References


10. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC et al. The diagnosis of mild cognitive impairment due to Alzheimer’s disease: Recommendations from


Between-chapter (2-3) rationale

In Chapter two, we found that increasing PA in individuals with AD does not appear to generate clinically relevant improvements in cognitive symptoms. It is possible that increasing PA may not be beneficial once the disease has reached an irreversible degree of neuropathology. A small number of studies have indicated that increasing PA in seniors without a diagnosis of AD but are at a higher risk of developing it, for example those with SCD, can improve cognitive performance. As such, increasing PA in this target population may be a more useful strategy in delaying the onset of AD. There is also evidence indicating that targeting only one risk factors (low PA) may be not be sufficient in delaying the onset of AD, however. Targeting another risk in combination with increasing PA may therefore generate more clinically relevant outcomes than targeting PA in isolation. Bilingualism (the ability to speak two or more languages) has been linked with a delayed onset of AD symptoms and diagnosis. Not all studies have documented such benefits, however. The purpose of the next Chapter is to determine whether bilingualism may be a potential modifiable risk factor in delaying AD.
Chapter 3 Bilingualism is associated with a delayed onset of dementia but not with a lower risk of developing it: a systematic review with meta-analyses.
Title: Bilingualism is associated with a delayed onset of dementia but not with a lower risk of developing it: a systematic review with meta-analyses.

Authors: Stefano Brini\textsuperscript{a,b,c,i}, Hamid, R. Sohrabi\textsuperscript{d,e,f}, Jeffrey, J. Hebert\textsuperscript{g}, Mitchell R. L. Forrest\textsuperscript{a}, Matti Lainen, Heikki Hämäläinen\textsuperscript{b,c}, Mira Karrasch\textsuperscript{b}, Jeremiah J. Peiffer\textsuperscript{a}, Ralph N. Martins\textsuperscript{d,e,f}, Timothy J. Fairchild\textsuperscript{a}

\textsuperscript{a}School of Psychology and Exercise Science, Murdoch University, Perth, Western Australia, Australia
\textsuperscript{b}Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland
\textsuperscript{c}Turku Brain and Mind Center, Turku, Finland
\textsuperscript{d}School of Medical and Health Sciences, Edith Cowan University, Perth, Western Australia, Australia
\textsuperscript{e}Department of Biomedical Sciences, Macquarie University, New South Wales, Australia
\textsuperscript{f}Australian Alzheimer’s Research Foundation, Perth, Western Australia, Australia
\textsuperscript{g}Faculty of Kinesiology, University of New Brunswick, Fredericton, Canada
\textsuperscript{h}Department of Psychology, Åbo Akademi University, Turku, Finland
\textsuperscript{i}Health Services Research and Management School of Health Sciences, City, University of London, London, United Kingdom

Conflict of interest: SB received a scholarship from Murdoch University. ML was supported by the Academy of Finland (grant #260276) and the Åbo Akademi University Endowment (the BrainTrain project).

Corresponding author.
Stefano Brini
School of Health Sciences
City, University of London
10 Northampton Square
London, EC1V OHB
Email: stefano.brini@city.ac.uk
ORCID ID: https://orcid.org/0000-0002-1909-1796
Abstract

Some studies have linked bilingualism with a later onset of dementia, Alzheimer’s disease (AD), and mild cognitive impairment (MCI). Not all studies have observed such relationships, however. Differences in study outcomes may be due to methodological limitations and the presence of confounding factors within studies such as immigration status and level of education. We conducted the first systematic review with meta-analysis combining cross-sectional studies to explore if bilingualism might delay symptom onset and diagnosis of dementia, AD, and MCI. Primary outcomes included the age of symptom onset, the age at diagnosis of MCI or dementia, and the risk of developing MCI or dementia. A secondary outcome included the degree of disease severity at dementia diagnosis. There was no difference in the age of MCI diagnosis between monolinguals and bilinguals [mean difference: 3.2; 95% confidence intervals (CI): −3.4, 9.7]. Bilinguals vs. monolinguals reported experiencing AD symptoms 4.7 years (95% CI: 3.3, 6.1) later. Bilinguals vs. monolinguals were diagnosed with dementia 3.3 years (95% CI: 1.7, 4.9) later. Here, 95% prediction intervals showed a large dispersion of effect sizes (−1.9 to 8.5). We investigated this dispersion with a subgroup meta-analysis comparing studies that had recruited participants with dementia to studies that had recruited participants with AD on the age of dementia and AD diagnosis between mono- and bilinguals. Results showed that bilinguals vs. monolinguals were 1.9 years (95% CI: −0.9, 4.7) and 4.2 (95% CI: 2.0, 6.4) older than monolinguals at the time of dementia and AD diagnosis, respectively. The mean difference between the two subgroups was not significant. There was no significant risk reduction (odds ratio: 0.89; 95% CI: 0.68–1.16) in developing dementia among bilinguals vs. monolinguals. Also, there was no significant difference (Hedges’ g = 0.05; 95% CI: −0.13, 0.24) in disease severity at dementia diagnosis between bilinguals and monolinguals, despite bilinguals being significantly older. The majority of studies had adjusted for level of education suggesting that
education might not have played a role in the observed delay in dementia among bilinguals vs. monolinguals. Although findings indicated that bilingualism was on average related to a delayed onset of dementia, the magnitude of this relationship varied across different settings. This variation may be due to unexplained heterogeneity and different sources of bias in the included studies. Registration: PROSPERO CRD42015019100.

Registration: PROSPERO CRD42015019100

*Keywords:* bilingualism; mild cognitive impairment; dementia; Alzheimer’s disease; meta-analysis

**Key points**
- Meta-analytic results showed that bilinguals vs. monolinguals were older at the time of Alzheimer’s symptom onset and dementia diagnosis
- Meta-analytic results did not show a reduction in the risk of developing dementia among bilinguals vs. monolinguals
- Several sources of bias were identified including poor measurement of participants’ language profiles and lack of control over confounding factors
Introduction

Rationale

Approximately 43.8 million people lived with dementia worldwide in the year 2016 (Nichols et al., 2019) and this number is projected to increase to 115.5 million people by 2050 (Prince et al., 2013). The global economic cost of dementia is estimated to surpass US$2 trillion per year by 2030 (Wimo et al., 2017). A five-year delay in the onset of Alzheimer’s disease (AD), the most common form of dementia, could reduce the number of patients living with the disease worldwide by 57%, thereby alleviating the associated economic costs by half (Sperling et al., 2011). Therefore, identifying modifiable lifestyle factors that can slow or delay the onset of dementia is a world’s public health priority (WHO, 2017; Wortmann, 2012).

One such factor may be bilingualism, which is the ability to speak two languages (Luk & Bialystok, 2013). This hypothesis comes from studies showing that bilinguals develop mild cognitive impairment (MCI), dementia, and AD, 4–7 years later than monolinguals (Alladi et al., 2013; Bialystok, Craik, Binns, Ossher, & Freedman, 2014; Bialystok, Craik, & Freedman, 2007). Others, however, have not documented such differences (Lawton, Gasquoine, & Weimer, 2015; Yeung, John, Menec, & Tyas, 2014). Also, while longitudinal prospective studies showed no risk reduction among bilinguals relative to monolinguals (Ljungberg, Hansson, Adolfsson, & Nilsson, 2016; Yeung et al., 2014; Zahodne, Schofield, Farrell, Stern, & Manly, 2014), foreign language education during adolescence has been associated with reduced risk of MCI later in life (Wilson, Boyle, Yang, James, & Bennett, 2015). Some authors have argued that confounding factors including migration status and education may explain some differences in study outcomes in cross-sectional and longitudinal studies (Fuller-Thomson, 2015; Fuller-Thomson & Kuh, 2014).
One systematic review concluded that “public health policy should… remove recommendations regarding bilingualism as a strategy to delay dementia” (Mukadam, Sommerlad, & Livingston, 2017). However, the authors conducted a meta-analysis of only four longitudinal prospective studies without performing meta-analyses on cross-sectional reports. Moreover, while studies without a monolingual control group were excluded from this review (Mukadam et al., 2017), their meta-analysis included one study (Sanders, Hall, Katz, & Lipton, 2012) which did not clearly define the control group as monolingual. That review did not include age at MCI diagnosis as an outcome or studies published more recently (Hack, Dubin, Fernandes, Costa, & Tyas, 2019; Ljungberg et al., 2016; Perani et al., 2017; Ramakrishnan et al., 2017; Zheng et al., 2018). As such, before suggesting that bilingualism should not be recommended as a strategy for delaying dementia, a careful re-evaluation of the available evidence is necessary (Del Maschio, Fedeli, & Abutalebi, 2018).

**Objectives**

Differences in study outcomes in the field of bilingualism and dementia research as well as the need to identify strategies to delay the onset of dementia as highlighted in the Global plan on the public health response to dementia 2017–2025 by the World Health Organization (WHO, 2017) prompted this systematic review. We assessed whether bilingualism relative to monolingualism might delay the age at which participants experienced the initial symptoms of AD and delay the age at which participants were diagnosed with MCI or dementia. We also examined whether bilingualism might be associated with a lower risk of dementia. The primary objectives were to review cross-sectional and longitudinal prospective studies investigating (i) differences in the age of symptom onset and age at diagnosis of MCI or dementia between older monolinguals and bilinguals, and (ii) the relationship between bilingualism relative to monolingualism and risk of dementia in older cognitively intact
adults. A secondary objective was to investigate differences in disease severity at dementia diagnosis between older monolinguals and bilinguals.

**Methods**

*Search strategy and selection criteria*

This systematic review with meta-analyses accords with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Statement (PRISMA; Moher, Liberati, Tetzlaff, & Altman, 2010). Eligible studies had to compare monolingual to bilingual participants on at least one of the following outcomes: reported age of symptom onset or age at diagnosis for MCI or dementia, degree of cognitive impairment at dementia diagnosis, or risk of dementia or MCI. Given the lack of a clear uniform definition of bilingualism in the literature, we included studies independently of the way bilingualism was operationalized or measured, or whether proficiency in the second language had been objectively assessed. We included studies that had recruited participants with MCI or dementia as assessed using clinical measures such as the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRDA) as well as cognitively intact individuals. We excluded studies without a group of monolinguals. We also excluded reports, conference abstracts, reviews, commentaries, editorials, letters, news articles, case series, and discussion forums, as well as grey literature including non-peer reviewed empirical studies. We searched cross-sectional, prospective and retrospective cohort studies, case-control studies, and randomized controlled trials across several databases including CINHAL, The Cochrane Library, PubMed, PsycINFO, LILACS, and Embase. Filters were used to exclude animal studies, but no restrictions were placed on time and language. The initial search was performed on September 3, 2015 and refreshed
several times with the last refresh being complete on December 5, 2018. We used similar keywords and criteria for each search.

The database searches were conducted by S.B., while the screening for title and abstract as well as the full-text screening was conducted independently by pairs of review authors (S.B., T.J.F., and J.J.H.). Data extraction was completed independently by pairs of review authors (S.B., M.F., T.J.F., and J.P.). We used Covidence software for each of these steps (Innovation, 2017). Disagreements were resolved through consensus and discussion with a third review author. We re-quested additional information from corresponding authors when necessary. Details of the protocol for this systematic review were registered a priori (PROSPERO 2015 CRD42015019100).


Study selection

The number of studies screened and included for quantitative synthesis is presented in Fig. 1. The qualitative synthesis included three cross-sectional studies with age at MCI diagnosis (Table 1) and 16 cross-sectional studies with age at AD symptom onset and age at dementia
or AD clinical diagnosis (Table 2). There was one longitudinal prospective study with the risk of MCI as the outcome (Table 3) and five with the risk of dementia as the outcome (Table 4).

*Figure 0.1 PRISMA flow chart*
Description of mild cognitive impairment and dementia

Years before clinical diagnosis, an individual may experience MCI which can either be of the amnestic or non-amnestic type (Pandya et al., 2016). The former is marked by memory impairment more severe than would be expected for the age of the individual and this is a risk factor for AD. In the non-amnestic type, other cognitive abilities (e.g., language) rather than memory are affected and this is a risk factor mainly for other types of dementia. However, some individuals who experience MCI of either type do not necessarily progress to AD or other forms of dementia (Pandya et al., 2016). Dementia is a progressive clinical syndrome presenting with impairment in cognition, daily functioning, and changes in behavior in the absence of any impairment in consciousness (Vinters, 2015). While dementia is an umbrella that describes a significant cognitive and functional decline usually caused by a wide range of neurodegenerative diseases, AD has a specific etiology marked by a progressive and irreversible amnestic disorder followed by a decline in other cognitive abilities and behavior as well as neuropsychiatric dysfunctions resulting in total dependence (Vinters, 2015).

Diagnosis of AD is based on clinical presentation (e.g., Diagnostic and Statistical Manual of Mental Disorders) and neuropsychological assessment while neuro-imaging is used to support clinical evaluation. However, a definite diagnosis can only be given by the NINCDS-ADRDA criteria with histopathological evidence supporting clinical diagnosis (Dubois et al., 2007).

Data extraction and risk of bias

We extracted information on sample size, sex, mean age at diagnosis, education level, language measure, measures to diagnose dementia, dementia subtype, degree of cognitive impairment outcomes, and study results. Two authors (SB and MF) independently assessed risk of bias at both the study and outcome level by using the modified version of the
Newcastle-Ottawa Scale (NOS) to assess risk of bias for cross-sectional studies (Wells et al., 2015). The modified NOS allows to allocate a maximum of 10 stars to each study across three domains: selection of study groups (range 0–5), comparability of study groups (range 0–2), and exposure/outcome ascertainment (range 0–3). For longitudinal studies, we used the original version of the NOS for cohort studies, which allows allocating a maximum of nine stars across the same domains as in the modified version.

Because the included studies were sufficiently similar regarding the research question, methodology, and outcome, we conducted a quantitative synthesis of the data by meta-analyzing effect sizes from included studies.

In cross-sectional studies, the authors reported the age of symptom onset for AD and age of clinical diagnosis for MCI, dementia, and AD as absolute numbers in years. Longitudinal prospective studies reported the risk of dementia as relative risk – the risk of developing dementia in bilinguals relative to monolingual controls and odds ratio – the odds of developing dementia given language status (i.e., bilingualism vs. monolingualism). One longitudinal prospective study reported the proportional hazard ratios in estimating the relationship between early foreign language instruction and the risk of developing MCI later in life (Wilson et al., 2015).

Data analysis

Our primary outcome measures were the age of symptom onset and age at diagnosis of MCI or dementia and the risk of developing dementia. A secondary outcome included the degree of disease severity at dementia diagnosis. Here, age at diagnosis was defined as the age at which participants were diagnosed with MCI, AD, or dementia and age of symptom onset was defined as the participants’ or informants’ retrospective recall of the age at which the first symptoms of cognitive impairment started. However, most studies that reported the age
of symptom onset included participants with AD, not dementia. Therefore, we could only conduct a meta-analysis on the age of symptom onset for participants with AD, not dementia. Our secondary outcome was the degree of cognitive impairment as measured by the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) during dementia or AD diagnosis, and dementia risk.

All the meta-analyses conducted here were based on random-effects models at an alpha level of .05 with the Knapp-Hartung adjustment (IntHout, Ioannidis, & Borm, 2014). Because studies did not provide individual-level data, we retrieved summary data. One study included two monolingual groups: one Mandarin and one Cantonese (Zheng et al., 2018). To increase the sample size, we combined the sample sizes, means, and standard deviations on the age of symptom onset, age of clinical diagnosis, and degree of cognitive impairment from the Mandarin and Cantonese group to form one monolingual group. For these calculations, we used the formula provided by the Cochrane Collaboration (Higgins & Green, 2016). For degree of cognitive impairment at dementia diagnosis, we presented MMSE scores (range: 0–30) as Hedges’ g between monolinguals and bilinguals because one study (Lawton et al., 2015) reported scores from the Modified Mini-Mental Status Examination (3MSE), which uses a scale from 0 to 100 points. Also, as not all prospective studies provided the same outcome results (one study provided hazard ratios and others provided log odds ratios), we extracted the unadjusted raw values of participants who had remained free of dementia and of those who had converted to dementia from the mono- and bilingual group. These values are unadjusted values but it was necessary to use these in order to combine results into a meta-analysis.

Data were analyzed using Comprehensive Meta-Analysis Software: version 3 (Borenstein, Rothstein, & Cohen, 2005). For cross-sectional studies, we presented mean differences between monolinguals and bilinguals for our primary and secondary outcomes.
(mean age in years and mean MMSE scores). For longitudinal studies, we presented odds ratio. We presented 95% confidence intervals (CI) around the pooled estimates (Riley, Higgins, & Deeks, 2011). We also computed 95% prediction intervals (PI), which reflect the distribution of effect sizes across different settings and estimate the expected effect sizes for future settings (IntHout, Ioannidis, Rovers, & Goeman, 2016). However, we computed PIs for meta-analyses with at least 10 studies (Hedges & Vevea, 1998). We used tau-squared (T2) to investigate between-study heterogeneity, with a non-zero T2 value indicating between-study heterogeneity. To investigate small- study effects, we generated funnel plots for meta-analyses that include at least 10 studies (Lau, Ioannidis, Terrin, Schmid, & Olkin, 2006). To explore the impact that imputing missing studies might have on the pooled estimate, we conducted Duval and Tweedie’s trim and fill test (Duval & Tweedie, 2000). We did not conduct formal tests for funnel plot asymmetry in meta-analyses with fewer than 10 studies (Sterne et al., 2011).

Because we did not pre-specify potential covariates and to avoid data dredging (Thompson & Higgins, 2002), we restricted our investigation of heterogeneity to immigration status (Fuller-Thomson & Kuh, 2014; Mukadam et al., 2017) and dementia etiology due to its clinical relevance (Bialystok, Abutalebi, Bak, Burke, & Kroll, 2016; IntHout et al., 2016). In two subgroup meta-analyses, we compared studies that had recruited participants with dementia (irrespective of etiology) to studies that had recruited participants with AD (specific etiology) on the age of dementia and AD diagnosis. In the other subgroup meta-analysis, we compared studies explicitly mentioning that the statistical analyses had been adjusted for immigration status or at least that the analytic cohort did not include migrants to studies not explicitly mentioning whether the statistical analyses had adjusted for migration status or whether the analytic sample had included migrants. We reported the pooled estimates for heterogeneity in subgroup meta-analyses.
**Results**

*Data collection process*

We extracted demographic data including sample size, percentage of females in each group, and education level. Moreover, we extracted methodological data including the operationalization and measurement of participants’ language profiles, type of diagnosis (i.e., MCI, dementia, or AD), as well as the measurement tools used for making the clinical diagnosis of MCI or any dementia. We also extracted data for each outcome in each group including mean age of dementia diagnosis, mean age of dementia symptom onset, risk of MCI or dementia, and degree of cognitive impairment. We were able to extract sufficient data on age of MCI (k = 4) and dementia clinical diagnosis (k = 13), AD symptom onset (k = 7), degree of cognitive impairment (k = 12), and risk of dementia (k = 5) to conduct a meta-analysis on each of these outcomes. The total number of participants in cross-sectional studies was 4671 including 2376 monolinguals and 2295 bilinguals (Table 1 and 2). There were 121 monolinguals and 159 bilinguals in cross-sectional studies with MCI diagnosis as an outcome (Table 1), and 2256 monolinguals and 2136 bilinguals in studies with dementia diagnosis as an outcome (Table 2). There were six longitudinal prospective studies comprising a total of 4227 participants (Tables 3 and 4).

*Study characteristics*

The operationalization of bilingualism differed across studies including: “had spent the majority of their lives, at least from early adulthood regularly using at least two languages” (Bialystok et al., 2007; Craik, Bialystok, & Freedman, 2010), “the ability to communicate in two or more languages in interaction with other speakers of these same languages” (Alladi et al., 2013; Alladi et al., 2017), “individuals had spent the majority of their lives, beginning at
least in early adulthood, speaking two or more languages fluently—ideally daily, but at least weekly” (Bialystok et al., 2014; Chertkow et al., 2010; Ossher, Bialystok, Craik, Murphy, & Troyer, 2012), “able to communicate fluently at least in 2 languages and made regular use for both” (Estanga et al., 2017), “ability to meet the communicative demands of the self and the society in their normal functioning in 2 or more languages in their interaction with other speakers of any or all of these languages” (Ramakrishnan et al., 2017), “fluent in a second language and had used both languages consistently throughout most of his or, her life” (Schweizer, Ware, Fischer, Craik, & Bialystok, 2012), “determined on the basis of second language proficiency and frequency of use” (Woumans et al., 2015) or did not apply a specific definition (Lawton et al., 2015; Ljungberg et al., 2016; Perani et al., 2017; Wilson et al., 2015; Yeung et al., 2014; Zahodne et al., 2014). One study used more strict definitions for monolingualism and bilingualism including “speaking English for all or most of one’s life and being fluent in English, but not in any other language” and “speaking both Welsh and English for all or most of one’s life and being fluent in both languages, but not in any other languages”, respectively (Clare et al., 2016).

Studies used different types of measurements for bilingualism (Tables 1, 2, 3, and 4). While several cross-sectional studies used validated measures including questionnaires to measure bilingualism (Bialystok et al., 2014; Clare et al., 2016; Estanga et al., 2017; Lawton et al., 2015; Ossher et al., 2012), others used non-validated methods (Alladi et al., 2013; Bialystok et al., 2007; Chertkow et al., 2010; Schweizer et al., 2012; Woumans et al., 2015), or did not report the method of collection (Craik et al., 2010; Ramakrishnan et al., 2017). Similarly, one longitudinal study assessed participants’ language profiles with a non-validated measure (Wilson et al., 2015), two used a questionnaire but did not report their psychometric properties (Hack et al., 2019; Ljungberg et al., 2016), while one study validated their measure as part of the study (Zahodne et al., 2014).
There were differences in the type of MCI and dementia across studies (Table 1–4). Four studies recruited participants with MCI (Bialystok et al., 2014; Ossher et al., 2012; Ramakrishnan et al., 2017; Wilson et al., 2015). The type of MCI differed across studies with two studies recruiting individuals with MCI without describing its subtype (Bialystok et al., 2014; Wilson et al., 2015), another study recruited individuals with single and multiple domain amnestic MCI (Ossher et al., 2012), while still another study recruited individuals with amnestic and non-amnestic MCI (Ramakrishnan et al., 2017).

The tools for diagnosing MCI and dementia as well as the dementia subtypes differed across studies (Table 1–4). For MCI, studies either did not report the method of diagnosis (Bialystok et al., 2014), diagnosed MCI during a clinical interview with neuropsychological tests (Ossher et al., 2012), adopted the Mayo Clinic MCI criteria [(Ramakrishnan et al., 2017) Table 1], or the NINCDS-ADRDA criteria (Wilson et al., 2015). The diagnosis of dementia was often based on a clinical interview conducted by medical staff (e.g., a neurologist) and a neuropsychological assessment and using NINCDS-ADRDA criteria (Bialystok et al., 2007; Chertkow et al., 2010; Lawton et al., 2015; Ljungberg et al., 2016), the International Classification of Diseases 10 (Clare et al., 2016), or the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (Alladi et al., 2013), among others (Table 2).

Studies recruited participants with a wide range of dementia subtypes including the behavioral variant frontotemporal dementia, progressive non-fluent aphasia, semantic dementia, frontotemporal dementia-motor neuron disease, corticobasal degeneration, and progressive supranuclear palsy (Alladi et al., 2017), AD (Clare et al., 2016; Ljungberg et al., 2016; Woumans et al., 2015), vascular dementia (Alladi et al., 2013; Ljungberg et al., 2016; Zahodne et al., 2014), mixed AD with cardiovascular disease, frontotemporal dementia, dementia with Lewy bodies (Alladi et al., 2013; Zahodne et al., 2014), probable AD (Bialystok et al., 2014; Chertkow et al., 2010; Craik et al., 2010; Lawton et al., 2015; Perani
et al., 2017; Schweizer et al., 2012; Zahodne et al., 2014), possible AD (Bialystok et al.,
2007; Lawton et al., 2015; Zahodne et al., 2014), dementia due to other neurodegenerative
disorders, cardiovascular disease (Bialystok et al., 2007), preclinical AD (Estanga et al.,
2017), frontal lobe dementia (Ljungberg et al., 2016), and dementia [(Hack et al., 2019;
Yeung et al., 2004) Table 2].

Studies also recruited participants with a wide range of dementia subtypes including
the behavioural variant frontotemporal dementia, progressive non-fluent aphasia, semantic
dementia, frontotemporal dementia-motor neuron disease, corticobasal degeneration, and
progressive supranuclear palsy (Alladi et al., 2017), AD (Clare et al., 2016; Ljungberg et al.,
2016; Woumans et al., 2015), vascular dementia (Alladi et al., 2013b; Ljungberg et al., 2016;
Zahodne et al., 2014), mixed AD with cardiovascular disease, frontotemporal dementia,
dementia with Lewy bodies (Alladi et al., 2013b; Zahodne et al., 2014), probable AD
(Bialystok et al., 2014; Chertkow et al., 2010; Craik et al., 2010; Lawton et al., 2015; Perani
et al., 2017; Schweizer et al., 2012; Zahodne et al., 2014), possible AD (Bialystok et al.,
2007; Lawton et al., 2015; Zahodne et al., 2014), dementia due to other neurodegenerative
disorders, cardiovascular disease (Bialystok et al., 2007), preclinical AD (Estanga et al.,
2017), frontal lobe dementia (Ljungberg et al., 2016), and dementia [(Yeung et al., 2014)
Table 2].
### Table 0-1 Cross-sectional studies investigating the relationship between bilingualism and MCI

<table>
<thead>
<tr>
<th>Study</th>
<th>N (% of females)</th>
<th>Mean age of dx &amp; sx onset</th>
<th>Education level</th>
<th>Language measure</th>
<th>MCI dx and type</th>
<th>MMSE scores</th>
<th>Effect sizes (MD = BL age minus ML age, 95% CI)</th>
<th>Age of dx</th>
<th>Age of onset</th>
<th>Cognitive impairment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bialystok 2014</td>
<td>ML: 28 (50%);</td>
<td>ML: 66.5 (12.3)</td>
<td>ML: 15.5 (3.8)</td>
<td>LSBQ</td>
<td>MCI</td>
<td>ML: 29 (1.4)</td>
<td>3.5 (-1.77-8.77)</td>
<td>4.7 (0.97-10.37)</td>
<td>0.6 (-0.17-1.37)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BL: 26 (56%)</td>
<td>BL: 70.0 (10.7)</td>
<td>BL: 14.3</td>
<td></td>
<td></td>
<td>BL: 28.4 (1.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ossher 2013</td>
<td>ML (SDaMCI):</td>
<td>ML (SDaMCI): 74.9 (6.9)</td>
<td>ML (SDaMCI): 14.7 (2.5)</td>
<td>Questionnaire</td>
<td>Clinical interview including neuropsychological tests; SDaMCI, MDaMCI</td>
<td>ML (SDaMCI): 27.7 (1.6)</td>
<td>SDaMCI: 4.50 (0.93-8.07)</td>
<td>NA</td>
<td>SDaMCI: -0.1</td>
<td>MDaMCI: -0.2</td>
</tr>
<tr>
<td></td>
<td>49 (55%)</td>
<td>BL (SDaMCI): 79.4 (6.3)</td>
<td>BL (SDaMCI): 14.5 (3.9)</td>
<td></td>
<td></td>
<td>BL (SDaMCI): 27.6 (1.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BL (SDaMCI): 19</td>
<td>ML (MDaMCI): 72.6 (7.2)</td>
<td>ML (MDaMCI): 14.9 (3.3)</td>
<td></td>
<td></td>
<td>ML (MDaMCI): 72.6 (1.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(32%)</td>
<td>BL (MDaMCI): 72.6 (7.2)</td>
<td>BL (MDaMCI): 15.0 (3.3)</td>
<td></td>
<td></td>
<td>BL (MDaMCI): 72.6 (1.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ramakrishnan</td>
<td>ML: 22 (18.2%)</td>
<td>ML: 58.1 (11.4); 55.8 (12.2)</td>
<td>ML: 10.4 (3.7)</td>
<td>NA</td>
<td>Clinicians used Petersen criteria for final dx; MCI: amnestic MCI &amp; non-amnestic MCI</td>
<td>ML: 27.7 (1.8)</td>
<td>7.1 (2.36-11.84)</td>
<td>7.4 (2.46-12.34)</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>BL: 93 (20.4%)</td>
<td>BL: 65.2 (9.9); 63.2 (10.1)</td>
<td>BL: 15.5 (3.3)</td>
<td></td>
<td></td>
<td>BL (MDaMCI): 72.6 (1.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ML: monolinguals; BL: bilinguals; Mild Cognitive Impairment; SDaMCI: Single domain amnestic MCI; MDaMCI: Multiple domain MCI; LSBQ Language and Social Background Questionnaire; NA: not available
<table>
<thead>
<tr>
<th>Study</th>
<th>N (% of females)</th>
<th>Mean age of dx &amp; sx onset</th>
<th>Education level</th>
<th>Language measure</th>
<th>Dementia dx and type</th>
<th>MMSE/3MSE scores</th>
<th>Effect sizes (MD = BL age minus ML age, 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Age of dx</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alladi 2017</td>
<td>ML: 72 (52.8%)</td>
<td>BL: 121 (36.4%)</td>
<td>Dementia</td>
<td>ML: 6.9 (5.3)</td>
<td>Case records</td>
<td>MMSE, ACE-R, FrSiBE; bvFTD, PNFA, SD, FTD-MND, CBD, PSP</td>
<td>ML: 15.9 (10.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BL: 13.9 (4.3)</td>
<td></td>
<td></td>
<td>BL: 18.1 (10.2)</td>
</tr>
<tr>
<td>Alladi 2013</td>
<td>ML: 257 (49.0%)</td>
<td>BL: 391 (25.1%)</td>
<td>Dementia</td>
<td>ML: 5.9 (5.1)</td>
<td>Family member interview</td>
<td>DSM-IV; AD, VaD, mixed AD with CVD, FTD, DLB</td>
<td>ML: 16.7 (7.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BL: 12.9 (4.9)</td>
<td></td>
<td></td>
<td>BL: 18.9 (8.0)</td>
</tr>
<tr>
<td>Bialystok 2014</td>
<td>ML: 35 (54%)</td>
<td>BL: 40 (55%)</td>
<td>Dementia</td>
<td>ML: 12.5 (3.7)</td>
<td>LSBQ</td>
<td>NA; Probable AD</td>
<td>ML: 23.4 (3.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BL: 12.2 (4.9)</td>
<td></td>
<td></td>
<td>BL: 22.3 (4.5)</td>
</tr>
<tr>
<td>Bialystok 2007</td>
<td>ML: 91 (53%)</td>
<td>BL: 93 (59%)</td>
<td>AD onset</td>
<td>ML: 10.4 (10.0)</td>
<td></td>
<td></td>
<td>ML: 6.7 (10.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BL: 68.6 (9.6)</td>
<td></td>
<td></td>
<td>BL: 70.6 (9.6)</td>
</tr>
<tr>
<td>Chertkow 2010</td>
<td>ML: 379 (63%)</td>
<td>BL: 253 (51%)</td>
<td>AD onset</td>
<td>ML: 76.7 (7.8)</td>
<td></td>
<td></td>
<td>ML: 10.9 (3.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BL: 77.6 (7.2)</td>
<td></td>
<td></td>
<td>BL: 10.7 (3.8)</td>
</tr>
<tr>
<td>Clare 2014</td>
<td>ML: 49 (45%)</td>
<td>BL: 37 (57%)</td>
<td>AD onset</td>
<td>ML: 12.3 (3.04)</td>
<td></td>
<td></td>
<td>ML: 12.31 (3.04)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BL: 11.84 (2.46)</td>
<td></td>
<td></td>
<td>BL: 11.84 (2.46)</td>
</tr>
<tr>
<td>Craik 2010</td>
<td>ML: 109 (55%)</td>
<td>BL: 102 (59%)</td>
<td>AD onset</td>
<td>ML: 76.5 (10.0)</td>
<td></td>
<td></td>
<td>ML: 12.6 (4.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BL: 76.5 (7.9)</td>
<td></td>
<td></td>
<td>BL: 10.6 (5.1)</td>
</tr>
<tr>
<td>Lawton 2015</td>
<td>ML: 54 (65%)</td>
<td>BL: 27 (63%)</td>
<td>AD onset</td>
<td>ML: 81.10 (NA)</td>
<td></td>
<td></td>
<td>ML: 4.99 (4.17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BL: 7.93 (NA)</td>
<td></td>
<td></td>
<td>BL: 7.70 (4.88)</td>
</tr>
<tr>
<td>Perani 2017</td>
<td>ML: 40 (52.5%)</td>
<td>BL: 45 (71%)</td>
<td>AD onset</td>
<td>ML: 71.4 (4.9)</td>
<td></td>
<td></td>
<td>ML: 10.5 (4.07)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BL: 8.26 (4.55)</td>
<td></td>
<td></td>
<td>BL: 21.10 (4.84)</td>
</tr>
<tr>
<td>Schweizer 2012</td>
<td>ML: 19 (70%)</td>
<td>BL: 20 (70%)</td>
<td>AD onset</td>
<td>ML: 77.3 (6.8)</td>
<td></td>
<td></td>
<td>ML: 13.6 (3.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BL: 7.89 (7.7)</td>
<td></td>
<td></td>
<td>BL: 11.6 (4.5)</td>
</tr>
<tr>
<td>Woumans 2015</td>
<td>ML: 69 (69%)</td>
<td>BL: 65 (69%)</td>
<td>AD onset</td>
<td>ML: 72.5 (9.4)</td>
<td></td>
<td></td>
<td>ML: 13.5 (2.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BL: 77.3 (10.5)</td>
<td></td>
<td></td>
<td>BL: 14.7 (3.1)</td>
</tr>
</tbody>
</table>

Table 0-2: Cross-sectional studies on the relationship between bilingualism and dementia or AD.
Table 0-3 Longitudinal prospective study investigating the relationship between bilingualism and MCI

<table>
<thead>
<tr>
<th>Study characteristics</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study</td>
<td>N (% female); mean age (SD), Ed, MMSE/3MS</td>
</tr>
<tr>
<td>Zheng 2018</td>
<td>ML (Cantonese): 48 (85%); ML (Mandarin): 20 (45%); BL: 61 (57%)</td>
</tr>
<tr>
<td></td>
<td>ML (Cantonese): 4.92 (3.85); ML (Mandarin): 10.95 (3.28); BL: 10.79 (4.32)</td>
</tr>
<tr>
<td>Estanga 2016</td>
<td>N (% of females)</td>
</tr>
<tr>
<td>Early BL: 81 (54.3%)</td>
<td>ML: 57.82 (6.42); Early BL: 56.82 (6.48); Late BL: 57.56 (6.57)</td>
</tr>
<tr>
<td>Late BL: 97 (60.8%)</td>
<td>ML: 77.4 (6.7); BL: 77.0 (6.5); ESL: 77.1 (7.1)</td>
</tr>
</tbody>
</table>

ML: monolinguals; BL: bilinguals; dx: diagnosis; sx: symptom; AD: Alzheimer’s Disease; FTD: Frontotemporal Dementia; bvFTD: behavioral variant Frontotemporal Dementia; PNFA: Progressive Non Fluent Aphasia; SD: Semantic Dementia; FTD-MND: Frontotemporal dementia-motor neuron disease; CBD: Cortico Basal Degeneration; PSP: Progressive Supranuclear Palsy; VaD: Vascular Dementia; CVD: Cardiovascular Disease; DLB: Dementia with Lewy bodies; SNAP: Suspected Non-Alzheimer Pathophysiology; LSBQ Language and Social Background Questionnaire; LQ-SV Language Questionnaire – Short Version; ARSMA-II Acculturation Rating Scale for Mexican Americans; BLPQ Bilingual Language Profile Questionnaire; BAT Bilingual Aphasia Test; MMSE: Mini-Mental State Examination; 3MS: Modified Mini-Mental State; DSM-IV: Diagnostic and Statistical Manual of Mental Disorders Version Four; (ADDTC) NINCDS–ADRDA: (Alzheimer Disease Diagnostic and Treatment Centers) National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer’s Disease and Related Disorders Association; ICD-10: International Statistical Classification of Diseases and Related Health Problems; CDR: Clinical Dementia Rating; DSM-III-R: Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised; NA: not available
During a mean of 5.8 years (SD = 3.5) of annual follow-up evaluations, 396 individuals (41.1%) developed MCI. Higher levels (>4 years) of foreign language instruction: HR = 0.687, 95% CI: 0.482, 0.961

Table 0-4 Longitudinal prospective studies investigating the relationship between bilingualism and dementia

<table>
<thead>
<tr>
<th>Study</th>
<th>N (% female); mean age (SD), Ed, MMSE/3MSE</th>
<th>Language measure, dementia dx and type</th>
<th>% with dementia ML/BL</th>
<th>% no dementia ML/BL</th>
<th>RR, OR, HR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lawton 2015</td>
<td>81(64%)</td>
<td>ARSMA-II</td>
<td>ML: 54/1154</td>
<td>BL: 27/624</td>
<td>BL did not decrease the risk of dementia p = .72, AD p = .59, or VaD p = .53</td>
</tr>
<tr>
<td></td>
<td>4.99 (4.17), 3MSE 78.87 (9.90)</td>
<td>Baseline</td>
<td>ML: 102 (13.86%), 634 (86.14%)</td>
<td>BL: 10 (12.2%), 72 (87.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.70 (4.88), 3MSE 79.56 (15.57)</td>
<td>Follow-up</td>
<td>ML: 54/1154</td>
<td>BL: 27/624</td>
<td>BL did not decrease risk of dementia (p = .50) or AD (p = .36), even after adjusting for age and sex (p = .29)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ML: 102 (13.86%), 634 (86.14%)</td>
<td>BL: 10 (12.2%), 72 (87.8)</td>
<td></td>
</tr>
<tr>
<td>Ljungberg 2016</td>
<td>818 (51%) 73.6 (8.9)</td>
<td>Language History Questionnaire</td>
<td>ML: 54.9 (9.4%), 492 (85.4%)</td>
<td>BL: 6 (11.1%), 46 (85.2%)</td>
<td>Model 1: 1.06 (0.69, 1.63)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DSM-IV, NINCDS–ADRDA</td>
<td>ML: 54.9 (9.4%), 492 (85.4%)</td>
<td>BL: 6 (11.1%), 46 (85.2%)</td>
<td>Model 2: .13 (0.73, 1.79) Model 3: .7 (0.67, 1.72)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AD, VaD, LBD, FLD, PD, and UD</td>
<td>ML: 54 (9.4%), 492 (85.4%)</td>
<td>BL: 6 (11.1%), 46 (85.2%)</td>
<td>Model 4: (0.61, 1.59)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BL did not decrease risk of dementia (p = .50) or AD (p = .36), even after adjusting for age and sex (p = .29)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yeung 2014</td>
<td>576 (61.6%), 76.1 (6.2), 10.7 (2.8), 3MS 91.2 (5.7)</td>
<td>Self-report</td>
<td>ML: 198/637</td>
<td>BL: 86/344</td>
<td>Better self-rated bilingualism was associated with lower odds of dementia conversion. Each point on the self-report scale was associated with 0.291 lower odds of conversion to dementia</td>
</tr>
<tr>
<td></td>
<td>ML: 576 (61.6%), 76.1 (6.2), 10.7 (2.8), 3MS 91.2 (5.7)</td>
<td>ML: Dementia 9.4%, 3MS 91.2 (5.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BL: 54 (70.4%), 75.5 (5.6), 12.4 (4.9), 91.1 (5.6)</td>
<td>BL: Dementia 11.1, 3MS 91.1 (5.6)</td>
<td>BL: 198/637</td>
<td>BL: 86/344</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ESI: 360 (60.6%), 75.7 (6.2), 8.7 (3.5), 87.4 (6.9)</td>
<td>ESI: Dementia 9.7%, 3MS 87.4 (6.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ESI: 360 (60.6%), 75.7 (6.2), 8.7 (3.5), 87.4 (6.9)</td>
<td>ESI: Dementia 9.7%, 3MS 87.4 (6.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zahodne 2014</td>
<td>637 (72%), 75.66 (5.79), 5.05 (3.61)</td>
<td>Self-report (four-point Likert-type)</td>
<td>ML: 198/637</td>
<td>BL: 86/344</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BL: 430 (64%), 74.78 (5.66), 8.30 (4.22)</td>
<td>DSM-III</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BL: 430 (64%), 74.78 (5.66), 8.30 (4.22)</td>
<td>Probable and possible AD, VaD, LBD, and other dementias</td>
<td>Model 1: 1.06 (0.69, 1.63)</td>
<td>Model 2: .13 (0.73, 1.79) Model 3: .7 (0.67, 1.72)</td>
<td>Time 1 3MS, Time 2 3MS, and Change in the 3MS: Unadjusted model, English bilingual: Time 1, 0.6 (-1.8, 2.9), Time 2, 2.50 (-0.7, 5.7), Changed in 3MS, -1.7 (-4.2, 0.8)</td>
</tr>
</tbody>
</table>

SD: Standard Deviation; CI: Confidence Intervals; ML: monolinguals; BL: bilinguals; ESL: English as a Second Language; AD: Alzheimer’s Disease; VaD: Vascular Dementia; FLD: Frontal Lobe Dementia; PD: Parkinson’s Disease; DLB: Dementia with Lewy bodies; UD: Unspecified dementia; SNAP: Suspected Non-Alzheimer Pathophysiology; MMSE: Mini-Mental State Examination; 3MS: Modified Mini-Mental State; DSM-IV: Diagnostic and Statistical Manual of Mental Disorders Version 4.
Four; (ADDTC) NINCDS-ADRDA: (Alzheimer Disease Diagnostic and Treatment Centers) National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association; ICD-10: International Statistical Classification of Diseases and Related Health Problems; CDR: Clinical Dementia Rating; DSM-III: Diagnostic and Statistical Manual of Mental Disorders, Third Edition
Risk of bias for cross-sectional studies

Risk of bias for cross-sectional studies is presented in Table 5. Most cross-sectional studies employed acceptable sampling methods (k = 16; 100%), but most did not provide evidence for power calculations (k = 14; 88%). While some studies (k = 5; 31%) administered a validated measure of language ability, half of all studies (k = 8; 50%) used non-validated measures, including self- or proxy-reported measures (e.g., family member) or did not report the method of data collection (k = 2, 13%). Some studies (k = 14, 88%) controlled for important covariates such as immigration status and education either methodologically or statistically while others did not control for any covariates (k = 2, 13%).
Risk of bias for longitudinal studies

Risk of bias for longitudinal studies is presented in Table 6. All longitudinal studies employed poor sampling methods and either administered a language questionnaire, of which there was no mention of the psychometric properties, or they relied on self-report during a structured interview. Potential confounding factors including age, sex, and apolipoprotein E (APOE) ε4 allele status (Ljungberg et al., 2016); age, sex, and years of formal education (Wilson et al., 2015); age, sex, education, and subjective memory loss (Yeung et al., 2014); country of origin, gender, education, time spent in the current home country (United States of America), recruitment wave, and age at enrollment (Zahodne et al., 2014); occupation, education, baseline age, immigration status, APOE ε4 allele status, idea density, and grammatical complexity (Hack et al., 2019) were controlled for. Finally, all studies had

Table 5 Risk of bias for cross-sectional studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Selection</th>
<th>Comparability</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Representativeness of the sample</td>
<td>Sample size calculation</td>
<td>Non-respondents</td>
</tr>
<tr>
<td>Alladi 2017</td>
<td>–</td>
<td>–</td>
<td>★</td>
</tr>
<tr>
<td>Alladi 2013</td>
<td>–</td>
<td>–</td>
<td>★</td>
</tr>
<tr>
<td>Białystok 2014</td>
<td>–</td>
<td>–</td>
<td>★</td>
</tr>
<tr>
<td>Białystok 2007</td>
<td>–</td>
<td>–</td>
<td>★</td>
</tr>
<tr>
<td>Chertkow 2010</td>
<td>–</td>
<td>–</td>
<td>★</td>
</tr>
<tr>
<td>Craik 2010</td>
<td>–</td>
<td>–</td>
<td>★</td>
</tr>
<tr>
<td>Clare 2014</td>
<td>–</td>
<td>–</td>
<td>★</td>
</tr>
<tr>
<td>Estanga 2016</td>
<td>–</td>
<td>–</td>
<td>★</td>
</tr>
<tr>
<td>Lawton 2015</td>
<td>–</td>
<td>–</td>
<td>★</td>
</tr>
<tr>
<td>Ossher 2013</td>
<td>–</td>
<td>–</td>
<td>★</td>
</tr>
<tr>
<td>Perani 2017</td>
<td>★</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Ramakrishnan 2017</td>
<td>–</td>
<td>–</td>
<td>★</td>
</tr>
<tr>
<td>Schweizer 2012</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Woumans 2015</td>
<td>–</td>
<td>–</td>
<td>★</td>
</tr>
<tr>
<td>Yeung 2014</td>
<td>–</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Zheng 2018</td>
<td>–</td>
<td>–</td>
<td>★</td>
</tr>
</tbody>
</table>

A maximum of 10 stars can be given to each study
adequate assessments of the outcome including blind assessments for dementia diagnosis and appropriate follow-up periods, as well as reported sufficient information on attrition rate.

Meta-analyses of cross-sectional studies: age of symptom onset, diagnosis and disease severity at dementia diagnosis

Age at Alzheimer’s disease symptom onset

The mean difference between mono- and bilinguals at the age of AD symptom onset was 4.7 years (95% CI: 3.3, 6.1; Fig. 2). The t value was 8.06 with a two-tailed p < 0.001. Therefore, bilinguals were significantly older than monolinguals at the time of AD symptom onset. The Q-value was 6 with 6 df and with p = 0.424. Also, I² 0.00 in true and the variance in true effect sizes was T2 = 0.00, with T = 0.00.

<table>
<thead>
<tr>
<th>Study</th>
<th>Selection</th>
<th>Comparability</th>
<th>Outcome</th>
<th>Total/9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hack 2019</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>★ ★ ★ ★ ★ ★</td>
</tr>
<tr>
<td>Ljungberg 2016</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>★ ★ ★ ★ ★ ★</td>
</tr>
<tr>
<td>Wilson 2015</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>★ ★ ★ ★ ★ ★</td>
</tr>
<tr>
<td>Yeung 2014</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>★ – – ★ ★ –</td>
</tr>
<tr>
<td>Zahodne 2014</td>
<td>–</td>
<td>★</td>
<td>–</td>
<td>★ ★ ★ ★ – –</td>
</tr>
</tbody>
</table>

A maximum of 9 stars can be given to each study
Age at MCI, dementia, and AD diagnosis

Bilinguals were on average 3.2 years (95% CI: −3.4, 9.7; Fig. 3) older than monolinguals at MCI diagnosis. This mean difference was not statistically significant (t = 1.53, two-tailed p = .223). There was evidence that studies did not share a common effect size but that the true effects varied (Q = 8.91, df = 3, p = .031). Approximately 66% of the observed variance reflected the difference in true effect sizes rather than sampling error (I² = 66.34). The variance in true effect sizes was T² = 11.13, with T = 3.34. Bilinguals were on average 3.3 years (95% CI: 1.7, 4.9; Fig. 4) older than monolinguals at dementia diagnosis. This mean difference was statistically significant (t = 4.3, two-tailed p < .001). There was evidence that studies in this analysis likely did not share a common effect size but that the true effects varied (Q = 48.24, df = 12, p < .001). The I² was 75.12 indicating that approximately 75% of the observed variance reflected the difference in true effect sizes rather than sampling error. The variance in true effect sizes was T² = 4.83, with T = 2.20. The 95% PIs ranged from −1.9 to 8.5 years. Overall, in this analysis, we observed a high degree of heterogeneity.
Subgroup Analysis: Type of Diagnosis (Dementia vs. AD)

We conducted a post hoc subgroup analysis to explore the source for this heterogeneity. We compared studies including participants with AD to studies including participants with dementia. Bilinguals in the AD subgroup (k = 8; Fig. 5) were on average 4.2 years (95% CI: 2.0, 6.4) significantly older than monolinguals (t = 4.13, two-tailed p = .002). Bilinguals in the dementia subgroup (k = 5; Fig. 5) were on average 1.9 years (95% CI: −0.9, 4.7)
older than monolinguals, but this between-group difference was not statistically significant (t = 1.52, two-tailed p = .157). We also compared the mean difference for the AD and dementia subgroups to explore whether there were any significant differences between the two subgroups (Fig. 5). The mean difference in years (2.3, 95% CI: −5.9, 1.2) between the two subgroups was not statistically different (F = 2.04, df = 1, 11, two-tailed p = 0.181). The pooled estimates for heterogeneity in this subgroup analysis were T2 = 4.83, T = 2.20, I2 = 75.12, Q = 48.24, with df = 12, and p < .001.

<table>
<thead>
<tr>
<th>Group by Diagnosis</th>
<th>Study name</th>
<th>Statistics for each study</th>
<th>MD and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>BL mean age</td>
<td>ML mean age</td>
</tr>
<tr>
<td>AD</td>
<td>Bialystok 2014</td>
<td>78.2</td>
<td>70.9</td>
</tr>
<tr>
<td>AD</td>
<td>Cherklow 2010</td>
<td>77.6</td>
<td>76.7</td>
</tr>
<tr>
<td>AD</td>
<td>Clare 2014</td>
<td>79.3</td>
<td>76.2</td>
</tr>
<tr>
<td>AD</td>
<td>Craik 2010</td>
<td>80.8</td>
<td>76.5</td>
</tr>
<tr>
<td>AD</td>
<td>Perani 2017</td>
<td>77.1</td>
<td>71.4</td>
</tr>
<tr>
<td>AD</td>
<td>Schweizer 2012</td>
<td>78.9</td>
<td>77.3</td>
</tr>
<tr>
<td>AD</td>
<td>Womans 2015</td>
<td>77.3</td>
<td>72.5</td>
</tr>
<tr>
<td>AD</td>
<td>Zheng 2018</td>
<td>74.4</td>
<td>67.5</td>
</tr>
<tr>
<td>AD: total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dementia</td>
<td>Alladi 2013</td>
<td>68.1</td>
<td>63.4</td>
</tr>
<tr>
<td>Dementia</td>
<td>Alladi 2017</td>
<td>64.2</td>
<td>61.0</td>
</tr>
<tr>
<td>Dementia</td>
<td>Bialystok 2007</td>
<td>78.6</td>
<td>75.4</td>
</tr>
<tr>
<td>Dementia</td>
<td>Lawton 2015</td>
<td>81.1</td>
<td>79.3</td>
</tr>
<tr>
<td>Dementia</td>
<td>Ljungberg 2016</td>
<td>80.7</td>
<td>81.9</td>
</tr>
<tr>
<td>Dementia: total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. Forest plot showing the mean difference (MD) in the subgroup meta-analysis comparing studies including participants with AD to studies including participants with dementia on the age of AD and dementia diagnosis between bilinguals (BL) and monolinguals (ML): AD: Alzheimer’s disease; LL: lower limit; UL: upper limit; CI: confidence interval

Subgroup analysis: immigration status (adjusted vs. did not adjust for immigration)

We conducted a post-hoc subgroup analysis (Fig. 6) exploring whether immigration status was a potential source of heterogeneity. Bilinguals in studies adjusting for immigration status (k = 8) were on average 3.1 years (95% CI: 0.9, 5.2) older than monolinguals at dementia diagnosis (t = 3.17, two-tailed p = .009). In studies that did not adjust for immigration status (k = 5), bilinguals were on average 3.6 years (95% CI: 0.8, 6.5) older than monolinguals at dementia diagnosis (t = 2.97; two-tailed p = .018). The mean difference in years (0.5, 95%
CI: −4.1, 3.0) between the two subgroups was not statistically different (F = 0.10, df = 1, 11, two-tailed p = 0.752). The pooled estimates for heterogeneity in this subgroup analysis were T2 = 4.83, T = 2.20, I2 = 75.12, Q = 48.24, with df = 12, and p 1.

<table>
<thead>
<tr>
<th>Group by Immigration status</th>
<th>Study name</th>
<th>BL mean age</th>
<th>ML mean age</th>
<th>MD</th>
<th>LL</th>
<th>UL</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Lawton 2015</td>
<td>81.1</td>
<td>79.3</td>
<td>-1.8</td>
<td>-4.9</td>
<td>-1.3</td>
<td>0.257</td>
</tr>
<tr>
<td>No</td>
<td>Perani 2017</td>
<td>77.1</td>
<td>71.4</td>
<td>5.7</td>
<td>3.7</td>
<td>7.7</td>
<td>0.000</td>
</tr>
<tr>
<td>No</td>
<td>Schweizer 2012</td>
<td>79.9</td>
<td>77.3</td>
<td>1.6</td>
<td>-3.0</td>
<td>6.2</td>
<td>0.492</td>
</tr>
<tr>
<td>No</td>
<td>Wosniak 2015</td>
<td>77.3</td>
<td>72.5</td>
<td>4.8</td>
<td>1.4</td>
<td>8.2</td>
<td>0.005</td>
</tr>
<tr>
<td>No</td>
<td>Zheng 2018</td>
<td>74.4</td>
<td>67.5</td>
<td>6.9</td>
<td>3.6</td>
<td>10.2</td>
<td>0.000</td>
</tr>
<tr>
<td>Yes total</td>
<td>Alladi 2013</td>
<td>68.1</td>
<td>63.4</td>
<td>4.7</td>
<td>3.0</td>
<td>6.4</td>
<td>0.000</td>
</tr>
<tr>
<td>Yes total</td>
<td>Alladi 2017</td>
<td>64.2</td>
<td>61.0</td>
<td>3.2</td>
<td>0.4</td>
<td>6.0</td>
<td>0.023</td>
</tr>
<tr>
<td>Yes total</td>
<td>Bialystok 2007</td>
<td>78.6</td>
<td>75.4</td>
<td>3.2</td>
<td>0.6</td>
<td>5.8</td>
<td>0.014</td>
</tr>
<tr>
<td>Yes total</td>
<td>Bialystok 2014</td>
<td>78.2</td>
<td>70.9</td>
<td>7.3</td>
<td>2.8</td>
<td>11.8</td>
<td>0.001</td>
</tr>
<tr>
<td>Yes total</td>
<td>Chertkow 2010</td>
<td>77.6</td>
<td>76.7</td>
<td>0.9</td>
<td>-0.3</td>
<td>2.1</td>
<td>0.143</td>
</tr>
<tr>
<td>Yes total</td>
<td>Clare 2014</td>
<td>79.3</td>
<td>76.2</td>
<td>3.0</td>
<td>-0.4</td>
<td>6.4</td>
<td>0.080</td>
</tr>
<tr>
<td>Yes total</td>
<td>Cunha 2010</td>
<td>80.8</td>
<td>76.5</td>
<td>4.3</td>
<td>1.9</td>
<td>6.7</td>
<td>0.000</td>
</tr>
<tr>
<td>Yes total</td>
<td>Ljungberg 2016</td>
<td>80.7</td>
<td>81.9</td>
<td>-1.2</td>
<td>-5.0</td>
<td>2.6</td>
<td>0.537</td>
</tr>
</tbody>
</table>

Figure 6. Forest plot showing the mean difference (MD) in the subgroup meta-analysis comparing studies that had adjusted for immigration status to studies that had not adjusted for immigration status on the age of dementia diagnosis between bilinguals (BL) and monolinguals (ML); LL: lower limit, UL: upper limit; CI: confidence interval. Studies that had not adjusted for immigration status are categorized as No and studies that had adjusted for immigration status are categorized as Yes.

### Disease severity

There was no significant difference between mono- and bilinguals in disease severity at the age of dementia diagnosis (Hedges’ g = 0.05; 95% CI: −0.13, 0.24; t = 0.62, two-tailed p = .547; Fig. 7). The Q-value was 33.82 with df = 11 and p < .001. Approximately 67% (I2) of the observed variance reflected the difference in true effect sizes rather than sampling error. The variance in true effect sizes was T2 = 0.05 and T = .21. The PIs ranged between -0.47 and 0.57 MMSE points.
Meta-analysis of prospective longitudinal studies: Risk of dementia

We performed a meta-analysis on longitudinal prospective studies (Fig. 8). Results from this meta-analysis (k = 5) showed that bilingualism was not associated with a reduction in the risk of dementia (OR: 0.89; 95% CI: 0.68, 1.16, t = −1.22, two-tailed p = 0.289) when compared to monolingualism. There was no evidence of heterogeneity (Q = 3.22, df = 4, p = .522; I2 = 0.00; T2 = 0.00; T = 0.00).
Small-study effects

To address small-study effects, we generated funnel plots (Figs. 9 and 10). One funnel plot (Fig. 9) shows the observed (white dots) and imputed (black dots) effect sizes. Here, visual inspection showed that the observed data points tend to cluster on the right-hand side of the funnel plot indicating a minor asymmetry, suggesting the presence of small-study effects. However, Egger’s test was not significant with an intercept of 1.03 and CIs including −2.15 and 4.20 and with a t value of 0.71, df = 11, and a 1-tailed p value of 0.246. The Duval and Tweedie’s Trim and Fill test showed that the adjusted effect size (black diamond) would be 2.7 (95% CIs 1.3, 4.1) if the imputed studies had been included in the analysis. This indicates that even the adjusted effect size remained statistically significant.

Visual inspection of the second funnel plot (Figure 9) for the meta-analysis on disease severity (Fig. 10) showed a slight asymmetry on the right-hand side of the plot indicating a minor asymmetry, suggesting the presence of small-study effects. However, Egger’s test was not significant (one-tailed p value of 0.420) with an intercept of −0.267 (95% CI: −3.158, 2.623) and a t-value of 0.21 with df = 11. The Duval and Tweedie’s Trim and Fill showed that the adjusted effect size (black diamond) would be 0.05 (95% CI: −0.10, 0.21) if the imputed studies had been included in the analysis. Even in the likelihood of small-study effects or publication bias (De Bruin, Treccani, & Della Sala, 2015), the adjusted effect size remained similar to the observed effect size.
Figure 9 Funnel plot on age of dementia diagnosis

Figure 10 Funnel plot on disease severity
Discussion

While some studies have linked bilingualism to a delay in AD symptom onset and dementia diagnosis (Perani et al., 2017; Perquin et al., 2013; Schweizer et al., 2012), others have not reported such benefits (Mukadam et al., 2017). Some authors have argued that education and immigration status, among other confounders, may influence the relationship between bilingualism and dementia in cross-sectional studies (Mukadam et al., 2017). As such, further research is needed (Del Maschio et al., 2018; Grundy & Anderson, 2017; Woumans, Versijpt, Sieben, Santens, & Duyck, 2017).

Mild cognitive impairment

Meta-analytic results did not suggest that bilingualism delays the diagnosis of MCI. Due to the small number of included studies (k = 4) and small sample sizes (monolinguals n = 131; bilinguals n = 169), it is likely that this meta-analysis was underpowered and consequently, a type II error is possible (Hedges & Pigott, 2001). Studies had recruited participants with different types of MCI including single-domain amnestic and multiple-domain amnestic MCI (Ossher et al., 2012), amnestic and non-amnestic MCI (Ramakrishnan et al., 2017), or did not specify the subtype (Bialystok et al., 2014). Given the low number of included studies in this meta-analysis, we could not conduct a subgroup analysis to explore whether bilingualism was associated with a delayed diagnosis of MCI in relationship to the different subtypes of MCI. Of note, while MCI is a risk factor for dementia and AD, not all individuals with MCI will progress to AD or dementia (Albert et al., 2011). Therefore, the putative beneficial effects of bilingualism may be more salient at the beginning of the AD clinical spectrum rather than in milder forms of cognitive impairment such as MCI. Notably, a longitudinal study showed that foreign language instruction during childhood and adolescence lowered the risk of non-amnestic MCI but not amnestic MCI (Wilson et al., 2015), which supports some of the
primary cross-sectional studies (Bialystok et al., 2014; Ossher et al., 2012; Ramakrishnan et al., 2017).

**Age of AD Symptom Onset**

Our meta-analysis showed that bilinguals experienced AD symptoms on average 4.7 years later than monolinguals. While we did not observe significant heterogeneity, given the low number of studies (k = 7), caution in interpreting these findings as homogenous is warranted (Ioannidis, Patsopoulos, & Evangelou, 2007). These findings are in line with previous studies which show that speaking multiple languages is associated with better cognitive health in old age (Ihle, Oris, Fagot, & Kliegel, 2016; Kavé, Eyal, Shorek, & Cohen-Mansfield, 2008). Notably, the included studies did not provide a comprehensive profile of participants’ spoken languages, and because of this, we could not further investigate whether second-language proficiency, frequency of use, and age of acquisition played a moderating role in the observed delay in AD symptom onset (Del Maschio et al., 2018). When assessing AD symptom onset, researchers asked participants to retrospectively recall the age at which participants first began noticing AD symptoms. However, participants’ recall is often inaccurate and recall bias might have distorted participants’ reported estimates questioning its accuracy (Van den Bergh & Walentynowicz, 2016). In this meta-analysis (Fig. 2) studies tended to have small sample sizes (N median: 68.5) questioning the precision of the observed effect sizes (Cumming, 2014). Consequently, whether the estimate is close to the true value in this meta-analysis remains uncertain.

**Age of Dementia and AD Diagnosis**

Bilinguals were diagnosed with dementia on average 3.3 years later than monolinguals. According to the 95% PI (−1.9 to 8.5), we could expect that in some 95% of all populations
comparable to those in this meta-analysis (Fig. 4), the association between bilingualism and dementia may be strong, while in others, this association may be absent or may even tend to be in the opposite direction (Riley et al., 2011). Therefore, the beneficial association between bilingualism and delayed dementia diagnosis may appear only in some populations. While there are several possible explanations for wide Ps such as high risk of bias, we explored whether clinical differences across studies may be associated with the magnitude of the observed effect size in the meta-analysis in Fig. 4 (Sterne et al., 2011; Thompson, 1994). To address this, we conducted a post hoc subgroup analysis (Borenstein & Higgins, 2013; Oxman & Guyatt, 1992) comparing studies including participants with dementia (irrespective of etiology) to studies including participants with AD (specific etiology). In this analysis, bilinguals were not older than monolinguals at dementia diagnosis (mean difference: 1.9 years) but were 4.2 years older at AD diagnosis. Here, the between-subgroup mean difference did not differ. Low statistical power, as indicated by wide CI, the low number of studies per subgroup (dementia: $k = 5$; AD: $k = 8$), and a low sample sizes per study might explain the lack of difference in the dementia subgroup and in the between-subgroup analysis (Riley et al., 2011).

However, subgroup analyses are by default observational and because of this, we cannot be certain that participants in each subgroup were similar other than in the type of diagnosis. For example, the AD subgroup might have included a large portion of participants who could speak several languages and the dementia subgroup might have included bilinguals who spoke only two languages. Therefore, while bilinguals vs. monolinguals were older at AD but not at dementia diagnosis, we cannot be certain that this was due to differences in the type of diagnosis, and that findings should only be interpreted as hypothesis-generating (Thompson & Higgins, 2002).
**Risk of dementia**

The meta-analysis including prospective studies showed no significant risk reduction in developing dementia among bilinguals compared to monolinguals. Our effect size favored bilinguals more than the effect size from the previous meta-analysis [odds ratio: 0.89; 95% CI: 0.68–1.16; (Mukadam et al., 2017)]. From our systematic review, we decided to exclude one study because it did not clearly define its control group as monolingual (Sanders et al., 2012), but it was included in the previous meta-analysis (Mukadam et al., 2017). The difference in the included studies between our and the previous meta-analysis might explain the difference in the magnitude of the effect sizes. Moreover, while results showed no risk reduction in dementia among bilinguals, the trend favoring bilinguals in our meta-analysis (Fig. 7) needs to be carefully considered. Given the low number of studies (k = 5), our meta-analysis might not have reached sufficient statistical power to detect a true effect (Hedges & Pigott, 2001). The CIs in each study were relatively wide indicating low statistical power and poor precision (Cumming, 2014). Therefore, each study was also likely under-powered to detect a true effect, if such an effect existed (Ioannidis, 2005, 2008). We did not find evidence of heterogeneity in this meta-analysis. Given the low number of studies in this meta-analysis, the Q statistic was likely underpowered, however. Notably, lack of heterogeneity does not necessarily indicate homogeneity (Ioannidis et al., 2007); interpreting a non-significant heterogeneity test in a meta-analysis with few studies is problematic (Rücker, Schwarzer, Carpenter, & Schumacher, 2008).

**Possible mechanisms and disease severity at dementia diagnosis**

Some authors have argued that while cross-sectional studies generally tend to show a later dementia diagnosis for bilinguals vs. monolinguals, these studies are more susceptible to the confounding effects of education or cultural differences (Mukadam et al., 2017). Given that
our meta-analyses included studies that had adjusted for education, it is unlikely that education had confounded the observed delays in dementia and AD diagnoses among bilinguals. We also conducted a subgroup meta-analysis comparing studies that had adjusted for immigration to those that did not explicitly mention participants’ immigration status. This analysis found that bilinguals were older than monolinguals at dementia diagnosis regardless of subgroup membership. This suggests that immigration might not have played a role in delaying the age of dementia diagnosis in bilinguals relative to monolinguals in these studies. It is noteworthy to highlight that while some studies had mentioned participants’ migration status, it was occasionally problematic to discern whether authors had in fact adjusted for migration status because there was no statement explicitly addressing the analytical approach for adjusting for this variable.

Even if bilinguals were delaying seeking medical attention due to cultural differences, we would still expect them to demonstrate greater cognitive impairment than monolinguals at dementia diagnosis. However, we found no difference (Hedges’ g = 0.05, 95% CI: −0.13, 0.24) between mono- and bilinguals on disease severity at dementia diagnosis. This suggests that in some settings, bilingualism may be more beneficial than monolingualism to help maintain cognitive function for a longer period of time despite the presence of ongoing neuropathology (Gold, 2015). It is possible that bilingualism may help in maintaining cognitive health for a longer period of time, protecting against the impending effects of AD on cognition (Bak et al., 2014; Gold, 2015). There is evidence to suggest that bilingualism is associated with higher cognitive function in old age (Ihle et al., 2016; Kavé et al., 2008) even after adjusting for differences in intelligence levels during childhood (Bak et al., 2014). Some authors have advanced the proposition that bilingualism may enhance cognitive reserve, which refers to the ability to maintain functioning levels of cognition despite the presence of a neurodegenerative disease such as AD (Perquin et al., 2013; Stern, 2012).
Supporting findings from our meta-analysis on disease severity, behavioral data indicated that mono- and bilinguals did not significantly differ in executive functions at AD diagnosis despite bilinguals being significantly older (Bialystok et al., 2014). Computed tomography scans also revealed greater atrophy of the medial temporal lobe at AD diagnosis in bilinguals vs. monolinguals matched for disease severity and despite monolinguals having higher education and job status (Schweizer et al., 2012). The medial temporal lobe is a region particularly affected by AD (Clerx et al., 2013; Visser et al., 2002). Moreover, bilinguals showed greater cerebral hypometabolism than monolinguals, which is indicative of greater neurodegeneration, and outperformed monolinguals in short- and long-term verbal and visuospatial memory, but not in language tasks (Perani et al., 2017). Given the disagreement in the field regarding the exact underlying mechanisms of bilingualism thought to promote cognitive reserve (García-Pentón et al., 2016a, 2016b), we are currently conducting a systematic review investigating the underlying brain mechanisms of bilingualism in non-clinical and clinical individuals with MCI or dementia (Brini et al., 2018a).

Because studies did not generally measure participants’ socioeconomic status, it was not possible to examine whether this factor might have contributed to the observed delays in dementia diagnosis among bilinguals. The incidence of dementia is higher in certain ethnic minorities than in Caucasian individuals (Mehta & Yeo, 2017), suggesting that socioeconomic and cultural factors may play a role in the observed relationship between bilingualism and dementia. Researchers have extensively debated how to quantify bilingualism (Del Maschio et al., 2018; Luk & Bialystok, 2013). While studies have tended to categorize participants between mono- and bilinguals (Del Maschio et al., 2018), bilingualism is a multidimensional variable that extends on a continuum (Luk & Bialystok, 2013). For example, factors including the number of languages one can speak, age of acquisition, proficiency, and frequency of use in the second language likely interact with one
another and may explain differences in the observed delay in dementia among bilinguals (Del Maschio). However, the studies included in our meta-analysis did not formally assess these factors (Table 5–6), and because of this, we could not examine whether the different dimensions of bilingualism (Del Maschio et al., 2018; Luk & Bialystok, 2013) contributed to the observed heterogeneity in some of our meta-analyses.

Sources of uncertainty and risk of bias in cross-sectional and prospective longitudinal studies

From our risk of bias assessment within studies, it is clear that one major source of uncertainty concerned how representative the samples were and whether the exposure had been measured appropriately (Table 5). In cross-sectional studies, because no study formally assessed monolingualism, the extent to which participants were truly monolinguals remains unclear. This would have been an important factor to assess because exposure to foreign languages through schooling or the media is ubiquitous nowadays (Laine & Lehtonen, 2018), questioning whether the monolinguals in our included studies were in fact, truly monolinguals. Generally, bilingualism was poorly defined, measured, or did not carry a specific definition. While researchers commonly defined bilingualism as “speaking two or more languages,” they did not routinely measure additional languages. This would have been a relevant factor to measure because some studies point to a dose-response relationship (Antoniou & Wright, 2017) with increasing number of languages generating a greater delay in the onset of dementia (Alladi et al., 2013; Chertkow et al., 2010; Clare et al., 2016), protection against cognitive impairment (Perquin et al., 2013), and greater cognitive health in older individuals (Ihle et al., 2016).

The assessment of participants’ language profiles and by extension, their representativeness, was also questionable in longitudinal prospective studies (Table 6). In one study, bilingualism was not associated with reduced dementia risk but those reporting
speaking a second language very well had a 14% lower risk of developing dementia than those who reported not at all well (Zahodne et al., 2014). This questions whether participants who reported speaking a second language “not at all well” should have been classed as bilinguals or monolinguals and supports the notion that participants’ language profiles should be treated as a continuous rather than a dichotomous variable (Luk & Bialystok, 2013). While other studies did not show a risk reduction in dementia among bilinguals, it is likely that they were underpowered. For example, one study included 736 monolinguals but only 82 bilinguals with 102 developing dementia in the monolingual group and 10 among bilinguals (Ljungberg et al., 2016). Furthermore, only three studies adopted a questionnaire to measure bilingualism (Hack et al., 2019; Ljungberg et al., 2016; Zahodne et al., 2014). Thus, differences in the operationalization and measurement of bilingualism, as well as relatively small sample sizes, question the internal validity of the longitudinal prospective studies (IntHout, Ioannidis, Borm, & Goeman, 2015).

Further, while some cross-sectional studies adjusted for important confounders such as education and occupation status, several studies did not specify whether they had adjusted for other likely confounders nor did they routinely report participants’ immigration status (Fuller-Thomson, 2015; Fuller-Thomson & Kuh, 2014) or levels of physical activity. Levels of physical activity may be an important factor to assess as bilingualism may benefit sedentary individuals (Brini et al., 2018b) differently than highly physically active individuals (Sterne et al., 2011). Since the majority of studies did not provide evidence of power calculations, it was unclear whether they had sufficient statistical power to detect an effect if one existed particularly when adjusting for genes (Ioannidis, 2008; Sham & Purcell, 2014). For example, bilingualism may benefit participants with the apolipoprotein E (APOE) ε4 allele, the main genetic risk factor for AD (Galimberti & Scarpini, 2016; Liu, Kanekiyo, Xu, & Bu, 2013), differently than those without the APOE ε4.
Similar to the cross-sectional studies, longitudinal prospective studies did not consider participants’ baseline risk of dementia. Although one study adjusted for the APOE ε4 (Ljungberg et al., 2016), which is a risk factor for AD (Brini, Sohrabi, et al., 2018b), no other prospective study considered other genetic risk factors implicated in AD (Naj, Schellenberg, & Consortium, 2017). Only 147 participants (across mono- and bilinguals) were APOE ε4 carriers (Ljungberg et al., 2016). Of note, whether APOE ε4 increases the risk of vascular dementia (Rohn, 2014), frontotemporal dementia (Verpillat et al., 2002), dementia with Lewy bodies (Lovati et al., 2010), and Parkinson’s disease (Fagan & Pihlstrøm, 2017) is unclear (Lovati et al., 2010). Therefore, adjustment for the APOE ε4 likely did not reach sufficient statistical power (Sham & Purcell, 2014) in this study (Ljungberg et al., 2016) and its clinical relevance to other dementia etiologies may have been limited (Lovati et al., 2010). The authors also did not analyze other variants of the APOE including the ε2, which may confer protection against AD (Liu et al., 2013).

*Small-study effects*

While our funnel plots showed slight asymmetry indicating possible small-study effects (Sterne, Egger, & Smith, 2001), Egger’s tests were not significant. However, results from Egger’s test should be interpreted with caution, because in the absence of severe bias, this test has low statistical power (Sterne, Gavaghan, & Egger, 2000). One possible source of small-study effects is publication bias (Egger, Smith, Schneider, & Minder, 1997), which is prevalent in the social (Franco, Malhotra, & Simonovits, 2014) and cognitive sciences (Ioannidis, Munafo, Fusch Poli, Nosek, & David, 2014). More notably for this systematic review, it is likely present in the field of bilingualism research, too (De Bruin et al., 2015); although others (Bialystok, Kroll, Green, MacWhinney, & Craik, 2015) have contested these findings (De Bruin et al., 2015). Therefore, despite the non-significant Egger’s tests, there are
reasons to believe that publication bias may be present in this field of research. In light of this, the Duval and Tweedie’s Trim and Fill test (Duval & Tweedie, 2000) showed that after imputing the estimated missing studies, bilinguals would be on average 2.7 years (95% CI: 1.3, 4.1) older than monolinguals at the time of dementia diagnosis. Even in the likelihood of publication bias (De Bruin et al., 2015), the observed effect size in this meta-analysis (Figure 4) would not change by a large margin.

Several of our included studies had small sample sizes, which can increase the risk of type I error (Ioannidis, 2005) and inflate the effect size (Ioannidis, 2008), which can result in funnel plot asymmetry (Sterne et al., 2011). For example, if the association between bilingualism and dementia is driven by a dose-response relationship (Alladi et al., 2013; Chertkow et al., 2010; Clare et al., 2016; Ihle et al., 2016), smaller studies with a higher portion of multilingual participants may generate greater effect sizes resulting in funnel plot asymmetry (Egger et al., 1997). Moreover, as mentioned previously, bilingualism may benefit participants who occupy a higher baseline risk of dementia (e.g., by virtue of genetic risk) differently, which could also explain funnel plot asymmetry (Sterne et al., 2011).

However, because most studies did not report data on the number of spoken languages or participants’ baseline dementia risk, we could not explore whether multilingualism may have contributed to funnel plot asymmetry.

**Limitations**

A limitation of our meta-analyses was the inclusion of all cross-sectional studies regardless of language status. Most studies did not precisely report how many languages were spoken by their bilingual cohorts. Therefore, from our meta-analyses, it remains unclear whether the number of languages a person can speak plays a role in delaying the onset of dementia. As noted above, however, some evidence suggests that the number of languages could play a
role in the risk and delay of dementia (Chertkow et al., 2010; Clare et al., 2016).

Additionally, in a subgroup meta-analysis, we compared studies that recruited participants with dementia and AD. In the dementia subgroup, however, participants were diagnosed with different forms of dementia. This is a limitation because, from this subgroup, it was not possible to discern whether bilingualism was distinctively related to different dementia etiologies. Furthermore, this subgroup analysis was likely underpowered given the small number of studies (k = 5) and the associated large CIs.

While the results of our meta-analyses on the age of dementia and AD diagnosis are interesting, it is crucial to stress that the observed relationship between bilingualism and dementia is not causal. Cross-sectional studies are useful when examining the relationship between two variables and help to generate hypotheses that may be further tested for causal effects in experimental studies. Particularly for this review, our risk of bias assessment uncovered several sources of uncertainty due to bias within studies. For example, several factors such as the poor measurement of bilingualism, the inclusion of varying types of dementia etiologies, and lack of control over confounding factors in several of the included studies, leave us questioning the beneficial link between bilingualism and dementia.

Most studies did not report how bilinguals had acquired the second or third language, or participants’ immigration status. This is a limitation in our meta-analyses because some participants might have acquired the second language through schooling whereas others might have acquired it due to migrating to a new country. In the former case, participants might have been diagnosed in their native language (e.g., English) whereas in the latter case, participants might have been diagnosed in their non-native language (e.g., a language other than the recipient country’s national language). As such, cultural differences (Chandra et al., 2001; Chin et al., 2011) or language barriers (Lindesay, 1998; Nielsen et al., 2011) might have contributed to the observed delays in dementia diagnosis and possibly confounded the
relationship between bilingualism and age of dementia diagnosis. Since there was insufficient information regarding what language was used to provide a diagnosis of MCI or dementia among bilinguals, we could not further explore whether the language of the assessment played a role in the observed delays in any of our outcomes.

**Strengths**

Unlike a previous systematic review (Mukadam et al., 2017), results from cross-sectional studies were meta-analyzed to determine whether bilingualism is associated with a delayed onset of dementia and AD. This allowed us to generate a more precise estimate of the effect size. In response to previous criticisms (Fuller-Thomson, 2015; Fuller-Thomson & Kuh, 2014; Mukadam et al., 2017), we explored whether immigration status might have been related to differences in the age of dementia diagnosis by conducting a sub-group meta-analysis. We have included more recently published studies that had not been included in the previous systematic re-view (Mukadam et al., 2017) and therefore, provide a more up to date review of the available literature. We also registered a study protocol a priori for this systematic review.

**Suggestions for future research**

Given the lack of a standard definition and measurement tool for mono- and bilingualism across our included studies, it is critical for future research to improve the measurement of participants’ language profile. This could mean quantifying the spectrum of language knowledge on a continuum and by measuring proficiency, frequency of use, and the age of second language acquisition (Anderson, Mak, Chahi, & Bialystok, 2018; Li, Zhang, Yu, & Zhao, 2019; Luk & Bialystok, 2013). Researchers could then apply multiple linear regression (Plonsky & Oswald, 2017) or Bayesian inference (Ross & Mackey, 2015) to explore whether
language skills can predict the age of dementia symptom onset and diagnosis. Researchers can apply objective measures for bilingualism (Clare et al., 2016; Estanga et al., 2017), rather than asking participants or family members to self-report language status (Alladi et al., 2013; Chertkow et al., 2010). Formal assessments of second language proficiency while treating the degree of bilingualism as a continuous variable (DeLuca, Rothman, Bialystok, & Pliatsikas, 2019; Laine & Lehtonen, 2018; Luk & Bialystok, 2013) should be applied. Authors have recently developed questionnaires to quantify participants’ language profiles on a continuum (Anderson et al., 2018) and to measure language proficiency, dominance, as well as immersion (Li et al., 2019). Researchers might want to establish a priori whether they wish to measure bilingualism, the ability to speak two languages (Anderson et al., 2018) or multilingualism, the ability to speak three or more languages (Li et al., 2019). Clearly reporting participants’ immigration status will also be beneficial.

Increasing statistical power will enable partitioning of participants into different dementia etiologies (Nelson et al., 2019) and to conduct sub-group analyses. While categorizing participants into dementia subtypes poses several challenges (De Reuck et al., 2016), applying biomarkers could help researchers in classifying dementia subtypes (Jack et al., 2016; Perneckzy et al., 2016). Future studies should consider adjusting for variables such as physical activity, which is associated with the risk of dementia (Brini, Sohrabi, et al., 2018b). Researchers wishing to adjust for genetic risk would need to recruit a large number of mono- and bilinguals to reach sufficient statistical power for this type of analysis (Sham & Purcell, 2014) and exclude participants for which the APOE ε4 may not be clinically relevant (Lovati et al., 2010). Additionally, while bilinguals vs. monolinguals were older at dementia diagnosis, the observed delay does not imply disease-modifying effects (Galimberti & Scarpini, 2016). Combining behavioral measures with surrogate biomarkers such as brain data will provide more robust evidence as to whether bilingualism can help maintain
cognitive function despite presence of neuropathology due to dementia (Bialystok, Anderson, & Grundy, 2018; Brini, Sewell, et al., 2018a) and could reveal potentially disease-modifying properties (Galimberti & Scarpini, 2016). Researchers wishing to conduct conditional power calculations for future studies based on our meta-analyses need to take into consideration heterogeneity when estimating sample size (Roloff, Higgins, & Sutton, 2013). Finally, to enhance reporting for observational studies, authors should follow and clearly state that their study adhered to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement (Von Elm et al., 2014).

**Implications and conclusion**

Identifying factors that can delay the onset of dementia and AD is a major public health priority (Winblad et al., 2016; Wortmann, 2012). This is because, a delay in the onset of AD of five years could reduce AD prevalence by 57% with concomitant savings of US$627 to US$344 billion in Medicare costs worldwide (Sperling et al., 2011). At the individual level, delaying the symptom onset of dementia and AD can also have important benefits for patients, families, and, by implication, the overall incidence of AD (Cummings, Morstorf, & Zhong, 2014). Our findings suggest that speaking two or more languages may be related to an ability to maintain functional cognition for a longer time compared to monolinguals. The observed effect sizes may be superior, under certain settings, to available pharmacological therapies that delay cognitive decline by 6–12 months and only target symptoms without modifying the pathogenic or clinical course of AD (Yiannopoulou & Papageorgiou, 2013).

While bilingualism appears to be associated with delayed AD symptom onset, dementia and AD diagnosis, the substantial heterogeneity and several sources of bias challenge the interpretation of our findings. Until future studies improve the measurement of participants’ language profiles, increase sample sizes, comprehensively report sample characteristics including participants’ ethnicity and birthplace, adjust for
baseline dementia and AD risk (separately), it will be problematic to discern under which settings and to what extent bilingualism may be beneficial. Precisely because of these unanswered questions, we think it is premature to remove public health policy recommendations on bilingualism as a strategy to delay dementia as previously suggested (Mukadam et al., 2017). We also disagree that longitudinal prospective studies were “large high quality prospective studies” (Mukadam et al., 2017). We argue that longitudinal prospective studies were likely underpowered and carried serious methodological limitations and that, it is incorrect to conclude evidence of no effect (Mukadam et al., 2017) from no evidence of an effect (Schünemann et al., 2019). Given that the observed effect sizes may be superior to available pharmacological therapies (Yiannopoulou & Papageorgiou, 2013), we agree with others that researchers should improve study methodology and continue investigating the link between bilingualism and dementia (Del Maschio et al., 2018).

**Conflict of Interest**

RNM is the co-Founder of the KaRa Institute of Neurological Diseases and has stocks on Alzhyme Ltd. HRS has received/is receiving remunerations from Australian Alzheimer's Research Foundation for working on several clinical trials associated with Takeda, Merck and AstraZeneca and Eli Lilly.
References


mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders*, 44(3-4), 222-231.


Ross, S. J., & Mackey, B. J. L. L. (2015). Bayesian approaches to imputation, hypothesis testing, and parameter estimation. 65(S1), 208-227.


Between-chapter (3-4) rationale

In Chapter 3, we found that bilinguals demonstrated a later onset of AD symptoms and diagnosis. The results indicate that bilingualism may be a potential factor that could be modified to delay the onset of AD in the community. Chapter 3 also revealed several sources (e.g. poor measurement of participants’ language profiles) of uncertainty which led us questioning whether our estimates are close to the true value. Given these uncertainties and to gather additional evidence to determine whether bilingualism may be used as a potentially modifiable risk factor for delaying AD, we sought to explore the underlying brain mechanisms implicated in bilingualism. We argued that bilingualism may render the brain areas affected by AD more resilient against neuropathology thereby also delaying the clinical manifestation of AD.
Chapter 4 The bilingual brain in healthy and neuropathological aging: a systematic review and brain mapping of neuroimaging studies.
Authors: Stefano Brini,a,b,c,d, Behnaz Shahtahmassebi,c, Hamid, R. Sohrabi,e,f,g, Mira Karrasch,h, Ralph N. Martins,e,f,g, Timothy J. Fairchild,e,

aDepartment of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland
bTurku Brain and Mind Center, Turku, Finland
cSchool of Psychology and Exercise Science, Murdoch University, Perth, Western Australia, Australia
dHealth Services Research and Management School of Health Sciences, City, University of London, London, United Kingdom
eSchool of Medical and Health Sciences, Edith Cowan University, Perth, Western Australia, Australia
fDepartment of Biomedical Sciences, Macquarie University, New South Wales, Australia
gAustralian Alzheimer’s Research Foundation, Perth, Western Australia, Australia
hDepartment of Psychology, Åbo Akademi University, Turku, Finland

Highlights

1. We systematically reviewed neuroimaging findings in aged bilinguals
2. We identified candidate brain regions protecting bilinguals against dementia
3. We identified several sources of uncertainty and inconsistent findings in the field
4. We provide future research directions on the research of bilingual brain in dementia
Abstract

There is evidence showing that speaking more than one language (bilingualism) as opposed to one (monolingualism) is associated with a delay in the diagnosis of dementia up to four years. However, it is not yet known what brain mechanisms may be responsible for supporting the observed delays in dementia diagnosis among bilinguals. Here, we systematically reviewed the literature assessing the relationship between bilingualism and the brain. We searched for studies comparing monolinguals and bilinguals on structural and functional neuroimaging outcomes. Our target population included healthy adults and those diagnosed with dementia. We applied brain mapping to synthesize the available evidence and show what brain areas related to bilingualism had been investigated in previous studies.

Results from our brain mapping showed that bilingualism is associated with stronger frontostriatal and frontoparietal circuits in healthy adults and those diagnosed with dementia, which are brain areas severely affected by dementia. Our findings indicate that bilingualism may strengthen these brain areas and render the brain more resilient against neuropathological changes present during natural course of dementia. However, risk of bias revealed several sources of uncertainty questioning the interval validity of the included studies.
Introduction

Identifying protective strategies against late onset dementia has become a global public health priority (Winblad et al., 2016; Wortmann, 2012). This is because age is a major risk factor for dementia and with increased global life expectancy (Kontis et al., 2017) the number of individuals expected to develop dementia will significantly increase (Prince et al., 2013). Projections from the World Health Organization (WHO) suggest the number of individuals aged 65 years old and over will rise from 524 million in 2010 to 1.5 billion by the year 2050, with a concomitant increase in the number of people with dementia from 35.6 million in 2010 to 115.5 million by the year 2050 (Prince et al., 2013). The worldwide costs of dementia are estimated to reach two trillion United States dollars by the year 2030 (Wimo et al., 2017).

Modifying particular everyday behaviours such as increasing physical activity (Brini et al., 2018) and attaining higher levels of education (Xu et al., 2016) are linked to maintenance of cognitive health across the lifespan and a lower risk of dementia. It has been suggested that modifying these behaviours can result in greater cognitive reserve (CR), which is the ability to maintain cognitive functioning in the presence of neuropathology (Stern 2012). Another everyday behaviour, bilingualism (the ability to speak two or more languages), has also been linked with greater CR (Gold, 2016), which is thought be responsible for the observed delayed in the onset of dementia among bilinguals (Bialystok, Abutalebi, Bak, Burke, & Kroll, 2016). While it has been suggested that bilingualism might increase CR and render the brain more robust against neuropathology, the specific mechanisms underlying this process in the bilingual brain are still unclear (García-Pentón et al., 2016a, 2016b).

Some behavioural evidence shows that bilinguals are older than monolinguals at the time of dementia diagnosis by up to five years but demonstrate similar global cognitive impairment as monolinguals (Gold, 2016). One study also found that bilinguals were on
average five years older than monolinguals at the time of Alzheimer's disease (AD) diagnosis, and they also demonstrated greater impairment in cerebral glucose uptake, which is a marker of neurodegeneration (Perani et al., 2017). Moreover, a study in participants with AD who were matched on disease severity, cognitive performance, and years of education demonstrated that several neural indices associated with AD were more prominent in bilinguals as compared to monolinguals (Schweizer et al., 2012). More specifically, temporal lobe atrophy as measured by the radial width of the temporal horn and the temporal horn ratio was more pronounced in bilinguals as compared to monolinguals (Schweizer et al., 2012). These parallel lines of research suggested that bilingualism might enhance CR, thereby allowing bilinguals to cope better with advancing neuropathology than monolinguals, at least up to a certain point of the disease progression.

The exact mechanisms responsible for this link however have not been systematically explained; although interesting hypotheses have been developed. For example, bilinguals learn to engage the context-appropriate language (Bialystok, Anderson, & Grundy, 2018) while inhibiting the context-inappropriate language without considerable effort (Calabria, Costa, Green, & Abutalebi, 2018; Kroll et al., 2014). While certain cognitive processes (e.g. adopting novel strategies and developing a routine through extensive practice in language selection) could alleviate the cognitive load arising from this linguistic conflict, bilinguals may still experience a higher cognitive load needed to resolve the struggle arising from the linguistic conflict (Calabria et al., 2018; Costa & Sebastián-Gallés, 2014). Therefore, we could expect the resultant increase in cognitive load to affect frontostriatal and frontoparietal circuits, which interestingly tend to be more structurally and functionally developed in bilinguals than in monolinguals (Li, Legault, & Litcofsky, 2014). For example, bilinguals had more grey matter in the parietal lobe (Abutalebi et al., 2015) and in the left anterior temporal lobe (Abutalebi et al., 2014) compared to monolinguals matched for level of general
cognitive impairment. Because other CR-related factors had been held constant, differences in the brain were attributed to bilingualism. Also, higher proficiency in the second language positively correlated with brain volume in the left anterior temporal lobe (Abutalebi et al., 2014). Moreover, despite similar performance in executive functioning, older bilinguals did not show the canonical posterior-anterior shift in aging in brain activity (Ansaldo et al., 2015) that has been associated with brain degeneration (PASA, Davis, Dennis, Daselaar, Fleck, & Cabeza, 2007). More specifically, bilinguals did not recruit prefrontal cortex as much as monolinguals, but instead showed higher activity in the posterior parietal cortex.

Conveniently, the language-related brain regions meeting higher demands in bilinguals vs. monolinguals are also responsible for both verbal and non-verbal executive functions (Abutalebi & Green, 2007; Garbin et al., 2010; Prior & Gollan, 2011). Both language processing and executive functions emerge from the interaction of complex brain networks that are highly dependent on one another (Salmi, Nyberg, & Laine, 2018). For example, while the prefrontal cortex plays a key role in several executive functions, parietal regions (e.g. inferior parietal lobule) are likely also important in tasks assessing updating in the domain of executive functions (Collette et al., 2007). Given that frontostriatal and frontoparietal circuits also responsible for processes involving executive functions, the constant monitoring of each language in bilinguals may indirectly enhance executive functions (Bialystok, Anderson, & Grundy, 2018). Indeed, there is some evidence that older bilinguals outperform monolinguals in conflict resolution (Bialystok & Craik, 2010), although this relationship may not be observed in younger bilinguals (Lehtonen et al. 2018).

Whether bilingualism can promote neuroplasticity and enhance CR is, however, still highly controversial. While others clearly argue for the benefits of bilingualism (Gold 2015; Grundy, Anderson, & Bialystok, 2017), there are several others who disagree (García-Pentón et al., 2016a, 2016b; Gold, Johnson, et al., 2013; Olsen et al., 2015). Importantly, previous
reviews on the bilingual brain in healthy aging and dementia have not been systematic nor
have pulled together the vast array of neuroimaging methods (García-Pentón et al., 2016a
2016b; Grundy, Anderson, & Bialystok, 2017). Moreover, while some argue that findings in
this field are largely consistent (Grundy, Anderson, & Bialystok, 2017), studies have
produced mixed findings with some showing that bilingualism might be related to changes in
frontostriatal and frontoparietal circuits as already mentioned and others showing no such
relationship (García-Pentón, Fernández García, Costello, Duñabeitia, & Carreiras, 2016a,
2016b; Gold, Johnson, & Powell, 2013; Li, Legault & Litcofsky, 2014; Olsen et al., 2015). In
particular, no previous review has systematically explored how the observed differences in
frontostriatal and frontoparietal circuits among bilinguals may translate into protective effects
against dementia. Also, small sample sizes as well as varying methods used in measuring
bilingualism and brain functions may partially explain the variability in study outcomes
(García-Pentón et al., 2016a, 2016b; Gold, Johnson, et al., 2013; Olsen et al., 2015). As such,
findings from single studies need to be interpreted with circumspection, and the available
evidence should be systematically reviewed (Bialystok, Anderson, & Grundy, 2018).

Objectives

There is some evidence linking bilingualism to differences in grey matter density in both
whole brain and regional white matter integrity across the lifespan (Li, Legault & Litcofsky,
2014). Others have identified candidate systems linked to bilingualism including the inferior
parietal cortex (Age of Acquisition, AoA and proficiency of the second language or L2),
temporal pole (L2 picture naming), anterior cingulate gyrus (conflict monitoring), inferior
prefrontal cortex (AoA), caudate nucleus (language switching), and cerebellum (procedural
memory; Li, Legault & Litcofsky, 2014). It is plausible that the brain areas linked with
bilingualism in healthy cohorts overlap with those involved in the neuropathology of
dementia, and development of these areas in the bilingual brain may directly promote resilience against dementia. In this case, the brain areas that are distinct for bilinguals with dementia would be expected to be similar to those in bilinguals at other age groups and no signs of differential development of dementia would be observed. Alternatively, bilinguals might use other brain areas to compensate for degeneration of dementia-affected regions. Therefore, we could expect bilinguals to demonstrate differences in frontostriatal and frontoparietal circuits compared to monolinguals in health cohorts. By extension, these differences should be reflected in the way these brain circuits cope with advancing neuropathology in bilinguals with dementia; for example, by showing a greater degree of neuropathology in age- and performance-matched bilinguals compared to monolinguals.

To clarify the existing evidence in terms of these two alternatives, we systematically reviewed i) the structural and functional brain differences between mono- and bilinguals in non-clinical older individuals, and ii) differences in brain pathology between older mono- and bilinguals with dementia. Despite the considerable variability of the brain imaging methods used in the prior studies (see Methods section and Table 2), brain mapping allowed us to condense the data from a heterogeneous pool of studies onto a single brain map. Actually, multiple different neural indices associated with bilingualism in dementia could even be considered as stronger evidence than replication of the findings with one particular approach. Without a replicable pattern of brain abnormality linked to a specific bilingualism- or dementia-related brain system, it would be difficult to determine the mechanism that might mediate the potential bilingual advantage in dementia. Because the previous review pulling together the data on bilingual brain in healthy aging was published was not systematic, did not include individuals with dementia, and since its publication (Li, Legault & Litcofsky, 2014) more studies have been published, here we provide the first systematic review.
exploring the link between bilingualism and the brain in healthy participants and those diagnosed with dementia.

Therefore, we hypothesized that i) bilinguals would demonstrate more structurally developed frontostriatal and frontoparietal circuits than monolinguals in healthy cohorts and that ii) bilinguals would demonstrate greater neuropathology than monolinguals at the time of dementia diagnosis in the same brain circuits. We also hypothesized that iii) bilinguals would demonstrate more efficient functional use of relevant brain regions than monolinguals in healthy cohorts and that iii) this more efficient use of brain resources would translate into neuroprotective effects against dementia.

Methods

Search strategy

This systematic review accords with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Statement (PRISMA; Moher, Liberati, Tetzlaff, Altman, & Group, 2009). Eligible studies had to compare monolingual vs. bilingual participants on functional or structural neuroimaging outcomes. We retrieved studies that mentioned having recruited mono- and bilinguals. We did not apply a specific definition of mono- and bilingualism. We included participants with mild cognitive impairment (diagnosed by standardized measures), dementia of all causes (diagnosed by standardized measures), any neuroimaging methods assessing brain morphology, biomarkers (tangles/plaques), or functional or structural changes using Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), Functional Magnetic Resonance Imaging (fMRI), or Diffusion Tensor Imaging (DTI) and participants with a mean age of 60 years or above as well as articles in any language. We excluded studies in participants with neurological disorders other than mild cognitive
impairment or dementia. We also excluded studies without neuroimaging measures and those that did not compare monolinguals to bilinguals/multilinguals. Finally, we excluded editorials, commentaries, case reports, and studies that had not been peer-reviewed. We did not place restrictions of the year of publication.

One author initially searched for studies without any resection on the type of study design in PubMed, Embase, Web of Science, Scopus, and PsychINFO. Filters were used to exclude animal studies, but no restrictions were placed on time and language. The initial search was performed on the 31st of January 2018. Screening for title and abstract as well as full-text, data extraction, and risk of bias were conducted independently by pairs of review authors (S.B and B.S.). Disagreements were resolved through consensus and discussion with a third review author when necessary. We requested further information from corresponding authors when necessary. Specifically, because some primary studies did not report the coordinate data to accurately locate the brain regions differentiating the demented bilinguals from healthy ones in a standard brain atlas, we contacted authors to retrieve such coordinates. However, authors either did not respond or did not have the specific coordinate data we needed. Details of the protocol for this systematic review were registered a priori on PROSPERO (CRD42017026839).

Data extraction and risk of bias

We extracted demographic data including sample size, percentage of female participants, mean age per group, degree of cognitive impairment, and mean level of education (Tables 1 and 2). Participants were considered cognitively healthy if they had not received a diagnosis of any dementia type. For studies in participants with dementia, we also extracted data on the type of dementia and the tools used to diagnose dementia and to assess the degree of cognitive impairment (Table 2). For risk of bias, we used the modified version of the
Newcastle-Ottawa Scale (NOS) for cross-sectional studies (Wells et al., 2015). The NOS allows to assign a total of 10 stars to each study on three domains: selection of study groups (range 0–5), comparability of study groups (range 0–2), and exposure/outcome ascertainment (range 0–3).

**Data analysis and synthesis**

The primary outcomes included brain structure and function such as differences in brain volume and regions of interest as well as surrogates of dementia neuropathology such as accumulation of beta amyloid. All studies reporting information about the affected brain areas were included to brain maps illustrating the key results (Figure 2). In addition, we performed a separate activation likelihood estimation (ALE) based analysis for the studies that reported coordinate data. ALE was performed with GingerALE software (version 2.3.6, www.brainmap.org, Eickhoff et al., 2009; Turkeltaub et al., 2012). This software models coordinate data as spatially smoothed 3D Gaussian probability distributions capturing the uncertainty associated with each locus. ALE values were calculated by the voxel-wise union of the probabilities in the modeled activation maps. ALE maps were thresholded at corrected $p < 0.05$ (cluster level $p < 0.05$, clusters defined through 100 permutations).

**Results**

**Overview of the included studies**

We retrieved a total of 17 studies including cross-sectional studies with a non-clinical population and studies with a clinical population. The total number of participants was 715 individuals including 468 healthy older adults (234 monolinguals and 234 bilinguals) and 247 older individuals with dementia [121 monolinguals and 126 bilinguals (Table 1)].
Studies including healthy participants

There were 12 studies that included healthy older individuals (Table 1). To measure bilingualism, some authors administered validated language questionnaires (Anderson et al., 2018; Ansaldo, Ghazi-Saidi, & Adrover-Roig, 2015; Berroir et al., 2017; Borsa et al., 2018), while in some studies it was unclear whether questionnaire or self-reporting measures were validated tools (Abutalebi, Canini, Della Rosa, Green, & Weekes, 2015; Abutalebi et al., 2014; Abutalebi, Guidi, et al., 2015). Some studies also used a picture-naming task including the Snodgrass and Vanderwart’s picture set to assess bilingualism (Abutalebi, Canini, et al., 2015; Abutalebi et al., 2014; Abutalebi, Guidi, et al., 2015; Borsa et al., 2018). Most studies also reported data on the age of second language acquisition which ranged from 10 years old or younger to 18 years of age (Abutalebi, Canini, et al., 2015; Abutalebi et al., 2014; Anderson et al., 2018; Berroir et al., 2017; Gold, Johnson, et al., 2013; Gold, Kim, Johnson, Kryscio, & Smith, 2013; Grady, Luk, Craik, & Bialystok, 2015; Luk et al., 2011; Olsen et al., 2015).

There were nine studies that employed structural neuroimaging including MRI which used techniques such as VBM and DTI (Abutalebi, Canini, et al., 2015; Abutalebi et al., 2014; Abutalebi, Guidi, et al., 2015; Anderson et al., 2018; Borsa et al., 2018; Gold, Johnson, et al., 2013; Gold, Kim, et al., 2013; Luk et al., 2011; Olsen et al., 2015). There were five studies that employed functional neuroimaging including functional fMRI (Ansaldo et al., 2015; Berroir et al., 2017; Gold, Johnson, et al., 2013; Gold, Kim, et al., 2013; Grady et al., 2015).
<table>
<thead>
<tr>
<th>Study</th>
<th>n/group (% of females); M age (SD); MMSE/MMSE short/MoCA; M education (SD)</th>
<th>Language measure; L2 AoA</th>
<th>Neuroimaging method</th>
<th>Brain differences between ML and BL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abutalebi 2014</td>
<td>23 (57%); 61.92 (6.80); 28.74 (0.92); 12 (4.41)</td>
<td>Picture naming (revised version of the Snodgrass and Vanderwart’s picture set) and translation tasks, questionnaire of daily language use</td>
<td>Structural (MRI - VBM)</td>
<td>↑ GMV in BL compared to ML in left temporal pole, right temporal pole, and left/right orbitofrontal region Positive correlation between ↑ GMV in left temporal pole and second language naming performance Overall brain analysis showed BL had ↓ GMV than ML</td>
</tr>
<tr>
<td>Abutalebi 2015a</td>
<td>30 (53%); 61.85 (6.71); 28.81 (0.95); 12.33 (4.54)</td>
<td>Self-report of daily exposure, picture naming task (revised version of the Snodgrass and Vanderwart’s picture set): L1: 80%; L2: 61%</td>
<td>Structural (MRI - VBM)</td>
<td>No correlation between GMV and age in either group in left inferior parietal lobule ↑ GMV in left and right inferior parietal lobules for BL compared to ML</td>
</tr>
<tr>
<td>Abutalebi 2015b</td>
<td>19 (53%); 60.93 (5.81); 28.95 (1.03); 13.16 (4.86)</td>
<td>L1 naming (% accuracy): 99 (0.01) 12.68 (10.71)</td>
<td>Structural (MRI - VBM)</td>
<td>↓ GMV in dorsolateral prefrontal cortex correlated with increased conflict effects in ML only ↑ GMV in the ACC among in BL compared to ML</td>
</tr>
<tr>
<td>Anderson 2018</td>
<td>23 (74%); 74.43 (2.95); 29.26 (0.86); 3.96 (0.98)</td>
<td>Phone interview; LSBQ 8.75 (6.21)</td>
<td>Structural (DTI)</td>
<td>↑ AD in BL compared to ML in the left superior temporal longitudinal fasciculus, bilateral superior posterior corona radiata, right external capsule, corpus callosum, and anterior inferior frontal occipital fasciculus ↑ RD in nearly all white matter brain regions in BL compared to ML, however this effect disappeared when the sample was matched ↑ FA in the internal capsule, the anterior corpus callosum, the corona radiata and the inferior and superior longitudinal fasciculi in BL compared to ML, however this effect disappeared when the sample was matched</td>
</tr>
<tr>
<td>Ansaldo 2015</td>
<td>10 (N/A); 74.5 (7.1); 27.7 (1.1); 16.1 (3.28)</td>
<td>LEAP N/A</td>
<td>Functional (MRI) (During the Simon task)</td>
<td>ML had ↑ activation in the right middle frontal gyrus compared to BL while completing the Simon task</td>
</tr>
<tr>
<td>Borsa 2018</td>
<td>20 (55%); 61.46 (7.26); 29 (0.94); 13.10 (4.17)</td>
<td>LlBQ and picture naming task (revised version of the Snodgrass and Vanderwart’s picture set) N/A</td>
<td>Structural (MRI)</td>
<td>No difference between BL and ML in mean GMV ↓ mean GMV in ACC predictive of ↓ cognitive control in BL but not ML Age negatively correlated with cognitive control in ML but not BL ↓ daily exposure to L2 correlated with ↓ GMV in the ACC and ↓ cognitive control Negative correlation between age and GMV for ML in inferior frontal gyrus, insula and inferior parietal lobule Negative correlation between age and GMV for BL in left inferior parietal gyrus, inferior parietal lobule, insula and caudate ML: Negative correlation between GMV and cognitive control except in ACC and right caudate</td>
</tr>
</tbody>
</table>
BL: Negative correlation between GMV and cognitive control except in the left and right caudate.

**Gold 2013a**

<table>
<thead>
<tr>
<th></th>
<th>Language questionnaire</th>
<th>Structural (MRI - VBM and DTI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ML</strong></td>
<td>20 (50%); 64.4 (5.1); 28.2 (1.6); 17.5 (2.6)</td>
<td>FA in BL compared to ML in inferior longitudinal fasciculus, inferior fronto-occipital fasciculus, fornix and corpus callosum. ( \uparrow ) FA in BL compared to ML in inferior longitudinal fasciculus, inferior fronto-occipital fasciculus, fornix and corpus callosum.</td>
</tr>
<tr>
<td><strong>BL</strong></td>
<td>20 (50%); 63.9 (4.0); 27.8 (1.2); 17.4 (2.2)</td>
<td>( \uparrow ) RD in BL compared to ML in inferior fronto-occipital fasciculus, corpus callosum and parietal and frontal lobes. No differences between ML and BL in AD or MD in any ROIs (inferior longitudinal fasciculus, inferior fronto-occipital fasciculus, fornix and corpus callosum).</td>
</tr>
</tbody>
</table>

**Gold 2013b**

<table>
<thead>
<tr>
<th></th>
<th>Language questionnaire</th>
<th>Functional (fMRI)</th>
<th>Structural (MRI - VBM)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ML</strong></td>
<td>15 (53%); 63.3 (3.8); N/A; 17.5 (2.6)</td>
<td>All measured while completing task switching</td>
<td>No main effects of language group</td>
</tr>
<tr>
<td><strong>BL</strong></td>
<td>15 (53%); 64.1 (4.4); N/A; 17.4 (2.2)</td>
<td>( \downarrow ) switching costs for BL compared to ML in left dorsolateral prefrontal cortex, left ventrolateral prefrontal cortex, and ACC. ( \uparrow ) activation in left middle temporal gyrus in BL compared to ML. BL had ( \downarrow ) activation in left DLPFC, left VLPFC, and ACC but better performance on task switching when compared to ML.</td>
<td>No differences in ROIs in either direction for both ML and BL.</td>
</tr>
</tbody>
</table>

**Grady 2015**

<table>
<thead>
<tr>
<th></th>
<th>Self-report (non-validated)</th>
<th>Functional (fMRI - Resting and during Simon Task)</th>
<th>Structural (DTI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ML</strong></td>
<td>14 (50%); 70.6; MMSE short form: 16.9 (NA); 16 (NA)</td>
<td>BL had stronger connections compared to ML in default mode network and frontoparietal control network. BL had stronger correlations compared to ML between executive control connectivity and task modulation ability. Stronger frontoparietal control network connectivity in BL associated with task related increases in frontal and parietal cortices.</td>
<td>( \uparrow ) FA and RA in corpus callosum in BL compared to ML.</td>
</tr>
<tr>
<td><strong>BL</strong></td>
<td>14 (50%); N/A; MMSE short form: 16.9 (0.4); 16 (2.8)</td>
<td>BL had stronger connections compared to ML between executive control connectivity and task modulation ability. Stronger frontoparietal control network connectivity in BL associated with task related increases in frontal and parietal cortices.</td>
<td>( \uparrow ) FA in BL compared to ML in right inferior frontal gyrus.</td>
</tr>
</tbody>
</table>

**Luk 2011**

<table>
<thead>
<tr>
<th></th>
<th>Self-report</th>
<th>Structural (DTI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ML</strong></td>
<td>14 (50%); N/A; MMSE short form: 16.9 (0.4); 16 (2.8)</td>
<td>BL had stronger functional connectivity between inferior frontal gyrus and posterior brain regions compared to ML. ML had stronger connections between inferior frontal gyrus and anterior regions compared to BL.</td>
</tr>
<tr>
<td><strong>BL</strong></td>
<td>14 (50%); N/A; MMSE short form: 17 (0); 17.7 (2.1)</td>
<td>No group differences in either direction for any ROIs (Left DLPFC, Left VLPFC, Right DLPFC, Right VLPFC, ACC, Left SMG, Right SMG).</td>
</tr>
</tbody>
</table>

**Olsen 2015**

<table>
<thead>
<tr>
<th></th>
<th>Self-report</th>
<th>Structural (MRI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ML</strong></td>
<td>14 (50%); 70.6 (3); MMSE short form: 16.9 (0.4); 16.0 (2.8)</td>
<td>BL had ( \uparrow ) WMV in frontal and temporal lobe compared to ML. Negative relationship between temporal pole thickness and age for ML but not BL.</td>
</tr>
<tr>
<td><strong>BL</strong></td>
<td>14 (50%); 70.4 (3.7); MMSE short form: 16.9 (0.3); 17.8 (2.0)</td>
<td>BL had ( \uparrow ) WMV in frontal and temporal lobe compared to ML. Negative relationship between temporal pole thickness and age for ML but not BL.</td>
</tr>
</tbody>
</table>

---

**ACC**: anterior cingulate cortex; **AD**: axial diffusivity; **ML**: monolinguals; **BL**: bilinguals; **DLPFC**: dorsolateral prefrontal cortex; **DTI**: diffusion tensor imaging; **FA**: fractional anisotropy; **MRI**: functional magnetic resonance imaging; **GMV**: grey matter volume; **LBQ**: language background questionnaire; **LEAP**: Language Experience and Proficiency Questionnaire; **LSBQ**: Language and social background questionnaire; **L1**: 1st language; **L2**: 2nd language; **MD**: mean diffusivity; **MMSE**: mini-mental state examination; **MoCA**: Montreal Cognitive Assessment; **MRI**: magnetic resonance imaging; **RD**: radial diffusivity; **ROI**: region of interest; **SMG**: supramarginal gyrus; **VBM**: voxel-based morphometry; **VLPFC**: ventrolateral prefrontal cortex; **WMV**: white matter volume; \( \uparrow \): higher; \( \downarrow \): lower.
Grey matter differences in healthy bilinguals vs. monolinguals

Some studies \((k = 3)\) found significantly greater GMV in bilingualism vs. monolinguals (Abutalebi et al., 2015a; Abutalebi et al., 2015b; Abutalebi et al., 2014), particularly in the in left and right inferior parietal lobules (Abutalebi et al., 2015a), anterior cingulate cortex (Abutalebi et al., 2015b), left temporal pole, right temporal pole, and left/right orbitofrontal cortex (Abutalebi et al., 2014). However, other studies \((k = 2)\) did not find differences in overall lower GMV (Abutalebi et al., 2014) or mean GMV (Borsa et al., 2018) between mono- and bilinguals. Moreover, one study found that decrease in grey matter volume in the dorsolateral prefrontal cortex was associated with a greater conflict effect in the Erikson Flanker task, which measures a participant’s ability to suppress inappropriate responses within a certain context, in monolinguals only (Abutalebi et al., 2015b). Similarly, another study found that a decrease in GMV in the inferior parietal lobule (IPL) was associated with greater age in monolinguals only, but bilinguals showed greater GMV in the left IPL compared to monolinguals (Abutalebi et al., 2015a). Another study found a significant and positive correlation between greater GMV in left temporal pole and performance in a task assessing second language naming (Abutalebi et al., 2014). A study found that daily exposure to a second language was associated with lower GMV in the anterior cingulate cortex (Borsa et al., 2008). The same study also found that age was negatively associated with GMV in the inferior frontal gyri, insula and inferior parietal lobule in monolinguals and in the in left inferior parietal gyrus, inferior parietal lobule, insula, and caudate in bilinguals (Borsa et al., 2008).

White matter differences between mono- and bilinguals in healthy participants

Some studies \((k = 3)\) applied DTI to investigate differences in WM integrity between mono- and bilinguals (Anderson et al., 2018; Gold et al., 2013a; Luk et al., 2011). Table 1 includes additional study details. One study found significantly greater axial diffusivity in bilinguals
compared to monolinguals across different WM regions (Anderson et al., 2018). They did not find significant differences between mono- and bilinguals in radial diffusivity and fractional anisotropy in several WM regions after matching participants across verbal as well as spatial intelligence quotient, Trail-Making-Task, MMSE scores, age, education, and gender (Anderson et al., 2018). Another study found lower fractional anisotropy in bilinguals vs. monolinguals but greater radial diffusivity across different WM regions as well as no significant differences in axial mean diffusivity in several WM regions of interest (Gold et al., 2013a). Finally, one study found significantly greater WM integrity particularly in the corpus callosum and in the superior and inferior longitudinal fasciculi lifelong bilinguals vs. monolinguals (Luk et al., 2011). Monolinguals however showed stronger connections in other brain regions; there we also no significant differences in axial diffusivity between mono- and bilinguals (Luk et al., 2011).

**Brain activity in healthy bilinguals vs. monolinguals**

Some studies ($k = 4$) used functional neuroimaging methods (see Table 1 for details on these methods). Details of the specific brain regions and cognitive tasks used during functional neuroimaging are presented in Table 1. Results suggested that bilinguals vs. monolinguals showed greater activation of the left IPL and monolinguals vs. bilinguals showed greater activation of the right middle frontal gyrus during The Simon task (Ansaldo et al., 2015). One study found that to resolve conflict in the Simon task, monolinguals compared to bilinguals showed greater connectivity in brain regions responsible for executive functions and interference control as well as motor and visual processing (Berroir et al., 2017). Bilinguals compared to monolinguals showed greater connectivity in one brain region responsible for visuospatial processing (Berroir et al., 2017). Bilinguals showed greater performance in switching-costs in a perceptual task-switching paradigm compared to monolinguals but lower
activation across the left lateral frontal cortex and cingulate cortex (Gold et al., 2013b). See Table 1 for specific brain regions. Finally, one study found that bilingualism showed a stronger correlation between frontoparietal control network and greater task-related activity in the prefrontal and parietal brain areas (Grady et al., 2015). Furthermore, bilinguals compared to monolinguals showed greater intrinsic functional connectivity in the frontoparietal control network and default mode network (Grady et al., 2015).

Studies including participants with mild cognitive impairment and AD

We retrieved four studies which included participants with Mild Cognitive Impairment (MCI) or dementia (Table 2). These studies used different methods assessing bilingualism including a single verbal question (e.g. whether a participant can speak a second language; Duncan et al., 2018) a validated questionnaire (Perani et al., 2017), and language status interview for the participant and significant other (Schweizer et al., 2012). One study did not report the bilingualism measure (Kowoll et al., 2016). With regards to the neuroimaging methods, studies used MRI to measure cortical thickness and tissue density (Duncan et al., 2018), two studies used $^{18}$F-fluorodeoxyglucose PET (FDG-PET) to measure cerebral glucose metabolism (Kowoll et al., 2016; Perani et al., 2017), and one study used Computed Tomography (CT) to measure brain atrophy (Schweizer et al., 2012). Participants’ diagnoses included MCI or AD (Duncan et al., 2018; Kowoll et al., 2016) and probable AD (Perani et al., 2017; Schweizer et al., 2012).

Grey matter differences between mono- and bilinguals with dementia

One study detected significantly thicker cortex and more grey matter density in multilinguals compared to monolinguals across language and cognitive control brain areas (see Table 2 for each brain region; Duncan et al., 2018). Multilinguals vs. monolinguals showed significantly
greater gray matter density in brain areas typically affected by AD (see Table 2 for each brain region; Duncan et al., 2018). While another study did not detect significant differences between mono- and bilinguals in cortical thickness across some brain regions (see Table 2 for each brain region), bilinguals showed significantly greater cortical thickness than monolinguals in brain areas typically affected by AD (Schweizer et al., 2012). Additional study details are presented in Table 2.
<table>
<thead>
<tr>
<th>Table 2</th>
<th>Cross-sectional studies on the relationship between bilingualism and brain pathology in older clinical populations using neuroimaging methods</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study</strong></td>
<td><strong>n/group (% of females); M age (SD)</strong></td>
<td><strong>Immigration status (n, %) Language measure; # of languages; L2 AoA</strong></td>
</tr>
<tr>
<td><strong>Duncan 2017</strong></td>
<td>ML: 34 (50%), 73.6 (SE 0.9); MT: 34 (44%), 73.7 (SE 1.0)</td>
<td>MCI</td>
</tr>
<tr>
<td></td>
<td>BL: 20, 59%, 12.3 (SE 0.7)</td>
<td>Two or more languages; ~ half spoke two languages (BL)</td>
</tr>
<tr>
<td></td>
<td>ML: 7, 21%, 12.5 (SE 0.7)</td>
<td>~ half spoke three or more (MLT)</td>
</tr>
<tr>
<td></td>
<td>AD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ML: 34, (44%), 73.7 (SE 1.0)</td>
<td>AD</td>
</tr>
<tr>
<td><strong>Kowall 2016</strong></td>
<td>ML: 14, (57%), 71.6 (7.9); ML: 14, (50%), 74.6 (6.8)</td>
<td>MCI</td>
</tr>
<tr>
<td></td>
<td>BL: 20, 59%, 12.3 (SE 0.7)</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>ML: 14, (57%), 71.6 (7.9); ML: 14, (50%), 74.6 (6.8)</td>
<td>Questionnaire derived from the Bilingual Aphasia Test; the bilingualism index ranging from 0 (completely ML) to 1 (completely BL) was obtained for each participant</td>
</tr>
<tr>
<td></td>
<td>BL: 12, (86%), 15.3 (3.6); BL: 11 participants were BL</td>
<td></td>
</tr>
<tr>
<td>Perani 2017</td>
<td>ML: 40, (52%), 71.42 (4.88); ML: 45, (71%), 77.13 (4.52)</td>
<td>MCI</td>
</tr>
<tr>
<td></td>
<td>BL: 20, (70%), 77.2 (7.6); BL: 20, (70%), 78.9 (7.6)</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>ML: 10.5 (4.07); ML: 8.26 (4.55)</td>
<td>Questionnaire derived from the Bilingual Aphasia Test; the bilingualism index ranging from 0 (completely ML) to 1 (completely BL) was obtained for each participant</td>
</tr>
<tr>
<td></td>
<td>BL: 45, (71%), 77.13 (4.52)</td>
<td>ML: 13.6 (3.5); ML: 11.6 (4.5)</td>
</tr>
<tr>
<td></td>
<td>BL: 119 participants were BL</td>
<td>Interview with patient and significant-other</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Schwizer 2012</strong></td>
<td>ML: 20, (70%), 77.2 (7.6); ML: 20, (70%), 78.9 (7.6)</td>
<td>MCI</td>
</tr>
<tr>
<td></td>
<td>BL: 20, (70%), 77.2 (7.6); BL: 20, (70%), 78.9 (7.6)</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ML: 13.6 (3.5); ML: 11.6 (4.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**MC1 paper**

AAD: aging-associated cognitive decline criteria; AD: Alzheimer’s disease; aMC1: amnestic mild cognitive impairment; BD: block design; BL: bilingual; BNA: behavioural neurology assessment; CDR: clinical dementia rating; CDT: clock drawing test; CERAD-NP: consortium to establish a registry for Alzheimer's disease neuropsychological assessment battery; CT: cortical thickness; CT*: computed tomography; DRVR: delayed recall visual reproduction; FDG-PET: fluorodeoxyglucose positron emission tomography; GDS: geriatric depression scale; GDS*: Global Deterioration Scale; IPR: immediate recall visual reproduction; L2 AoA: 2~ language age of acquisition; M: mean; MCI: mild cognitive impairment; ML: monolinguals; MLT: multilinguals; MMSE: mini-mental state examination; MRI: magnetic resonance imaging; N/A: not available; NIA-AA: National Institute on Aging and Alzheimer's Association; NINCDS-ADRDA: National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association; NS: non-significant; SCW: Stroop Colour Words; SD: standard deviation; SDVR: short delay verbal recall; SE: standard error; SE Stroop Interference: SST: spatial span total; TMT: trail making test; VLT: verbal long-term memory; VSTM: visuospatial short-term memory; WMS-R: Wechsler memory scale revised; (LIFG) Left inferior frontal gyrus; (RIFG) Right inferior frontal gyrus; (LATG) Left anterior temporal gyrus; (RATG) Right anterior temporal gyrus; (LIPL) Left inferior parietal lobule; (LC) Left cerebellum; (RC) Right cerebellum; (RCT) Right cerebellar tonsil; (LSG) Left supramarginal gyrus; (RSG) Right supramarginal gyrus.
Brain physiological function between mono- and bilinguals with dementia

In one study, bilinguals showed significantly lower glucose uptake as measured by $^{18}$F-fluorodeoxyglucose, which is a surrogate of neurodegeneration, across brain areas responsible for speech, language, and areas typically affected by AD such as frontotemporal and parietal areas including Brocas’ areas 9 (right), 21 (right), 40 (right and left), and 47 (left) (Kowoll et al., 2016). In the other study, cerebral hypometabolism was more severe in bilinguals vs. monolinguals in brain regions typically affected by AD (Perani et al., 2017). Regarding the executive control network, bilinguals showed greater metabolic connectivity in brain areas responsible for cognitive control and in default mode networks relative to monolinguals (Perani et al., 2017). Indeed, there was increased long-distance metabolic connectivity between the inferior and superior lobules, angular gyrus and the dorsolateral prefrontal cortex (Perani et al., 2017). Also, bilinguals but not monolinguals showed increased metabolic connectivity between the bilateral middle and superior frontal gyri with the superior parietal cortices (Perani et al., 2017). Increased metabolic connectivity was also observed between the bilateral caudate nucleus with the anterior, middle, and posterior cingulate cortex, the right insula, the right inferior frontal gyrus, and the right parietal operculum in bilinguals but not monolinguals (Perani et al., 2017). Regarding the dorsal default mode network there were significant metabolic autocorrelations within the posterior cingulum/precuneus in both cohorts but only bilinguals showed additional metabolic correlations within the cingulate cortex, the orbitofrontal cortex, the caudate nucleus, and the thalamus, bilaterally (Perani et al., 2017). Also, only the bilinguals showed significant correlation in the anterior default mode network between the anterior cingulate cortex and medial frontal cortex with the posterior cingulum (Perani et al., 2017). Additional study details are presented in Table 2.
Figure 0.1 Brain map

A summary of the regional effects in brain imaging studies in bilinguals vs. monolinguals without dementia (top) and bilingual advantage in dementia (bottom). The figure illustrates all regional effects reported in relevant studies (see Tables 1 and 2). The scale for the coloured circles represents the number of studies reporting significant group differences (bilinguals vs. monolinguals) in the respective brain regions. At the top image, the activation cluster at the left parietal cortex comes from an ALE analysis for the coordinate data (corrected p<0.05). To make this cluster visible, the centre of the circle illustrating the number of studies with significant effects for this particular brain region is empty. A) refers to maximal intracranial width, B) to temporal horn diameter, and C) to minimal intercaudate distance (Schweizer et al. 2012). Other indices reported by Schweizer et al. (2012) are not visualized in the image as they either do not reflect regional brain differences (width of the third ventricle and frontal skull) or are calculated as ratios between multiple indice
Risk of bias

Studies including non-clinical participants

Results from risk of bias for non-clinical participants are presented in Table 3. All studies employed an acceptable sampling method, but no study reported power calculations. Most studies including non-clinical participants applied a validated measure of bilingualism while others either use a non-validated tool or did not provide specific information on how they had obtained data on bilingualism. For example, some studies applied validated measures of bilingualism while other studies only asked participants whether they could speak a second language. All studies controlled for important confounders either statistically or methodologically. We considered confounders to be important if previous literature identified them as confounders in the field of bilingualism research. All studies applied the appropriate statistical techniques including between-group statistical models.

Studies including participant with dementia

Results from risk of bias are for studies including participants with dementia are presented in Table 4. Every study employed an acceptable sampling method, but no study provided information of power calculation. While only one study adopted a validated questionnaire to measure bilingualism, another did not provide information on how bilingualism was measured, and another two studies used non-validated measures including interviews. Each study controlled for important covariates including degree of cognitive impairment, education, age,
Discussion

Our aim was to systematically review the literature exploring the relationship between bilingualism and the brain. We sought to clarify whether bilingualism is associated with differences in certain brain structures generally implicated in language processing and executive function in healthy individuals. We also sought to explore whether bilingualism was associated with differences in brain areas typically affected by AD in individuals with a diagnosis of AD. Previous behavioural evidence has shown that despite a similar degree of cognitive impairment, bilinguals vs. monolinguals were older at the time of dementia and AD diagnosis by up to five years, respectively (Bialystok, Craik, Binns, Ossher, & Freedman, 2014; Bialystok, Craik, & Freedman, 2007). This evidence indicates that bilingualism contributes to CR allowing bilinguals to maintain functional cognition longer and delay clinical symptom onset (Gold, 2016). The executive control network might regulate the activation of multiple languages in the language network resulting in functional and structural changes in relevant brain areas (Green & Abutalebi, 2013). However, studies that investigate the underlying mechanisms responsible for the observed relationship between bilingualism and CR have not been previously systematically reviewed.

In the brain maps (Figure 1) we outlined several key brain regions implicated in bilingualism. While most studies did not find between-group differences in overall GMV, two studies showed that older healthy bilinguals vs. monolinguals demonstrated greater GMV, particularly in the inferior parietal lobule after matching participants for several CR-related variables and cognition (Abutalebi et al., 2014; Abutalebi et al., 2015). However, the measure used for matching participants on cognition (Mini-Mental Status Examination) in these studies (Abutalebi et al., 2014; Abutalebi et al., 2015), lacks sensitivity in assessing cognition in healthy cohorts; therefore, similar scores on this measure does not signify equivalent levels of cognition (Bialystok, Anderson, & Grundy, 2018). Interestingly, in
studies assessing cognition more comprehensively (Olsen et al., 2015; Gold, Johnson, & Powell, 2013), there were no differences between older mono- and bilinguals in grey matter volume, questioning findings from others (Abutalebi et al., 2014; Abutalebi et al., 2015).

While our included studies did not always observe differences between mono- and bilinguals on global measures of grey matter, several studies consistently showed that bilinguals had greater volume in the inferior parietal lobule. Previous studies in young cognitively intact adults support our findings by showing that the inferior parietal lobule may be implicated in bilingualism (Li et al., 2014), particularly when the second language is learnt before the age of five years and the proficiency in speaking the second language is high (Mechelli et al., 2004). Also, while performance on a reaction time task was similar between mono- and bilinguals, when completing the task, monolinguals showed activation of the prefrontal cortex, which is responsible for cognitive control, and bilinguals showed activation of the inferior parietal lobule, which is responsible for language (Ansaldo et al., 2015).

Moreover, one study found that multilinguals vs. monolinguals with MCI and AD matched for memory performance demonstrated greater cortical thickness in the left inferior parietal lobule (Duncan et al., 2018). Another study found greater cerebral glucose uptake in parietal brain regions in bilinguals vs. monolinguals after adjusting for disease severity (Kowoll et al., 2016). Another study using a more robust measure of bilingualism documented more severe glucose hypometabolism in the left inferior parietal lobule among bilinguals vs. monolinguals (Perani et al., 2017). The inferior parietal lobule is a brain region which is particularly vulnerable to aging and AD (Perani et al., 2017). Overall, our included studies indicate that bilingualism may promote neuroplasticity in the inferior parietal lobule and may protect against MCI and AD.

Bilinguals also demonstrated greater WM connectivity compared to monolinguals (Luk et al., 2011; Olsen et al., 2015). For example, in one study, bilinguals compared to
monolinguals showed greater WM integrity projecting to the bilateral superior longitudinal fasciculi, the right fronto-occipital fasciculus, and the uncinate fasciculus (Luk et al., 2011). Also, WM connectivity was associated with resting-state functional connectivity in frontal areas which might indicate that lifelong bilingualism might have promoted changes in brain structure and function simultaneously (Luk et al., 2011). By contrast, in another study with a larger sample size, monolinguals, as opposed to bilinguals, showed greater WM integrity inferior and superior fronto-occipital fasciculi (Gold, Johnson, et al., 2013). More recently however, bilinguals vs. monolinguals matched on verbal and spatial IQ, age, education, Trail-Making-Task performance, disease severity, and gender, showed greater WM integrity in the superior longitudinal fasciculus (Anderson et al., 2018).

There are also alternative explanations for some of the observed relationships between bilingualism and structural and well as functional changes to relevant brain areas. For example, participants in the included studies might vary in factors associated with brain health and risk of dementia such as physical activity (Brini et al., 2018). Moreover, what “WM integrity” precisely means and its clinical relevance have been a point of contention and some have argued that its use should be discontinued (Jones, Knösche, & Turner, 2013). Finally, evidence gathered herein are all cross-sectional and therefore should only interpreted as hypothesis generating. More studies with longitudinal design need to be conducted before concluding that i) bilingualism modifies relevant brain regions and ii) that such changes in the brain carry any translational impact on the onset and risk of dementia.

While our findings indicate that bilingualism may promote certain brain areas implicated in the neuropathology of AD, our findings need to be interpreted in light of clinical and methodological differences as well as limitations within studies. For example, lack of validated measures of bilingualism, particularly in studies including participants with dementia, was a major limitation. Bilingualism is increasingly recognised as a
multidimensional construct encompassing several factors such as frequency, proficiency, and age of acquisition. Therefore, bilingualism is likely not a dichotomous variable, rather it is a construct that extends on a continuum from monolingualism to bilingualism (ability to speak two languages) to multilingualism (ability to speak three or more languages) while including several dimensions (Luk & Bialystok, 2013). Some studies also did not collect data on age of acquisition. Moreover, some studies did not assess the extent to which monolinguals had been exposed to foreign languages or whether they spoke a dialect. This questions whether monolingual participants were in fact, truly monolinguals (Laine & Lehtonen, 2018). As such, future research should consider applying validated measures of bilingualism and as well as measure it on a continuum rather than dichotomizing participants between mono- and bilinguals (Luk & Bialystok, 2013). Future studies will also benefit from increasing sample sizes, particularly in light mixed findings in this field of research.

Differences in sample size may have contributed to the conflicting findings in this field of research (García-Pentón et al., 2016a, 2016b). The average total sample size in studies with non-clinical and clinical populations was 18 (SD = 5.69) and 27.63 (SD = 12.09) participants, respectively. Small samples increase the probability of Type I and Type II error and inflate the estimated effect size thereby affecting precision (Ioannidis, 2005, 2008). Conflicting results from some of our included studies is not surprising considering that statistical power in the field of neuroscience is generally low ranging from approximately 8% to 31% (Button et al., 2013; however, see Bacchetti, 2013 for an opposing view on the importance of sample size in neuroscience). Interestingly, not one of our included studies reported power calculations. Also, in a field where studies have small samples such as in neuroscience, a non-significant finding may be less likely to be published resulting in publication bias (Button et al., 2013). While, we could not conduct formal statistical tests for assessing publication bias, it is likely present in the social sciences (Franco, Malhotra, &
Simonovits, 2014), neuroimaging research (Jennings & Van Horn, 2012), and bilingualism research (De Bruin, Treccani, & Della Sala, 2015). Therefore, it is a possibility that publication bias be found in the field of bilingualism neuroscience as well.

In conclusion, the extant limited evidence indicates that several brain areas consistently appear to be implicated in bilingualism. Generally, bilingualism was linked with differences in affect frontostriatal and frontoparietal circuits, particularly in brain regions responsible for language and executive functions. This supports the hypothesis that changes to brain areas responsible for language including the temporoparietal lobes, may extend to brain areas implicated in cognitive control including the prefrontal cortex (Green & Abutalebi, 2013). These changes appear consistent across the lifespan and may promote neuroplasticity and protect against neurodegeneration. However, the included studies carried important limitations such as small sample sizes and poor measurement of both monolingualism and bilingualism. Also, in this field of research we could identify only a limited number of studies. As such, while the evidence appeared somewhat consistent and agrees with some studies in younger populations, when considering the limited number of studies and their limitations, it may be early to provide conclusive recommendations.
References


**Between-chapter (4-5) rationale**

In Chapter 3, we observed that bilinguals experienced a later onset of AD symptoms than monolinguals. In Chapter 4 we also observed that bilinguals tend to show greater frontostriatal and frontoparietal circuits as well as greater white and grey matter than monolinguals in brain that are severely affected by AD. Taken together these findings suggest that bilingualism may render the brain more resilient against AD neuropathology and translate into a later onset of AD symptoms and diagnosis. However, in Chapter 3 and 4 four, participants were individuals with a diagnosis of dementia or AD and studies generally operationalised bilingualism as speaking two or more languages. None of the included studies had considered language acculturation as a variable that might be associated with cognition. Therefore, in the following chapter, we explored whether language acculturation was linked with cognitive performance among Hispanic and Asian seniors.
Chapter 5 Language acculturation predicts cognitive performance in older Hispanic and Asian individuals in the United States: a cross-sectional NHANES study.
Authors: Stefano Brini\textsubscript{a,b,c,d}, Philip Veliz\textsubscript{e}, Timothy J. Fairchild\textsubscript{c}, Hamid R. Sohrabi\textsubscript{f,g}, Ralph N. Martins\textsubscript{f,g}, Jeremiah J. Peiffer\textsubscript{c}, Jeffrey, J. Hebert\textsubscript{c,j}

\textsuperscript{a}Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland

\textsuperscript{b}Turku Brain and Mind Center, Turku, Finland

\textsuperscript{c}School of Psychology and Exercise Science, Murdoch University, Perth, Western Australia, Australia

\textsuperscript{d}Health Services Research and Management School of Health Sciences, City, University of London, London, United Kingdom

\textsuperscript{e}University of Michigan School of Nursing

\textsuperscript{f}School of Medical and Health Sciences, Edith Cowan University, Perth, Western Australia, Australia

\textsuperscript{g}Department of Biomedical Sciences, Macquarie University, New South Wales, Australia

\textsuperscript{j}Faculty of Kinesiology, University of New Brunswick, Fredericton, Canada

Keywords acculturation, NHANES, cognition, language, bilingualism, immigration

Declarations of interest: none
Abstract

Ethnic minorities living in the United States experience a higher incidence of dementia compared to their Caucasian counterparts. Therefore, identifying modifiable factors associated with cognitive health in seniors from ethnic minorities is an important public health priority. Acculturation may be one such factor. Although some studies have linked higher acculturation with better performance in verbal but not in non-verbal tasks, others have observed the opposite trend. Differences in the operationalization of acculturation and varying sample sizes (range: 50-503) may explain the mixed findings. Here, we explored the link between self-reported language use at home, one index of acculturation, and cognition in a larger sample (N = 890) of Hispanic and Asian participants living in the United States. This is the first study adopting language preference at home as a proxy for acculturation using a large representative sample of the civilian non-institutionalized United States population. We retrieved data from the National Health and Nutrition Examination Survey on self-reported language use at home, verbal fluency, psychomotor speed and memory performance, as well as demographic information. We replicated previous evidence by showing that higher level of acculturation predicted higher verbal fluency. However, contrary to previous evidence, we found that higher levels of acculturation also predicted higher psychomotor speed. In conclusion, language acculturation is an important factor contributing to cognitive health in older Hispanic and Asian individuals living in the United States.
Introduction

Individuals from Hispanic and Asian backgrounds living in the United States experience poorer health when compared to their Caucasian counterparts (Mehta & Yeo, 2017; Vega, Rodriguez, & Gruskin, 2009). For example, the incidence of dementia, particularly among ethnic minorities, is significantly higher than among Caucasian Americans (Mehta & Yeo, 2017). Given the projected increase in life expectancy (Kontis et al., 2017) and prevalence of dementia (Prince et al., 2013) in industrialized countries, identifying factors that can be modified to maintain cognitive health into old age has become a public health priority (Wortmann, 2012). Importantly however, research exploring the relationship between ethnic disparities and cognitive health is still lacking (Babulal et al., 2019). One factor linked with cognitive health among older individuals from ethnic minorities is acculturation (Lara, Gamboa, Kahramanian, Morales, & Hayes Bautista, 2005; Xu, Zhang, & Wu, 2017). This refers to the acquisition of certain cultural properties (e.g. language) of the recipient society (Lara et al., 2005).

However, acculturation is a complex phenomenon including multiple cultural properties (Portes & Rumbaut, 2001; Zhou, 2014). Importantly, there is little consistency in the way researchers have measured acculturation. For example, years lived in a country, years lived and educated in a country and age at which the local language was first learned, or language spoken every day at home, have been commonly used to measure acculturation (Boone, Victor, Wen, Razani, & Pontón, 2007; Flores et al., 2017; Touradjí, Manly, Jacobs, & Stern, 2001). Studies adopting language as a proxy for acculturation (Flores et al., 2017; Lara et al., 2005; Xu et al., 2017) have occasionally measured it with a language questionnaire (Coffey, Marmol, Schock, & Adams, 2005; Razani, Murcia, Tabares, & Wong, 2007). In addition to the variability in the operationalization and measurement of acculturation (Lara et al., 2005), studies investigating the link between acculturation and
cognition have generally included relatively small sample sizes ranging from 50 to 503 participants (Arnold, Montgomery, Castañeda, & Longoria, 1994; Flores et al., 2017; Manly, Byrd, Touradji, & Stern, 2004; Touradji et al., 2001; Xu et al., 2017). Importantly, studies linking acculturation with cognition have produced mixed findings.

For example, greater acculturation predicted higher performance in a verbal task among 87 foreigner-born non-Hispanic participants (Touradji et al., 2001). However, it is unclear if higher acculturation reflected a better command of English or greater cognitive performance (Lehtonen et al., 2018) because the language of the cognitive tasks was not reported (Touradji et al., 2001). Across different ethnic minorities (N range: 50-161), acculturation positively correlated and predicted scores on verbal tasks and executive functions (Boone et al., 2007; Coffey et al., 2005) but not non-verbal intelligence (Razani et al., 2007). Conversely, in 94 Hispanic-Spanish individuals living in the United States, and for whom tests were administered in Spanish, lower acculturation predicted lower attention and inhibition control while higher acculturation predicted higher working memory (Flores et al., 2017). In a larger sample (N = 503) however, the effect size between acculturation and cognition diminished and became non-significant after multivariate adjustment of covariates (Manly et al., 2004). Recently, a systematic review linked acculturation with greater cognition in migrants, but the mechanism of this association remained unclear (Xu et al., 2017).

When language use is adopted as a proxy for acculturation, there is evidence showing that higher levels of acculturation (higher use of the local language) is associated with better health outcomes (Perez-Escamilla, 2010). This link may relate to individuals from ethnic minorities in the United States who reported greater use of English also having greater access to the healthcare system (Derose & Baker, 2000; Fiscella, Franks, Doescher, & Saver, 2002). However, speaking more than two languages has, in itself, also been linked with greater
cognition in older healthy individuals even after adjusting for multiple covariates (Kavé et al., 2008; Padilla, Mendez, Jimenez, & Teng, 2016; Razani et al., 2007). Indeed, the phrase ‘bilingual advantage’ is often used to refer to this association (Bialystok, 2017). There is also evidence indicating that language preference at home in ethnic minorities may be linked with cognition (Flores et al., 2017).

Given the link between language and cognition, the small sample sizes in previous studies, and the equivocal results in the extant literature, it is timely to investigate the role of language acculturation with a larger and representative sample of the civilian non-institutionalized United States population. As such, we hypothesized that participants who reported speaking English at home (high acculturation) would show greater verbal and non-verbal skills compared to participants who reported speaking either their native language (low acculturation) or both English and another language at home (middle acculturation).

**Methods**

**Design**

We used data from the *National Health and Nutrition Examination Survey (NHANES)*, which is an ongoing survey that is conducted in the United States by the National Center for Health Statistics of the Centers for Disease Control and Prevention. The NHANES Survey Methods and Analytic Guidelines can be found here: [https://wwwn.cdc.gov/nchs/nhanes/analyticguidelines.aspx](https://wwwn.cdc.gov/nchs/nhanes/analyticguidelines.aspx). We combined data from the 2011-2012 and 2013-2014 cycles of the study. This study adopted a cross-sectional design. The predictor variable included language status with three levels: *mostly non-English* (reference group), *both equally*, and *mostly English*. The outcomes included psychomotor speed performance and verbal fluency.
To identify individuals who are representative of the civilian non-institutionalized United States population, participants were selected through a complex stratified, multistage national probability sampling design. Individuals living in nursing homes, soldiers, institutionalized persons, or United States nationals living outside of the United States were excluded. Each sample person is given a sample weight. A sample weight reflects the number of individuals in the populations that are represented by that sample person in NHANES while taking into account unequal probability of selection, nonresponse adjustment, and adjustment to independent population controls. Also, the NHANES oversamples certain subgroups from within the population to increase reliability and precision of the outcome estimates. Therefore, to account for the complex sampling design and oversampling we created the appropriate weights.

We retrieved data from the NHANES website on demographics, language status, and cognitive function. The cognitive tasks were administered only to participants aged 60 years and above; therefore, we restricted our analysis to this age group. While cognitive decline is present across most of an individual’s lifespan, the rate of decline accelerates from approximately the age 60 years onward in healthy individuals (Salthouse, 2009, 2019). The response rate for the years 2011-2014 in the unweighted interviewed sample ranged between 64.4 and 65.6% for participants aged 60-69 years, 59.2 and 61.4% for those aged 70-79, and 47.9 and 51.7% in participants aged 80 years and above. Ethics approval for the NHANES was given by the National Center for Health Statistics Research Ethics Review Board.

**Materials**

The NHANES provides a freely accessible dataset from which researchers can retrieve variables of interest. For this study, we retrieved three datasets from the NHANES website including the demographic, acculturation, and cognitive functioning. From the demographic
dataset we retrieved age, sex, education, country of birth, ethnicity, citizenship status, and annual family income. From the acculturation variables we retrieved information on the languages spoken at home for Hispanic and Asian participants. From the cognitive functioning dataset, we retrieved the Digit Symbol Substitution Test (DSST), Animal Fluency (AF), and the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) neuropsychological battery Word Learning (W-L) subtest. Participants could choose their preferred language for the cognitive testing. The languages from which they could choose included English, Spanish, Korean, Vietnamese, and Chinese.

Demographics and covariates

Age was recorded as the age in years at screening. Sex was a categorical variable in which participants reported being males or females. This variable was coded as male = 1 or female = 2. Therefore, the reference group for this variable was males.

The variables education and annual family income included multiple categories and the number of participants within some categories was small. Therefore, to increase the number of participants in each category so as to increase statistical power, we collapsed several categories and created variables with fewer categories. Details of each category in the education and annual family income variables are explained below.

To attain participants’ level of education, participants were asked “What is the highest grade or level of school completed or the highest degree {you have/s/he has} received?” with the choices being: “Less than 9th grade”, “9-11th grade (includes 12th grade with no diploma)”, “High school graduate/GED or equivalent”, “Some college or AA degree”, or “College graduate or above”. We collapsed the first three categories into a single category and the last two into one category to differentiate between participants whose highest levels of education was high school and tertiary education.
We also retrieved data on country of birth with two levels: 1) born in 50 United States or Washington, District of Columbia; and 2) born in other countries, including United States territories. Therefore, country of birth was entered as a categorical variable.

Participants were asked to report their annual family income by selecting one of the following categories: “$0 to $ 4,999”, “$ 5,000 to $ 9,999”, “$10,000 to $14,999”, “$15,000 to $19,999”, “$20,000 to $24,999”, “$25,000 to $34,999”, “$35,000 to $44,999”, “$45,000 to $54,999”, “$55,000 to $64,999”, “$65,000 to $74,999”, “$20,000 and Over”, “Under $20,000”, “$75,000 to $99,999”, or “$100,000 and Over”. We collapsed categories to create three groups: i) between $0 and $24,999; ii) $25,000 and $64,999; and iii) over $65,000. We removed the category “$20,000 and Over” because it did not provide an upper limit. We chose these cut points because they provided the most even distributions of income.

Language Acculturation

At the home of participants, trained interviewers had used the Computer-Assisted Personal Interviewing system to ask participants questions about their language status. We retrieved two variables including information on language use at home for Mexican American, other Hispanic and non-Hispanic Asian participants. Those who self-identified as Mexican American, other Hispanic, or were asked to report what languages they use at home by selecting one of the following four options: only Spanish, more Spanish than English, both equally, more English than Spanish, or only English. For participants who self-identified as non-Hispanic Asian they were asked whether they spoke only a non-English language, more non-English than English, both equally, more English than non-English at home.

However, the acculturation variable includes questions about language use for Hispanic and Asian populations. Therefore, to increase the overall sample size for the acculturation variable, we combined the two variables (1: Mexican-American, other Hispanic
and 2: non-Hispanic Asian) and collapsed the Only Spanish and More Spanish than English categories into one category named Mostly Spanish and collapsed More English than Spanish and Only English and created another category named Mostly English. We applied the same procedure for the non-Hispanic Asian variable. Therefore, this procedure allowed to increase sample size in each language category by generating three categories: mostly non-English, both equally, and mostly English.

**Digit Symbol Substitution Test (DSST)**

The DSST is a subscale of the Wechsler Adult Intelligence Scale (WAIS III) that measures processing speed, sustained attention, and working memory (Wechsler, 1981). In this test, participants are given a paper at the top of which there was a key including nine numbers coupled with symbols where participants were asked to copy the symbols in 133 boxes that match the numbers. Before starting the test, participants were instructed to complete one trial run and those who could not match the symbol with the number, did not participate in the main trial.

**Animal Fluency (AF)**

The AF is a test that measures categorical verbal fluency (Strauss, Sherman, & Spreen, 2006). In this test, participants were instructed to name as many animals as possible and for each correct animal they received one point. Before starting the test, participants were instructed to name three pieces of clothing as a practice trial. Those who were unable to name three items of clothing did not continue with the AF test.
Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) neuropsychological battery Word Learning (W-L)

The CERAD W-L is used to measure immediate and delayed learning ability of newly learnt information (Fillenbaum et al., 2008). This test included three consecutive trials and one delayed recall in which participants were instructed to learn 10 unrelated words. In the immediate learning condition, participants were presented the 10 unrelated words and were asked to read them aloud one by one. At the end of each trial, participants were asked to recall as many words as they could. The sequence of the words was different in each trial. In each trial, participants could achieve a maximum score of 10 points. We calculated the summary score by adding the scores of each trial. Participants could achieve a maximum total score of 30 (i.e. 10 correct responses across the three trials). Higher scores indicate better learning performance. Participants completed the delayed learning condition after completing the DSST and AF tasks (8-10 minutes later). Also, data on the number of incorrect words that were not part of the list (intrusions) were collected. These are provided as supplementary data.

Statistical analyses

Data from the NHANES 2011-2012 and 2013-2014 cycles were used because these were the only NHANES cycles that included information on cognitive function. The primary objective was to explore whether language status was associated with cognitive function. To explore this question, we conducted a complete-case analysis using Multiple Linear Regression while adjusting for age, sex, education, annual family income, and country of birth. We identified participants for whom sufficient data on language status and cognitive function were available. We used the Mostly non-English as the reference group in each analysis. We also
conducted multiple imputation for each outcome and conducted Multiple Linear Regression using the imputed dataset. All data were analysed at an alpha level of .05 with STATA: Data Analysis and Statistical Software Version 15. Of note, STATA does not allow the analyst to calculate β values with the appropriate weights. Therefore, each β value was calculated without appropriate weights and, as such, does not necessarily reflect a precise estimation of the magnitude of each relationship.

Results

We analysed data from 890 Hispanic and Asian participants. The median age of the whole sample was 69 years. Participants’ demographic profiles are presented in Table 1.

Table 1. Demographic characteristics of the analytic sample.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mostly non-English</th>
<th>Both equally</th>
<th>Mostly English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (M; SD years); range</td>
<td>68.04 (6.24) 60</td>
<td>67.53 (6.26) 60</td>
<td>67.50 (6.08) 60</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>320</td>
<td>52</td>
<td>133</td>
</tr>
<tr>
<td>Females</td>
<td>341</td>
<td>47</td>
<td>153</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>485</td>
<td>63</td>
<td>111</td>
</tr>
<tr>
<td>Tertiary</td>
<td>176</td>
<td>35</td>
<td>175</td>
</tr>
<tr>
<td>Country of birth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Born in the US</td>
<td>24</td>
<td>45</td>
<td>168</td>
</tr>
<tr>
<td>Born outside the US</td>
<td>637</td>
<td>53</td>
<td>116</td>
</tr>
<tr>
<td>Ethnic origin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mexican American</td>
<td>172</td>
<td>46</td>
<td>116</td>
</tr>
<tr>
<td>Other Hispanic</td>
<td>249</td>
<td>41</td>
<td>73</td>
</tr>
<tr>
<td>Non-Hispanic Asian</td>
<td>240</td>
<td>12</td>
<td>97</td>
</tr>
<tr>
<td>Citizenship status</td>
<td>657</td>
<td>95</td>
<td>284</td>
</tr>
<tr>
<td>Citizen by birth/naturalization</td>
<td>434</td>
<td>90</td>
<td>265</td>
</tr>
<tr>
<td>Not a citizen of the US</td>
<td>223</td>
<td>5</td>
<td>19</td>
</tr>
<tr>
<td>Annual family income</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>US$0 - $24,999</td>
<td>576</td>
<td>94</td>
<td>258</td>
</tr>
<tr>
<td>US$25,000 - $64,999</td>
<td>280</td>
<td>48</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>196</td>
<td>34</td>
<td>102</td>
</tr>
</tbody>
</table>
We conducted the Multiple Linear Regression to test whether acculturation including covariates significantly predicted DSST, AF, and CERAD scores. The regression model was statistically significant ($R^2 = .42, F(7, 27) = 115.73, p < .001$) expanding 42% of the variance in DSST scores. Participants who reported speaking mostly English at home scored 5.57 points higher ($p = .002$) in the DSST task compared to the Mostly non-English group (reference group; Table 2). The regression equation for the AF task was also statistically significant ($R^2 = .16, F(7, 27) = 17.00, p < .001$), explaining 16% of the variance in AF scores. Here, participants who reported speaking both languages equally at home generated 1.72 significantly ($p = .006$) fewer words relative to the reference group (Table 2). The regression model for the CERAD word learning ($R^2 = .21, F(8, 25) = 47.43, p < .001$), delayed recall ($R^2 = .21, F(8, 25) = 78.53, p < .001$), and intrusion ($R^2 = .04, F(8, 25) = 4.41, p = .002$) were also statistically significant. Participants who reported speaking mostly English at home scored 1.05 points significantly higher than participants reporting speaking Mostly non-English (reference group; Table 2). We conducted multiple imputation on each predictor and outcome and conducted additional analyses on the imputed dataset. Results did not change in the analysis using the imputed dataset relative to the results in the unimputed dataset.

Table 2. Results of the Multiple Linear Regression ($N = 890$).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef.</th>
<th>Linearized SE</th>
<th>$T$</th>
<th>$p$</th>
<th>95% CI</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DSST</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-1.00</td>
<td>0.10</td>
<td>-10.37</td>
<td><strong>0.000</strong></td>
<td>-1.20, -.80</td>
<td>-.31</td>
</tr>
<tr>
<td>Sex</td>
<td>3.02</td>
<td>0.88</td>
<td>3.45</td>
<td><strong>0.002</strong></td>
<td>1.23, 4.80</td>
<td>0.08</td>
</tr>
<tr>
<td>Education</td>
<td>14.81</td>
<td>1.44</td>
<td>10.30</td>
<td><strong>0.000</strong></td>
<td>11.89, 17.74</td>
<td>0.39</td>
</tr>
<tr>
<td>Country of birth</td>
<td>-4.64</td>
<td>1.73</td>
<td>-2.68</td>
<td><strong>0.012</strong></td>
<td>-8.17, -1.11</td>
<td>-0.11</td>
</tr>
<tr>
<td>Annual family income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$25,000 and $64,999</td>
<td>over $65,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------------------</td>
<td>--------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Age</td>
<td>-0.22</td>
<td>0.03</td>
<td>-7.45</td>
<td>0.000</td>
<td>-2.8, -1.6</td>
<td>-0.27</td>
</tr>
<tr>
<td>Sex</td>
<td>-0.44</td>
<td>0.23</td>
<td>-1.90</td>
<td>0.067</td>
<td>-0.91, -0.03</td>
<td>-0.04</td>
</tr>
<tr>
<td>Education</td>
<td>1.42</td>
<td>0.44</td>
<td>3.23</td>
<td>0.003</td>
<td>0.52, 2.31</td>
<td>0.14</td>
</tr>
<tr>
<td>Country of birth</td>
<td>-2.10</td>
<td>0.41</td>
<td>-5.07</td>
<td>0.000</td>
<td>-2.94, -1.26</td>
<td>-0.19</td>
</tr>
<tr>
<td>Annual family income</td>
<td>$25,000 and $64,999</td>
<td>0.35</td>
<td>0.34</td>
<td>1.04</td>
<td>0.305</td>
<td>-0.33, 1.04</td>
</tr>
<tr>
<td></td>
<td>over $65,000</td>
<td>1.16</td>
<td>0.55</td>
<td>2.11</td>
<td>0.043</td>
<td>0.04, 2.29</td>
</tr>
<tr>
<td>Acculturation</td>
<td>Both equally</td>
<td>-1.75</td>
<td>0.63</td>
<td>-2.77</td>
<td>0.009</td>
<td>-3.04, -0.46</td>
</tr>
<tr>
<td></td>
<td>Mostly English</td>
<td>0.05</td>
<td>0.50</td>
<td>0.10</td>
<td>0.920</td>
<td>-0.97, 1.07</td>
</tr>
<tr>
<td>CERAD word learning</td>
<td>Age</td>
<td>-0.27</td>
<td>0.03</td>
<td>-8.76</td>
<td>0.000</td>
<td>-3.37, -2.10</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>1.31</td>
<td>0.43</td>
<td>3.08</td>
<td>0.004</td>
<td>0.443, 2.176</td>
</tr>
<tr>
<td></td>
<td>Education</td>
<td>2.22</td>
<td>0.43</td>
<td>5.20</td>
<td>0.000</td>
<td>1.348, 3.084</td>
</tr>
<tr>
<td></td>
<td>Country of birth</td>
<td>-0.36</td>
<td>0.62</td>
<td>-0.58</td>
<td>0.565</td>
<td>-1.634, 0.908</td>
</tr>
<tr>
<td></td>
<td>Annual family income</td>
<td>$25,000 and $64,999</td>
<td>0.48</td>
<td>0.44</td>
<td>1.09</td>
<td>0.283</td>
</tr>
<tr>
<td></td>
<td>over $65,000</td>
<td>0.06</td>
<td>0.68</td>
<td>0.09</td>
<td>0.925</td>
<td>-1.316, 1.445</td>
</tr>
<tr>
<td>Acculturation</td>
<td>Both equally</td>
<td>1.41</td>
<td>0.75</td>
<td>1.88</td>
<td>0.069</td>
<td>-0.114, 2.928</td>
</tr>
<tr>
<td></td>
<td>Mostly English</td>
<td>1.05</td>
<td>0.47</td>
<td>2.22</td>
<td>0.033</td>
<td>0.088, 2.018</td>
</tr>
<tr>
<td>CERAD delayed recall</td>
<td>Age</td>
<td>-0.12</td>
<td>0.01</td>
<td>-7.89</td>
<td>0.000</td>
<td>-0.15, -0.09</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>0.73</td>
<td>0.18</td>
<td>3.96</td>
<td>0.000</td>
<td>0.35, 1.10</td>
</tr>
<tr>
<td></td>
<td>Education</td>
<td>1.08</td>
<td>0.21</td>
<td>5.10</td>
<td>0.000</td>
<td>0.65, 1.51</td>
</tr>
<tr>
<td></td>
<td>Country of birth</td>
<td>0.01</td>
<td>0.28</td>
<td>0.04</td>
<td>0.965</td>
<td>-0.56, 0.58</td>
</tr>
<tr>
<td></td>
<td>Annual family income</td>
<td>$25,000 and $64,999</td>
<td>0.11</td>
<td>0.28</td>
<td>0.39</td>
<td>0.696</td>
</tr>
<tr>
<td></td>
<td>over $65,000</td>
<td>0.42</td>
<td>0.32</td>
<td>1.33</td>
<td>0.194</td>
<td>-0.23, 1.07</td>
</tr>
<tr>
<td>Acculturation</td>
<td>Both equally</td>
<td>0.63</td>
<td>0.29</td>
<td>2.19</td>
<td>0.036</td>
<td>0.04, 1.22</td>
</tr>
<tr>
<td></td>
<td>Mostly English</td>
<td>0.55</td>
<td>0.30</td>
<td>1.86</td>
<td>0.072</td>
<td>-0.05, 1.16</td>
</tr>
<tr>
<td>CERAD intrusion</td>
<td>Age</td>
<td>0.01</td>
<td>0.01</td>
<td>1.83</td>
<td>0.077</td>
<td>-0.00, 0.03</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>0.05</td>
<td>0.09</td>
<td>0.54</td>
<td>0.591</td>
<td>-0.13, 0.22</td>
</tr>
<tr>
<td></td>
<td>Education</td>
<td>-0.28</td>
<td>0.10</td>
<td>-2.63</td>
<td>0.013</td>
<td>-0.49, -0.06</td>
</tr>
<tr>
<td></td>
<td>Country of birth</td>
<td>0.31</td>
<td>0.10</td>
<td>2.97</td>
<td>0.006</td>
<td>0.10, 0.52</td>
</tr>
<tr>
<td></td>
<td>Annual family income</td>
<td>$25,000 and $64,999</td>
<td>-0.14</td>
<td>0.10</td>
<td>-1.38</td>
<td>0.176</td>
</tr>
</tbody>
</table>
## Discussion

We explored the relationship between language acculturation and cognition among older Hispanic and Asian individuals living in the United States. Our findings revealed that language acculturation is an important predictor of verbal and non-verbal cognitive performance. Overall, our findings add to the extant evidence linking language acculturation to cognition by showing that language use at home may contribute to maintaining cognitive health among ethnic minorities living in the United States (Coffey et al., 2005; Derose & Baker, 2000; DuBard & Gizlice, 2008; Fiscella et al., 2002; Lara et al., 2005; Solis, Marks, Garcia, & Shelton, 1990; Weinick & Krauss, 2000; Xu et al., 2017). Results from the imputed dataset were not largely different compared to the results from the original dataset.

**Digit Symbol Substitution Test**

Results supported our hypothesis that higher acculturation significantly predicted greater performance in the DSST, which assesses non-verbal skills (Wechsler, 1981). Our findings disagree with some previous studies, which did not observe associations between acculturation and non-verbal tasks (Boone et al., 2007; Razani et al., 2007). Results however, agree with a study that had adopted the Acculturation Rating Scales for Mexican Americans for measuring acculturation, which includes items related to language use (Coffey et al., 2005) and with a study showing that acculturation significantly predicted performance on the Rosen Drawing Test (Rosen, 1981), even after adjusting for covariates (Manly et al., 2004).
In this study however, acculturation accounted for only 2% of the variance in Rosen Drawing Test scores (Manly et al., 2004). Furthermore, as the DSST assesses psychomotor speed, it is widely used to investigate cognitive changes as a function of age (Hoyer, Stawski, Wasylyshyn, & Verhaeghen, 2004), and is a robust predictor of mortality and incident disability (Rosano, Newman, Katz, Hirsch, & Kuller, 2008). Clinically, our findings are interesting because they show that language use at home in ethnic minorities may contribute to psychomotor speed and memory in old age.

**Verbal tasks**

Contrary to our hypothesis, the middle acculturation group generated 1.72 fewer words on average than the reference group in the AF task. The confidence intervals in this group however, were wide indicating poor precision and low statistical power (Cumming, 2014). Analyses of the imputed dataset did not reveal any between-group differences in the AF task. Supporting our hypothesis, the higher acculturation group remembered on average one additional word than the reference group in the CERAD word learning task. Moreover, the middle acculturation group remembered on average 0.6 more words than the reference group in the CERAD delayed recall task. The effect size was small suggesting that the strength of the relationship between language acculturation and verbal fluency including its clinical relevance may be trivial. Also, acculturation did not predict performance in the intrusion and percent savings outcomes. Our findings disagree with a previous study with a smaller sample size (N = 279) showing worse performance in Hispanic compared to Non-Hispanic White individuals in the List Sorting Test which has a heavy verbal load component (Flores et al., 2017). These findings were observed irrespective of which language participants were tested in (Flores et al., 2017).
Some of our results are at variance with previous studies showing that higher levels of acculturation are linked with greater performance in verbal tasks (Boone et al., 2007; Nielsen, Vogel, & Waldemar, 2012; Razani et al., 2007). Our findings disagree with a study linking higher acculturation with greater performance in the Supermarket Fluency test, which is similar to our AF task (Nielsen et al., 2012). Also, because the authors did not report in what language the cognitive task was administered (Nielsen et al., 2012), it is unclear whether higher acculturation reflected a better command of the language of the recipient country or greater cognition (Lehtonen et al., 2018). Alternatively, less acculturated individuals might have received less exposure to the language of the recipient country and may have been less familiar with the test items (Razani et al., 2007). Importantly however, our findings are consistent with a large study (N = 503), which did not detect a link between acculturation and verbal fluency (Manly et al., 2004). Possibly, previous studies with smaller samples were statistically underpowered and committed a type I or II error (Ioannidis, 2005).

Possible mechanisms

One explanation for some of the positive findings relates to bilingualism, which is the ability to speak two languages (G. Luk & E. Bialystok, 2013). Bilingualism is associated with protection against cognitive decline in older adults (Bialystok et al., 2018), a lower risk of mild cognitive impairment (Wilson et al., 2015), and a delayed onset of dementia (Bialystok et al., 2007; Craik et al., 2010). Although a recent meta-analysis of four longitudinal prospective studies concluded that bilingualism relative to monolingualism does not reduce the risk of dementia (Mukadam et al., 2017). Of note, given the low number of studies included in the meta-analysis and the poor operationalization as well as measurement of monolingualism and bilingualism within each study among other of this review (Mukadam et al., 2017), some researchers have criticised it questioning the validity of its conclusions.
Another study found that among Japanese individuals living in the United States, exposure to the Japanese language in childhood and adulthood was associated with a lower risk of cognitive decline (Graves et al., 1999). Also, while we could not partition participants between mono- and bilinguals, 96% of participants in the reference group, 54% in the both equally group, and 41% in the mostly English group reported having been born outside of the United States. This indicates that over half of the whole sample could have been bilingual. Therefore, in our study, we cannot discount the possibility that bilingualism played a role.

**Limitations**

Our measure of language acculturation did not include language proficiency, which may be an important factor contributing to cognition (Lara et al., 2005). We also did not have data on whether participants were bilinguals or monolinguals and the degree of biculturalism, which could be considered when investigating the relationship between language and cognition (G. Luk & E. Bialystok, 2013). Also, we did not control for length of residence in the United States due to the large portion of missing data in this variable. This is a limitation because length of residence in the United States is linked to several health outcomes (Lara et al., 2005). Also, we did not have data to investigate whether language use acted as a proximal or distal factor in contributing to cognitive performance. For example, better use of English in ethnic minority groups could facilitate communication with public health officials resulting in greater use of healthcare services which may be associated with maintenance of cognitive health (Derose & Baker, 2000; DuBard & Gizlice, 2008; Fiscella et al., 2002; Solis et al., 1990; Weinick & Krauss, 2000). It is also possible however, that learning the language of the recipient country while still using the heritage language on a daily basis could help maintain cognitive health (Opdebeeck, Martyr, & Clare, 2016). From our findings however, it was not
possible to discern which of the two mechanisms, or whether the combination of the two, was responsible for the observed relationships.

**Strengths**

Unlike previous studies (Boone et al., 2007; Coffey et al., 2005; Razani et al., 2007; Touradji et al., 2001), we had access to a larger sample of participants and were able to adjust for multiple covariates thereby generating more precise estimates of our outcomes. Moreover, our measure of acculturation is clinically relevant because collecting data on language use at home is practical and time-effective and public health practitioners can easily collect these type of data at the patient- and population-level (Lara et al., 2005). Additionally, the NHANES adopts a complex stratified, multistage national probability sampling design which allows to sample individuals who are representative of the civilian non-institutionalized populations of the United States. Also, while our sample included only two ethnic minorities (Hispanic and Asian individuals), each cohort included multiple nationalities rendering our analytic sample ethnically diverse (Mexican-American or other Hispanic, Korean, Vietnamese, and Chinese individuals). Finally, unlike a previous studies (Touradji et al., 2001), participants in our study were tested in their chosen language.

**Suggestions for future research**

Future research could aim to also collect data on bilingualism, biculturalism, and language proficiency as additional predictors to explore whether these factors also contribute to cognitive performance (G. Luk & E. Bialystok, 2013) in ethnic minorities. Because language acculturation and bilingualism may be mutually inclusive (Portes & Rumbaut, 2001), their interaction could produce an additive effect on cognitive health. Future research may also benefit from employing larger epidemiological samples to generate more precise estimates.
and allow for multivariate adjustments of important covariates such as sex (Coffey et al., 2005; Nielsen et al., 2012; Razani et al., 2007). While sex difference in cognitive abilities are generally small (Hyde, 2005), women tend to have marginally better verbal skills with the greatest magnitude found for verbal fluency than men while men tend to show marginally greater spatial performance than women (Hyde, 2014; Zell, Krizan, & Teeter, 2015). Future research should also consider conducting prospective longitudinal studies to assess whether within-group levels of acculturation and between-group differences across migrants and native-born individuals account for differences in cognition (Xu et al., 2017).

**Implications**

Results from our study have some implications for clinical practice and governments. First, our findings show that choice of language use in an ethnically diverse minority group living in the United States was an important predictor of cognition. From a neuropsychological perspective, this finding is clinically relevant because performance on the DSST task has been shown to predict individuals at risk of mortality and incident disability (Rosano et al., 2008) as well as age-related cognitive decline (Hoyer et al., 2004; Salthouse, 2019). Therefore, encouraging individuals in ethnic minorities from attaining higher levels of language acculturation may be a useful strategy for protecting against age-related decline in psychomotor speed performance. Second, from a public health perspective, governments may be interested in considering preference of language use when collecting data for large-scale epidemiological surveys in ethnically diverse and immigrant populations (Lara et al., 2005). Some of our results are consistent with some evidence showing that after adjusting for acculturation, between-group differences across ethnic minorities disappeared, indicating that language acculturation can be applied to increase the accuracy of diagnosis in clinical populations (Manly et al., 1998).
**Conclusion**

To our knowledge, this is the first study exploring the relationship between language use at home and cognition using a large epidemiological and ethnically diverse sample of older cognitively intact Hispanic and Asian individuals living in the United States. Our results add to the growing body of literature by showing that language acculturation may be an important factor linked with cognition in older individuals who are part of an ethnic minority. Therefore, language acculturation may be important to consider when exploring factors linked with cognitive health in older individuals at the patient and population level.

**Acknowledgements**

This study was supported by University of Turku, Turku Brain and Mind Center, Murdoch University, University of Michigan, University of New Brunswick, and City, University of London.
References


Between-chapter (5-6) rationale

Throughout the previous chapters, we found that i) increasing PA in seniors at a higher risk of developing AD might generate more clinically relevant outcomes than in individuals with a diagnosis of AD, ii) bilingualism as opposed to monolingualism is linked with a delayed onset of AD possibly by rendering the brain areas affected by AD more resilient against neuropathology, and iii) language acculturation is linked with greater cognition in seniors from ethnic minorities living in the USA. The evidence gathered in this thesis indicates that modifying PA and a person’s language profile could be used as potential strategies for improving cognitive outcomes in individuals at a higher risk of developing AD. However, modifying just one putative risk factor may not be sufficient for improving health-related outcomes. To our knowledge, there is not experimental study that would have investigated the effects of increasing PA in combination with studying a foreign language on cognitive outcomes in individuals at a higher risk of AD such as those with SCD. The following chapter outlines the proposal for a randomized-controlled trial investigating whether combining increased PA with studying a foreign language in monolingual English seniors who report experiencing SCD can improve cognitive performance.
Chapter 6: Study protocol for a randomized controlled trial testing the efficacy of studying a foreign language with physical activity on cognition in older monolinguals with subjective cognitive decline.
This Chapter was drafted for submission to the journal *Trials*. The referencing style accords with the journal’s requirements. This manuscript accords with The Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) 2013 Checklist: Recommended Items to Address in a Clinical Trial Protocol and Related Documents. This chapter was written following a Structured Study Protocol Template (https://trialsjournal.biomedcentral.com/submission-guidelines/preparing-your-manuscript/study-protocol/structured-study-protocol-template).
**Title** Study protocol for a randomized controlled trial testing the efficacy of studying a foreign language with physical activity on cognition in older monolinguals with subjective cognitive decline.

**Names protocol contributors**
Stefano Brini\textsuperscript{a,b,c}, Behnaz Shahtahmassebi\textsuperscript{a}, Jeffrey J. Hebert\textsuperscript{a,d}, Jeremiah Peiffer\textsuperscript{a}, Timothy J. Fairchild\textsuperscript{a}

\textsuperscript{a}School of Psychology and Exercise Science, Murdoch University, Perth, Western Australia, Australia

\textsuperscript{b}Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland

\textsuperscript{c}Turku Brain and Mind Center, Turku, Finland

\textsuperscript{d}Faculty of Kinesiology, University of New Brunswick, Fredericton, Canada

**Corresponding author.**

E-mail address: stefano.brini@city.ac.uk (S. Brini).
Abstract

Background

Some observational evidence shows that bilingualism is associated with delayed symptom onset and clinical diagnosis of dementia of all types and specifically, dementia due to AD. Moreover, increased physical activity has been shown to improve cognition in older individuals at a higher risk of AD. Whether studying a foreign language (bilingualism) combined with increased physical activity can be implemented and is beneficial to cognition in older adults at a higher risk of AD has not been explored yet. This paper is the study protocol describing the methodology for the proposed randomized-controlled trial aiming to address this question.

Method

This 18-month randomized-controlled trial will investigate the effects of studying a foreign language and increasing levels of physical activity in SCD individuals. Participants will be randomly allocated to either a language-learning with physical activity condition or a physical activity-only condition. The primary outcome is cognitive performance.

Conclusion

This protocol will provide a study design for a randomized-controlled trial investigating the impact of foreign language-learning and increased PA on cognitive performance in older adults with subjective cognitive decline.

Keywords

subjective cognitive decline, Alzheimer’s disease, bilingualism, physical activity, randomized-controlled trial
Administrative information

Note: the numbers in curly brackets in this protocol refer to Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) checklist item numbers. The order of the items has been modified to group similar items (see http://www.equator-network.org/reporting-guidelines/spirit-2013-statement-defining-standard-protocol-items-for-clinical-trials/).

<table>
<thead>
<tr>
<th>Title {1}</th>
<th>Study protocol for a randomized controlled trial testing the efficacy of studying a foreign language with physical activity on cognition in older monolinguals with subjective cognitive decline.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial registration {2a and 2b}.</td>
<td>Australian New Zealand Clinical Trials Registry (ANZCTR)</td>
</tr>
<tr>
<td>Protocol version {3}</td>
<td>Version 1</td>
</tr>
<tr>
<td>Funding {4}</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Author details {5a}</td>
<td>Stefano Brini, Behnaz Shahtahmassebi, Jeffrey J. Hebert, Jeremiah Peiffer, Timothy J. Fairchild School of Psychology and Exercise Science, Murdoch University, Perth, Western Australia, Australia Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland Turku Brain and Mind Center, Turku, Finland Faculty of Kinesiology, University of New Brunswick, Fredericton, Canada</td>
</tr>
<tr>
<td>Name and contact information for the trial sponsor {5b}</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Role of sponsor {5c}</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>
Introduction

*Background and rationale (6a)*

The effects of modifying certain behavioral factors including physical and mental activity on cognition in people diagnosed with Alzheimer’s disease (AD), the most common form of dementia, have shown limited translation impact (1, 2). Individuals with AD have likely accumulated such an advanced degree of neuropathology that behavioral interventions may fail to generate clinically relevant improvements in health-related outcomes (1). Therefore, implementing interventions before widespread brain damage and the clinical symptoms of AD manifest, for example, those with SCD, might have some translational impact (3-6).

Here, SCD refers to self or informant reported departure from the normal state of cognitive performance into a chronic subjective experience of cognitive decline (4) despite normal neuropsychological performance (3). Although there is some evidence of lower performance in immediate and delayed verbal recall (7) as well as psychomotor performance (8) before the onset of SCD. Nevertheless, individuals with SCD are considered at a higher risk of AD (7) and therefore, SCD is likely an important time point in the clinical spectrum of AD (1, 3).

The National Academy of Medicine has suggested that learning a foreign language and increasing physical activity (PA) as potentially protective factors could be applied as interventions (9), as they may be particularly beneficial in the preclinical phase of AD (3, 5, 6). Moreover, there is some evidence showing that when cognitive interventions are delivered together with structured PA programs, the magnitude of the effects on cognition is greater relative to when each is delivered in isolation (10). Therefore, we argued that increasing PA in combination with studying a foreign language might generate a greater effect on cognition than delivering each intervention on its own (6, 11).
Bilingualism and foreign-language learning

Broadly defined, bilingualism refers to the ability to speak at least two languages as opposed to one (12). Bilinguals, compared to monolinguals, are diagnosed with AD about four to seven years later (13, 14) and studying a foreign language in adolescence has been linked to a lower risk of mild cognitive impairment (MCI), the prodromal phase of AD (15). These studies indicate that bilingualism may promote cognitive reserve (CR), which is the ability to maintain cognitive performance despite the presence of AD pathology (16). For these reasons, studying a foreign language may be an important strategy for maintaining cognitive function in individuals with SCD (9, 17).

To our knowledge, however, only one pilot study has investigated the feasibility and effects of studying a foreign language on cognition in elderly individuals (18). Results showed that language learning was feasible in this age group ($M=75.42$; $SD=8.93$) and that participants reported that the program was stimulating and enjoyable. However, there was no significant improvement in global cognitive function as measured via the Montreal Cognitive Assessment (MoCA). This is likely due to the small sample size ($N=14$), lack of a control group, and low sensitivity of the MoCA to detect cognitive changes in healthy individuals. One ongoing RCT is testing the effects of cognitively stimulating activities (e.g. studying Spanish) on cognition (19). In this trial, however, participants will not engage in a structured PA program and researchers will not specifically monitor language learning among participants (19). Therefore, it is necessary to replicate the methodology of these studies while addressing their limitations (18, 19).

Mechanisms of action in bilingualism

During oral communication, bilinguals demonstrate simultaneous activation of both languages in brain areas responsible for speech production (20), and as such, both languages
compete to control speech output (21). Therefore, when speaking, the bilingual needs to inhibit the context-inappropriate language and focus on the target language (21). By implication, the administration of competing languages should exercise brain areas responsible for cognitive control and promote executive function (21) with some evidence supporting this theory (22).

For example, older bilinguals had greater white matter integrity in both the left and right prefrontal cortex (23) and greater white matter in the frontal lobe (24) compared to monolinguals. Higher grey matter volume in brain areas responsible for cognitive control have also been observed in bilinguals vs. monolinguals (25). However, others were either unable to replicate previous evidence, for example, the greater grey matter volume (24), or detected the opposite trend with bilinguals showing reduced microstructural integrity in white matter despite similar cognitive performance relative to monolinguals (26). This evidence suggests that bilingualism may strengthen white matter rendering the brain more resilient against neuropathological changes associated with AD (26, 27). The neuropathological effects of AD on white matter integrity may already be present at the SCD phase (28).

Further, in participants with mild cognitive impairment and early AD, bilinguals demonstrated more impairment than monolinguals in cerebral glucose uptake in brain areas responsible for language, which is a marker of neurodegeneration (29). Participants also showed similar levels of cognitive impairment suggesting that bilingualism might have helped participants in maintaining functional cognition despite the presence of neurodegeneration associated with AD (29). These findings have also been replicated in participants occupying a more advanced stage in the clinical spectrum of AD (30). Structural neuroimaging studies have also shown that in participants with mild cognitive impairment (31) and AD (32) matched for degree of cognitive impairment, bilinguals demonstrated greater brain atrophy in brain areas typically affected by AD.
There is also experimental evidence showing that studying a foreign language can promote neuroplasticity (22). For example, after a nine-month intensive course in Modern Standard Chinese, participants in the experimental group showed changes in language brain areas; namely the left and right hemisphere as well as in the frontal lobe compared to controls (33). In another study, military interpreters taking a three-month intensive language course showed increased cortical thickness across brain areas responsible for language as well as in right hippocampal volume (34). The hippocampus is one of the first brain areas that is severely affected by AD and individuals with memory complaints or SCD relative to controls show more hippocampal atrophy (35). Also, atrophy of the hippocampus has been shown to precede SCD (36). Overall, from these findings, it appears that studying a foreign language may be linked with changes in brain areas involved in the neuropathology of AD. What remains unknown, however, is whether these putative changes in relevant brain areas following a foreign language course translate into improved cognition in individuals already experiencing SCD.

Moreover, cross-sectional studies assessing differences in relevant brain areas between monolinguals and bilinguals generally often applied non-validated measures of bilingualism (22). Also, studies testing the effects of studying a foreign language on the brain tended to have small sample sizes (range $N = 27$ to 31), did not apply random allocation of participants (33, 34), and one study did not assess changes in cognition (33). Therefore, while some cross-sectional and experimental studies appear to suggest that bilingualism and studying a foreign language might be linked with changes in brain areas typically affected by AD (22), whether studying a foreign language might improve cognition in individuals with SCD has not yet been explored.
Physical activity

Physical inactivity is one of the greatest modifiable risks factor of AD (37). While increasing PA has only modest positive effects on cognition among individuals in the clinical phase of AD (5), there is reason to believe that increasing PA may be more beneficial on cognition in individuals at the preclinical phase of AD (5, 38, 39). However, there is limited research assessing the effects of PA on cognition in individuals with SCD (5), particularly with studies combining non-pharmacological and PA interventions.

One randomized-controlled trial (RCT) found that increased PA improved cognitive performance in individuals with memory complaints (39). Here, participants were classed as memory complainers if they had answered yes to the question: “Do you have any difficulty with your memory?”. Another RCT tested the effects of a multidomain intervention of diet, mental activity, and exercise in older individuals at a higher risk of AD and found improvements in cognitive tasks assessing processing speed and memory (38). In another RCT, older participants (including some with SCD) showed improved cognition and cerebral glucose metabolism after receiving a combination of mental and physical exercise than participants in the control group or those receiving each treatment in insolation (40). Here, while participants were at higher risk by virtue of their advanced age, SCD was not specifically assessed, however (40). Although one study did assess memory complaints, the authors did not apply other important criteria for establishing SCD in their cohort of participants (39). Therefore, in this study, it remains unclear whether SCD was due to early AD or other reasons such as depression, anxiety, or another type of dementia (41).

**Objective {7}**

Our objective will be to test the effects of studying a foreign language together with increasing PA on cognitive performance in older monolingual adults with SCD. Therefore,
we hypothesize that by the end of the intervention we will observe between-group differences in cognitive performance favouring the intervention relative to the control group.

**Trial design** {8}

We will adopt a two-group parallel design, with a 12-month intervention and a six-month follow-up period. This study will be reported using the guidelines for reporting RCTs according to the Consolidated Standards of Reporting Trials (CONSORT) statement (42). The recommended content for the schedule of enrolment, interventions, and assessments is presented in Table 2 (43). The SPIRIT 2013 Checklist: Recommended Items to Address in a Clinical Trial Protocol and Related Documents is presented in Appendix A (43).

**Methods: participants, interventions, and outcomes**

**Study settings** {9}

The data will be collected at the Exercise Science laboratories at Murdoch University in Perth, Western Australia.

**Eligibility criteria** {10}

According to the United States National Institute on Aging-Alzheimer's Association framework criteria for research, SCD refers to the transitional stage from normal cognition to AD presenting with biomarker indicators of AD in the absence of objective cognitive impairment (44). The working group of the Subjective Cognitive Decline Initiative provides a list of factors to apply in order to increase the likelihood of greater specificity for AD in individuals with SCD (45). We will also follow selection criteria according to published recommendations for generating a more uniform definition of SCD (41). Details are
presented below. See Table 1 for a list of selection criteria.

Participants
Psychiatric disorders are often responsible for SCD and therefore the presence of a current or past major psychiatric disorder according to the ICD-10 or DMS-V should be considered an exclusion criterion. Age is another factor that should be considered. As such we will recruit participants within the age range (60 and 85 inclusive) in which the presence of SCD is more likely due to pre-clinical AD rather than other factors (45). To increase the probability that our target sample will include participants at a higher risk of developing AD, participants will be screened with the Cardiovascular Risk Factors, Aging and Dementia (CAIDE) application (45). The total scores (possible range: 0–15 points) will be based on factors including age, sex, education, systolic blood pressure, body-mass index, total cholesterol, and physical activity. This application has been used previously to screen for participants' baseline AD risk (38). A list of selection criteria is presented in Table 1.

Inclusion criteria
We will include individuals who report subjective decline in memory, rather than other domains of cognition, who experienced the onset of SCD five years before screening for study eligibility, are aged 60 years or above but not older than 85, have concerns associated with SCD and carry a feeling of worse performance than others of the same age group, with at least one the apolipoprotein ε4 allele and have biomarker indicators of AD. Participants answering yes to both: “Do you feel like your memory has become declined” and “Are you worried about this?”. We will recruit male and female English monolinguals. Monolingual status will be determined at the screening phase during a phone interview by asking participants whether they can speak a second language and whether they have received any
form of language instruction (e.g. face-to-face or online) in the last 30 years. Prospective participants will also need to have access to an electronic device with an internet connection because the language course will be delivered online. We will recruit individuals within the age range of 60 and 80 years inclusive. This is because this age range is more predictive of AD than other dementia aetiologies (41, 46). Participants with a CAIDE scored of six or above will be eligible to participate (38, 47). We will include participants with a Mini-Mental State Examination (MMSE) score of 25 points or above, a global Clinical Dementia Rating (CDR) score of 0 points, and a Montreal Cognitive Assessment (MoCA) total score above 20 points (19).

Exclusion criteria

Factors such as different dialects within a language, having studied a foreign language in school, or exposure to other languages in the media question the extent to which an individual may be truly monolingual. Therefore, we will exclude individuals who report having studied a foreign language within the last 30 years. Also, participants with uncontrolled hypertension or type 2 diabetes mellitus will be excluded as they may increase cognitive complaints due to neuropathy unrelated to AD (4). Participants who have currently been diagnosed with a major psychiatric condition using the International Statistical Classification of Diseases and Related Health Problems or the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Further, because participants will be asked to increase their PA levels, those who report the presence of an acute or chronic condition that may be negatively affected by increased PA or report being unable to engage in PA without help will be excluded from enrolling.
Table 1 Selection Criteria for study protocol

**Inclusion criteria**

1. Age: 65 to 85 years
2. Sex: all
3. Age at SCD onset ≥ 60 years
4. Onset of SCD within the last 5 years
5. SCD in memory performance as opposed to other cognitive domains
6. English monolinguals
7. MMSE score ≥ 25
8. CDR score of 0
9. MoCA score of above 20

**Exclusion criteria**

1. Received foreigner language instruction within the last 30 years
2. Presence of substance use, psychiatric, neurological, or medical conditions known to affect cognition
3. Presence of major psychiatric conditions according to ICD-10
4. Unable to perform PA (e.g. walking) unhindered
5. Acute or chronic conditions that may be exacerbated by physical exercise: stage one of the APSS screening tool
6. A CAIDE risk score of 5 or below
7. Inaccessibility to an electronic device (e.g. computer, tablet, etc.) with an internet connection
8. Known history of HIV

SCD: Subjective Cognitive Decline; MMSE: Mini-Mental State Examination; CDR: Clinical dementia rating scale; MoCA: Montreal Cognitive Assessment; ICD-10: International Statistical Classification of Diseases and Related Health Problems; APSS: Adult Pre-Exercise
Who will take informed consent? {26a}

A trial coordinator will be responsible for screening and recruiting as well as taking informed consent from eligible participants.

Additional consent provisions for collection and use of participant data and biological specimens {26b}

Not applicable.

Interventions

Explanation for the choice of comparators {6b}

In this trial, we compare one group receiving a PA program combined attending an online foreign language course to a group receiving a low dose PA program combined with attending an online foreign language course without receiving language instruction. The reasons for comparing these two groups are threefold: 1) moderate-high PA with an aerobic component with high frequency is necessary to generate improvements in health-related outcomes, 2) since the language course is delivered online, there is a possibility that engaging with the online environment may be mentally stimulating for some of the participants and as such, it is necessary to control for this by allowing participants to engage with the online content without studying a language, and 3), both components of the intervention group expose participants to social engagement which may contribute to an observable improvements in our outcomes and as such it is necessary to render both groups similar in the
level of social engagement participants receive in each group (48).

\textit{Intervention description [11a]}

\textbf{Language}

We will let participants chose from Italian, Spanish, and French. Participants will be able to choose which one of these three languages they prefer to study by using an online language course using a free-of-charge application (www.duolingo.com). Upon receiving an e-mail inviting participants to sign up to the website, they will be able to choose their preferred language. Participants will be instructed to work from the beginner through to the intermediate and advanced levels. Each of these three levels includes multiple language-learning-based activities. To progress forward, the user needs to successfully complete each activity. The successful completion of one activity unlocks a subsequent activity. Once all activities within one level (e.g. beginner) have been successfully completed, the user can move forward to the following level (e.g. intermediate). Participants will be able to progress at their own pace but will be encouraged to work at least three to four hours per week for 12 months. Moreover, this website allows researchers to track participants’ progress through the language course. This will provide information as to whether participants are adhering to the language program in real-time. A researcher will call participants who do not comply with the language-learning tasks and encourage them to continue studying the language.

\textbf{Physical activity}

The 12-month PA protocol will consist of moderate-high intensity aerobic and resistance training activities (5). The overall volume of PA will be at least 150 minutes of PA per week, which is the minimum amount of PA recommended by the American College of Sports Medicine. In all activities, the intensity will be assessed using the rating of perceived exertion
(RPE) scale (49), and all participants will be trained in the use of this scale during the assessment and supervised exercise sessions. This training is particularly relevant to individuals unaccustomed to moderate-intensity PA and will occur by monitoring heart-rate in conjunction with the RPE during supervised sessions, so that participants may directly compare their perceived-rating relative to the physiological measures. Participants will be (i) encouraged to maintain their current PA activities; (ii) asked to complete one session per week of up to 30 minutes of high-intensity aerobic exercise in an exercise laboratory; (iii) complete home-based resistance training exercises two times per week.

The 30 minutes of high-intensity aerobic exercise will comprise cycling on a Velotron bicycle for five minutes at 60% of maximal oxygen consumption (V\textsubscript{O\textscript{2}}\textsubscript{peak}) and then increasing to V\textsubscript{O\textscript{2}}\textsubscript{peak} to 95% across four intervals of four minutes. In between each interval, participants will be asked to decrease the exercise intensity back to 60% for two minutes. The home-based resistance exercise sessions will be supported by five exercise sessions in an exercise laboratory, which will be supervised by an accredited exercise physiologist. These sessions will be weekly for the first three weeks and repeated in week 6 and week 12 (50). In these sessions, home-based resistance (bodyweight) exercises will be demonstrated and participants will be educated on is the movements as well as regressions (how to make these easier) and progressions (how to make these harder; e.g. single limb progression or increasing resistance with therabands). The home-based resistance exercises will comprise i) three exercises for the lower body, sit-to-stand, lying leg extensions, heel raises; ii) glute bridge; iii) two upper body exercises, wall push-up, band row (using a therapy band). Participants will be asked to perform 8-12 repetitions of each exercise and to perform each exercise twice, on at least two occasions per week.

The circuit is expected to take 20 minutes, and the total target for PA is, therefore, >200 minutes of PA per week. Participant adherence will be recorded in a training diary,
which will include additional questions concerning motivation to exercise and follow-up phone calls. We set the compliance criterion as adhering to 80% of the PA program. To increase adherence to the PA intervention we will develop a written schedule tailored for each participant outlining the dates and times of each PA session in the laboratory and at home for the study duration as well as send prompts and reminders via mobile phone and follow-up with phone calls throughout the trial (50).

Active control

Previous evidence has shown that seniors learning how to use a computer showed improvements in general wellbeing and a sense of empowerment (51). As such, to render the intervention and control group as similar as possible, participants in the active control condition will be asked to sign up to www.duolingo.com and create an account. They will also be asked to engage with the website by completing exercise without engaging with the language learning component. Participants in the active control group will be asked to engage with the activities within the beginner’s level but without actually studying the language and passing the activity. For example, the participant will be able to answer questions within the activity incorrectly which will prevent the subsequent activity to be unlocked. As such, we will be able to monitor if some participants in the active control group engage in studying the language by verifying that subsequent activities remain unlocked. In case we observe that some participants in the active control group unlock subsequent activities, this would signal that the participant has engaged in language learning and will be asked to discontinue the language learning by unlocking subsequent activities. To match the control group to the intervention group, participants in the control group will be instructed to attend an organized walking group around the university campus.
Criteria for discontinuing or modifying allocated interventions \(11b\)

We expect some participants might experience adverse events during the PA program in either group. Any adverse event which may include but is not limited to physical injury that may occur in participants during the trial will be recorded and reported as an adverse event. In case a participant cannot adhere to the level of intensity or frequency of the PA program, trial the exercise scientist will modify intensity or frequency or both to a level that is manageable to the participant.

Strategies to improve adherence to interventions \(11c\)

Participants in the intervention group will be asked to progress in the online language course at their own pace so as to ensure that participates do not feel overburden and decide to dropout of the study. Similarly, missing data due to participants who need to discontinue participating in the trial because of adverse events or any other reason will be handled at the analysis stage with the appropriate statistical techniques including multiple imputations and intention-to-treat analysis with sensitivity analysis to assess the robustness (52). More details about handling missing data at the statistical level are reported in the Statistical methods section below. We will also apply several recommendations for limiting missing data during the trial including hiring researchers with previous knowledge on recruiting and monitoring participants and training them in retaining participants in the trial until completion, reducing burden to participants that may be due to completing the intervention by allowing participants to progress at their own pace, among others (52).

Relevant concomitant care permitted or prohibited during the trial \(11d\)

Participants will be permitted to continue using any medications for comorbid conditions such as Antihypertensives or any over-the-counter drugs that may have had started taking
before or during the trial. We will collect information regarding any medications that participants had begun taking before starting the trial will any will continue to do so regarding medications that participants might have started using during the trial. Participants who are already maintaining the recommended 150 of PA per week or attending PA-based classes will be encouraged to continue maintaining their habitual level of PA.

**Provisions for post-trial care {30}**

During the completion of this trial, the university’s insurance will cover all participants for indemnity due to negligent harm or researchers’ departure from conducting the trial in accordance with this protocol. The university will cover costs or compensations related to any harm due to protocol violation form the part of the researchers.

**Outcomes {12}**

*Primary outcome*

The primary outcome will be the between-group difference at post-intervention in cognition using the Alzheimer Disease Cooperative Study Preclinical Alzheimer Cognitive Composite (ADCS-PACC), which can be used to assess cognition in individuals at the asymptomatic phase of AD such as those with SCD (53). A composite score component for each participant is calculated from four cognitive tests: 1) Total Recall score from the Free and Cued Selective Reminding Test (FCSRT; 0-48 words), Delayed Recall score on the Logical Memory IIA subtest from the Wechsler Memory Scale (0-25 story units), The MMSE total score (0-30 points). Change scores from each component are divided by the standard deviation of the baseline sample of that particular component. This will generate standardized \( z \) scores, which will be summed to generate the composite scores. A change of one standard deviation in a component will equal a change of four points on the composite
score.

Secondary outcome

As secondary outcome will include changes from baseline in Alzheimer's Disease Cooperative Study—Activities Daily Living—Prevention Questionnaire (ADCS-ADL-Prevention Questionnaire) scores (54). Another secondary outcome will include an official language test at the end of the trial to assess participants’ knowledge in the chosen language after completing the course. Scores will be presented descriptively for the intervention group for the post-intervention time point at six months. This language test will be administered by a qualified language teacher in the language that the participant chose to study during the trial.

Demographic factors

We will collect demographic data on age, sex, occupation, socio-economic status, education, and marital status. These will be collected with standardized forms asking participants to report their age (continuous), sex (categorical: male or female), job type (nominal), years of formal education (continuous), and marital status (categorical: whether they are married, divorced, etc.). We will use the Adult Pre-Exercise Screening System (APSS) Screening Tool to identify participants with pre-existing conditions that may be exacerbated by increasing PA. To identify participants at a higher risk of AD, we will use the CAIDE application, which collects information on age, educational level, hypertension, hypercholesterolemia, obesity, and PA (47).

Neuropsychiatric symptoms

Depression and anxiety are highly prevalent in individuals with SCD and have been found to influence the relationship between SCD and future risk of AD (5). Therefore, we will exclude
participants with major neuropsychiatric disorders including depression and anxiety according to the International Statistical Classification of Diseases and Related Health Problems (ICD-10). We will screen participants using the Geriatric Depression Scale, which is a self-report measure of depression including 30 items consisting of yes/no questions (55). Scores range from 0 to 30 and we will exclude participants scoring 10 points or above. To screen participants for the presence of anxiety, we will administer the Geriatric Anxiety Scale with higher scores indicating higher anxiety and exclude participants scoring > 9 points (56).

**Participant timeline {13}**

A description of the timeline content is presented in Table 2.

*Assessment of eligibility*

Researchers will screen potentially eligible participants against selection criteria over the phone and explain the procedure of the experiment. Participants who meet inclusion criteria and are interested in participating will be asked to attend baseline testing. Participants meeting any of the exclusion criteria will not be eligible to participate.

*Baseline*

Baseline testing is estimated to last approximately 60 minutes. Within this timeframe, participants will be asked to read the information letter and sign the consent form. The baseline testing will involve collecting demographic data and cognitive tests. The order of administering the cognitive measures will be counterbalanced across participants to account for fatigue effects. We will also collect PA levels using Actigraph accelerometers which will be worn for two weeks. They will also be instructed to remove the Actigraph when coming
into contact with water (e.g. while showering or swimming). At the end of testing, researchers will book a suitable date for post-intervention testing.

Post-intervention testing

At the end of the intervention, participants will be asked to come back and complete the same tests that were administered during the baseline testing.
### Table 2 Participants’ timeline for study protocol

<table>
<thead>
<tr>
<th>TIMEPOINTS:</th>
<th>Study period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Enrolment</td>
</tr>
<tr>
<td>Assessment of eligibility</td>
<td>(-t_1)</td>
</tr>
<tr>
<td>Informed consent</td>
<td>X</td>
</tr>
<tr>
<td>Allocation</td>
<td></td>
</tr>
</tbody>
</table>

### INTERVENTIONS:
- PALL
- PA

### ASSESSMENTS:

#### Primary
- Demographics | X
- Memory Complaints Questionnaire | X
- Geriatric Depression Inventory | X
- Executive functioning | X | X | X
- Processing speed | X | X | X
- Memory | X | X | X

PALL: physical activity + language learning; PA: physical activity
Sample size \{14\}

We used GLIMMPSE to conduct power analysis (https://glimmpse.samplesizeshop.org/#/) for LMM. We will assume that the between-group difference, if one exists, will be relatively small. A priori power calculation with an alpha level of .01, power at 90%, to detect a small effect size gave a sample size of 216 participants. We will include an additional 20% of participants to account for possible attrition rate, which will result in an additional 54 participants. Therefore, the total sample size will be 270 (135 participants per group).

Recruitment \{15\}

The method of recruiting participants with SCD can affect the characteristics of the study sample. To increase the probability that SCD in our participants is due to pre-clinical AD, we will recruit participants from memory clinics (57) in and outside the metropolitan area of Perth, Western Australia. We will send recruitment fliers to multiple memory clinics and retirement centers in Perth and ask the clinical team to distribute the fliers to participants who appear to meet our selection criteria in the first instance.

Assignment of interventions

Sequence generation \{16a\}

To randomly allocate participants to the intervention and control group, we will use the method of simple randomization (58) as it is an appropriate method in trials with at least 60 participants (59) using an online random number generator (https://www.random.org).
**Concealment mechanism** [16b]

A researcher who will not be involved in any other aspect of the trial will generate the random sequence and prepare opaque envelopes and number them sequentially containing the participants’ allocation.

**Implementation** [16c]

These envelopes will be opened and given to participants after completion of the baseline assessment. Given the nature of the intervention, it will not be possible to blind participants to the interventions.

**Data collection, management, and analysis**

**Plans for assessment and collection of outcomes** [18a]

Research fellows qualified in delivering the PA activities and administering cognitive tests will be responsible for implementing the intervention and collecting data. The researchers will collect data at baseline, post-intervention, and follow-up.

**Plans to promote participant retention and complete follow-up** [18b]

The plan for increase adherence in the PA program is described under item 1.1 Intervention description {11a}. The study coordinator will also monitor participants’ language learning progress in real time online and will identify those not adhering to it. The coordinator will help participants not adhering to the language program find appropriate solutions in dealing with potential barriers. The study coordinator will also send reminders and prompts to increase participants’ adherence to the PA and language programs.
Data management \{19\}

The data will be saved and stored in a secure database whose access will be restricted with a password to the researchers involved in the trial. Two researchers not involved in data analyses will conduct checks for quality and range checks for data values.

Statistical methods

Statistical methods for primary and secondary outcomes \{20a\}

To minimize bias, one investigator who will be blind to the conditions, will receive the dataset and conduct the statistical analyses. A second investigator also blinded to the conditions, will re-analyze the dataset to confirm the results. The data will be analyzed at an alpha level of 0.05.

The statistical analysis will involve a Linear Mixed Modelling (LMM), which will also include an intention-to-treat analysis (ITT) on each outcome variable. The LMM with an AR(1)/AD(1) covariance matrix will be used to explore the effects of studying a foreigner language coupled with increasing PA on cognitive performance. Time will be one factor in the analysis and will be entered as a random factor. We will have three time-points: baseline (T1), post-intervention (T2: 12 months since baseline), and follow-up (T3: 18 months since baseline). We will examine the main effect of treatment group vs control group, the main effect of time (from baseline across all follow-ups), and the interaction between group and time to examine if the effects of the treatment are dependent on time factors. Post hoc comparisons will be conducted using Bonferroni’s correction. Time will be nested within participants and therefore we will restructure the dataset from wide to long format. Time will be entered as a fixed effect with participants’ identification number as random effects in each LMM. Means with standard deviations and 95% confidence intervals for demographic
variables and outcome measures will be calculated and presented in a table. Means with standard deviations and 95% confidence intervals for demographic variables and outcome measures will be calculated and presented in a table.

To assess changes from baseline, between-group differences at post-intervention, and interaction effects, we will calculate reliable change indices in the primary outcome (60), as this is the preferred method when accounting for practice effects (61) in assessing clinically relevant changes in trials that include participants with SCD (6).

Methods in analysis to handle protocol non-adherence and any statistical methods to handle missing data \( \{20c\} \)

Before analyzing the dataset, the investigators will conduct the Missing Completely at Random (MCAR) test to determine whether the data are MCAR. If the data are MCAR, to handle missing covariate data we will conduct multiple imputations (MI) including all predictors and time points to fill the missing data. Following this, we will conduct MI for all relevant outcomes and covariates. The imputation model will include all variables from each time point, which will be imputed individually. We will create two datasets: pre-imputed and imputed \( (m = 10) \); the 10 sets of results will be combined with Rubin’s rule.

Plans to give access to the full protocol, participant level-data and statistical code \( \{31c\} \)

The dataset will be made public within one year from the completion of the trial. The SPSS syntaxes will be made publicly available along with participant-level data.
Discussion

Despite several decades of research, there is still no treatment or cure for AD (62). Identifying modifiable risk factors that can alter AD disease progression and ultimately the risk of developing it, has become a global public health priority (62). Individuals with AD have likely accumulated widespread neuropathology limiting the potential beneficial effects of non-pharmacological interventions on cognition (63). As such, targeting individuals with SCD before the accumulation of advanced neuropathology may generate greater translational impact (3, 5, 7, 41, 44, 45, 63).

While there is currently no universally accepted measure of SCD, recruiting participants with certain key demographic features can increase specificity for AD (4, 6). Despite the need for recruiting individuals with SCD, previous RCTs testing the impact of non-pharmacological intervention (e.g. PA programs or studying a foreign language) in individuals described as having SCD, did not apply strict criteria for assessing SCD (6, 39). In line with one ongoing trial (19), we propose the application of specific selection criteria (Table 1) in the recruitment of participants with SCD (41, 45). Furthermore, there is evidence indicating that only modifying PA levels may not be sufficient in preventing AD (11) and while non-pharmacological intervention can improve cognition in individuals with SCD ($d = 0.22$), the inclusion of a cognitive component (e.g. cognitive rehabilitation and stimulation) can generate larger improvements ($d = 0.37$) on cognition (6). Therefore, combining a PA program with studying a foreign language (17), might be a cost-effective and useful strategy to improve cognition in individuals with SCD.

The rationale for this study is based on previous evidence showing that bilingualism is associated with a later onset of AD (14) and that studying a foreign language early in life is linked to a lower risk of mild cognitive impairment (15). Also, that increasing levels of PA are associated with a lower risk of AD (5) and that it has been shown to provide modest
improvements on cognition among individuals in the preclinical phase of AD (39). To our knowledge, no RCT has yet been conducted testing the effects and feasibility of studying a foreign language and increasing PA on cognition in older adults with SCD. This study will also provide information regarding whether the combination of studying a foreign language and increasing PA is comparable or greater effects on the currently documented effects sizes of non-pharmacological interventions (6).

To maximize the impact of the intervention on cognition, the exercise program needs to include several components. For example, while there is some evidence suggesting that both aerobic and anaerobic exercise in isolation can improve cognition, combining aerobic and anaerobic exercise for at least 150 minutes per week appears to generate greater effects on cognition in individuals at the preclinical phase of AD (5, 64). Moreover, multimodal (e.g. running, strength training, etc.) exercise program including an aerobic component appears to be more beneficial than exercise program with only one component such as running (5, 64). Previous studies testing the impact of PA on cognition in individuals with preclinical AD did not consider these details in the delivery of the PA program (39, 40). For this RCT, we purposely designed the PA program while addressing the methodological weaknesses in previous trials.

In conclusion, we will conduct an RCT investigating the impact of increasing PA and studying a foreign language on cognition in older individuals with SCD. The choice of our intervention was based on previous evidence suggesting that increasing PA and bilingualism appear to be associated with a later onset and a lower risk of AD.
References


Chapter 7 General Discussion
Due to the lack of an effective treatment or cure for AD, identifying modifiable risk factors to reduce its future incidence has become a global public health priority (Brookmeyer, Abdalla, Kawas, & Corrada, 2018; Livingston et al., 2017; Nichols et al., 2019). While the link between several modifiable risk factors including poor diet, obesity, smoking, and sedentary behavior with AD has already been established (Baumgart et al., 2015; Lee et al., 2010), the role of risk factors such as PA and monolingualism may play in preventing AD remains equivocal (Gold, 2015; M. Kivimäki et al., 2019; National Academies of Sciences, Division, Policy, & Impairment, 2017). For example, experimental studies testing the impact of increasing PA on cognition in individuals with AD have shown limited improvements in cognitive performance questioning the clinical utility of PA interventions in clinical phases of AD (Sperling, Jack, & Aisen, 2011; Öhman, Savikko, Strandberg, & Pitkälä, 2014). Also, while there is evidence suggesting that bilingualism may delay AD diagnosis, longitudinal prospective studies do not show a risk reduction in dementia among bilinguals relative to monolinguals (Gold, 2015; Mukadam, Sommerlad, & Livingston, 2017). There is also limited evidence testing the impact of studying a foreign language on cognition in older individuals at risk of AD (Ware et al., 2017).

Therefore, the purpose of this thesis was to i) review the available evidence exploring the link between PA and dementia across the clinical spectrum by conducting a narrative review of the literature (Chapter 2), ii) conduct a systematic review of the literature exploring the link between bilingualism and dementia onset as well as risk (Chapter 3), iii) conduct a systematic review of literature exploring the underlying link between bilingualism and the brain (Chapter 4), and iv) conduct a cross-sectional study epidemiological study investigating the link between language acculturation and cognition in older individuals from ethnic minorities.
While the link between physical inactivity and poor cardiometabolic outcomes is well documented, the evidence linking PA and dementia is less clear (M. Kivimäki et al., 2019; Wahid et al., 2016). Studies testing the impact of increasing PA on cognition and AD symptoms have shown limited translational impact (Öhman et al., 2014). It has been argued that increasing PA at the clinical phase of AD is too late in the natural history of the disease and that targeting the preclinical phase might be more clinically meaningful (Sperling et al., 2011; Öhman et al., 2014). The brain damage caused by AD in the clinical phase is too advanced to reverse or slow down and therefore the application of behavioral interventions to improve clinical symptoms is less likely to carry translational impact. By contrast, modifying risk factors in the preclinical phase where brain damage is still limited is likely to generate more clinically meaningful outcomes. Although there are several existing reviews looking at the effects of PA interventions on cognition in individuals with MCI and AD on cognition (Smart et al., 2017; Öhman et al., 2014), to our knowledge there is no review synthesizing the available evidence focusing on the impact of PA interventions across the clinical spectrum of AD.

In Chapter Two, this thesis found that high-moderate PA across the lifespan was linked with a lower risk of developing AD but increasing PA in individuals who had already reached a diagnosis of AD did not improve cognition or symptoms. This is in line with previous evidence showing that increasing PA earlier in the natural history of AD may be more beneficial than at the clinical phase (Sperling et al., 2011; Öhman et al., 2014). However, multimodal PA programs that included different types of exercises with at least one aerobic component delivered at a moderate-high intensity with high frequency (at least 150 minutes.
per week) across a long period of time (at least three to six months) was more beneficial than PA program only targeting anaerobic or aerobic exercise in isolation of low intensity and frequency across a short period of time. High-intensity PA was found to increase the production of growth factors (e.g. brain-derived neurotrophic factors) which may be implicated in neuroprotective effects in brain areas severely affected by AD including the hippocampus (Wang & Holsinger, 2018). However, the evidence indicated that increasing PA had a limited clinical impact on cognitive outcomes in individuals with AD. It was found that increasing PA in the earlier phases of AD such as in those with MCI or SCD might be more useful than targeting the clinical phase (Chapter 2). Previous studies show that non-pharmacological interventions can improve cognition in individuals with SCD, particularly multimodal interventions that include at least on cognitively stimulating activity (Rabin, Smart, & Amariglio, 2017). Recent evidence suggests that modifying only PA is likely not sufficient in preventing AD (M. Kivimäki et al., 2019). As such, there is an urgent need to test the impact of multicomponent interventions on cognition in individuals with SCD (Mika Kivimäki et al., 2019; Smart et al., 2017).

Modifying a risk factor such as PA in combination with promoting a cognitive stimulating behavior may be more effective in improving cognition in older individuals with SCD rather than targeting one risk factor at a time (Smart et al., 2017). There is some evidence suggesting that bilingualism, as opposed to monolingualism, may delay the onset of AD (Bialystok, Abutalebi, Bak, Burke, & Kroll, 2016; Gold, 2015) and that studying foreign language may be a useful strategy for preventing AD (Antoniou, Gunasekera, & Wong, 2013; Bialystok et al., 2016; National Academies of Sciences et al., 2017). However, one recent systematic review (Mukadam et al., 2017) concluded that bilingualism is not protective against AD and that confounding factors such as education and immigration status are
responsible for the observed delays in AD diagnosis among bilinguals that have been documented in previous studies (Bialystok, Craik, Binns, Ossher, & Freedman, 2014; Bialystok, Craik, & Freedman, 2007). However, this systematic review only meta-analyzed longitudinal prospective studies without meta-analyzing cross-sectional studies (Grundy & Anderson, 2017) and does not include recently published longitudinal prospective studies. Moreover, there is currently no systematic review exploring the underlying brain mechanisms thought to be responsible for the observed delays in AD diagnosis among bilinguals relative to monolinguals. Therefore, the equivocal findings in the literature on bilingualism and AD prompted the completion of a systematic review to determine whether bilingualism is linked with a delayed onset of AD (Chapter 3) and a second systematic review exploring the underlying brain processes linking bilingualism with a delay in AD diagnosis (Chapter 4).

Contrary to the conclusion of a previous systematic review (Mukadam et al., 2017), meta-analytic results from Chapter Three showed that bilingualism, as opposed to monolingualism, may delay the age in which the first symptoms of AD appear and the age of diagnosis. Although immigration status and education levels might have been responsible for the observed delays in AD diagnosis in bilinguals relative to monolinguals (Fuller-Thomson & Kuh, 2014; Mukadam et al., 2017), the majority of studies included in the meta-analyses had adjusted for education levels (Chapter three). Also, a subgroup meta-analysis indicated that immigration status did not appear to contribute to the observed relationship between bilingualism and AD. There was no evidence to suggest that bilingualism relative to monolingualism is linked with a lower risk of developing dementia. It is possible that while no difference in the incidence of dementia between bilinguals and monolinguals exists, symptoms will manifest at a later age in bilingualism relative to monolinguals. There was also no difference in the degree of cognitive impairment between mono- and bilinguals at the
time of dementia diagnosis despite bilinguals being older than monolinguals. This suggests that bilinguals were able to maintain a functional level of cognition for more years before reaching a similar level of disease severity to warrant a dementia diagnosis as monolinguals.

The evidence showing that bilingualism is linked with a later onset of AD symptoms and diagnosis (Chapter 3) was further supported by brain mapping showing that cognitively intact older bilinguals demonstrated structural differences in frontostriatal and frontoparietal circuits relative to their monolingual counterparts (Chapter 4). Bilinguals exhibited enhanced white matter integrity, for example, in the corpus callosum as well as stronger anterior/posterior functional connectivity and greater grey matter density than monolinguals, particularly in the anterior cingulate cortex and in the left inferior parietal lobule (Chapter 4). Each of these brain regions is severely affected by dementia and AD (Daianu et al., 2013; Palop, Chin, & Mucke, 2006). Furthermore, older bilinguals with AD demonstrated greater atrophy in the radial width of the temporal horn and the temporal horn ratio, each of which is severely affected in AD, despite bilinguals having similar cognitive performance and levels of education as monolinguals. Bilinguals with AD also exhibited more severe cerebral hypometabolism than their monolingual counterparts despite the bilinguals being older. Overall, the findings suggest that bilingualism may strengthen neural connectivity in frontostriatal and frontoparietal circuits (Gold, 2015) which are severely affected in dementia and AD (Daianu et al., 2013; Palop et al., 2006). The observed differences between older mono- and bilinguals in frontostriatal and frontoparietal circuits have already been observed in younger participants (Li, Legault, & Litcofsky, 2014).

It is possible that bilingualism could be used to maintain healthy neural circuits and forestall the neuropathological processes of AD (Canter, Penney, & Tsai, 2016; Gold, 2015).
Results from risk of bias, however, revealed that bilingualism was generally poorly operationalized and measured (Chapters 3 and 4). For example, most studies simply asked participants whether they could speak a second language without assessing how frequently and how well they could speak that second language or at what age they had acquired it. Studies generally did not assess how many languages bilinguals could speak but included those who could speak two or more languages in the same analytic cohort. Participants who reported being unable to speak a second language were categorized as monolinguals. However, it is conceivable that some of these participants were able to perhaps understand a second language (receptive bilinguals) rendering effectively non-monolinguals. These participants, however, were included as monolinguals in all studies in Chapter Three. Assessing bilingualism on a continuum would take into account the large variability present within this construct including differing levels of proficiency, frequency of second language use, and age of acquisition (Luk & Bialystok, 2013).

Findings from Chapter five support results from Chapter three and four showing that language acculturation is linked with cognition in older individuals from ethnic minority groups living in the United States of America. In this chapter, the evidence indicated that higher levels of language acculturation were linked with higher psychomotor speed and verbal fluency (Chapter 5). Low psychomotor speed (Amieva, Meillon, Proust-Lima, & Dartigues, 2019) and verbal fluency (Szlloczki, Hoffmann, Vincze, Kalman, & Pakaski, 2015) in old age are associated with a greater risk of developing AD. Decline in psychomotor speed can lead to widespread impairment in other cognitive domains because changes in processing speed limit the ability to execute higher-order thought processes (Salthouse, 1996). Since psychomotor speed declines as a function of age (Salthouse, 1996), can increase
the risk of AD (Amieva et al., 2019), and maybe present more than 15 years before AD is diagnosed (Amieva et al., 2014), preventing its decline may be a critical factor in promoting cognitive health in old age. Notably, moderate-high intensity aerobic PA can result in improved psychomotor speed with concomitant greater activation of frontostriatal and frontoparietal circuits (Rosano et al., 2010), which are the same brain circuits also found to be structurally more developed in bilinguals relative to monolinguals in Chapter four.

**Limitations of this thesis**

Most studies in the field of bilingualism-based research have tended to dichotomously categorize participants between monolinguals and bilinguals. Authors now recognize that bilingualism includes several dimensions such as how well and how frequently a bilingual speaks the second or third language, age of second language acquisition. In Chapters three and four, it was not possible to establish whether any of these dimensions played a role in the observed delays in dementia as well as differences in brain structure and function between mono- and bilinguals. For example, while most studies applied a definition suggesting that bilinguals had used the second language for most of their lives, it was not possible to discern whether there is a critical age at which learning a second language might be relevant in protective against dementia. The issue of measuring participants’ language profile was also evident in Chapter five. Here, participants were simply asked what languages they spoke at home and as such, it was not possible to establish whether the observed relationship between language acculturation and greater cognitive performance in the analytic sample was due to bilingualism promoting cognitive health has the findings from Chapter three and four would indicate, or greater accessibility to the healthcare system, which is also a likely possibility in this ethnic minority (Derose & Baker, 2000; Fiscella, Franks, Doescher, & Saver, 2002).
Suggestions for future research

In this thesis, we identified several sources of bias and uncertainty in the field of exercise science- and bilingualism-dementia based research. Generally, there were common limitations across both fields of research that challenged the interpretation of our findings. For example, limitations including small sample sizes, lack of operationalization and poor measurement of the exposure of the independent variable, and recruitment of participants regardless of dementia type were commonly found across both fields of research. Therefore, studies should consider improving upon the abovementioned limitations. Moreover, PA-dementia based research should consider monitoring participants’ adherence (van der Wardt et al., 2017), deliver high-intensity PA with an aerobic component for at least six months, and continue monitor participants for at least another six months post-intervention. In bilingualism-dementia based research, studies might benefit from treating participants’ language skills as a continuous variable while measuring dimensions such as proficiency, frequency of second language use, and age of acquisition, rather than as a dichotomous variable.

By extension, the purpose of Chapter 6 was to propose a trial that addresses the limitations identified in the studies we reviewed in the preceding chapters. Specifically, we propose an RCT investigating the effects of increasing PA levels and studying a foreign language in English-speaking monolingual individuals with SCD while addressing the limitations of previous studies (Chapter: 2-5). Unlike previous studies, in this RCT we will apply recently published guidelines for recruiting individuals with SCD, sufficiently power our study based on previous evidence including similar populations while using similar cognitive measures, exclude participants who are not truly monolinguals, include a post-intervention follow-up period, and design the PA intervention to include moderate-high intensity, aerobic exercise as
well as apply tailor the PA intervention to each participant and provide follow-up phone calls to remind and motivate participants to adherence to PA (van der Wardt et al., 2017).

**Clinical significance**

Given the future expected increase in the number of people with dementia and the heavy economic burden that societies are expected to face, our findings carry important clinical implications. Overall our findings carry important clinical relevance at the individual and population level. We found that targeting PA in individuals with SCD or who are otherwise at a higher risk of developing dementia or AD may be more clinically useful than targeting PA in the clinical phase of dementia (Chapter 2). Findings from Chapters three and four indicate that speaking two or more languages may strengthen certain brain areas that are severely affected by AD and that it can delay the onset of AD symptoms and diagnosis by 4.2 years. These are clinically relevant findings because delaying the onset of AD by approximately five years, can reduce the prevalence by 57% and halve its associated economic costs (Sperling et al., 2011). We also found that language acculturation may be associated with greater psychomotor speed and verbal fluency in older Hispanic and Asian individuals living in the United States of America (Chapter 5). This is also a clinically relevant finding because psychomotor speed is an important predictor of mortality and incident disability. Thus, findings from this Ph.D. indicate that maintaining habitual PA, particularly increasing PA before the development of dementia or AD and that bilingualism as well as language acculturation likely to contribute to cognitive health in old age and may modify the brain and render it more robust against neuropathology associated with dementia.

**Conclusion**
Our findings show that high-intensity PA of a high frequency for longer than six months including an aerobic component is associated with a lower risk of dementia and improvements in cognition in individuals with dementia. However, the literature carries several notable limitations including the poor measurement of PA, the inclusion of participants with differing dementia aetiology into the same analytic sample, and small sample sizes. While a large body of evidence supports a strong link between PA levels and dementia risk, these limitations question what the true strength of this relationship may be and whether increasing PA levels in the preclinical phase of AD may carry any translational impact in individuals with AD.

Furthermore, experimental studies showed that increasing PA levels improved global cognition in several studies, but these changes tended to be small in magnitude questioning their clinical relevance. However, because most studies had tested PA intervention individuals at the advanced stage of AD with only a small number of studies recruiting participants at the preclinical phase, it is problematic to discern whether increasing PA levels before the clinical onset of AD may generate clinically relevant improvements in cognition. Therefore, additional research in individuals with subjective cognitive decline is urgently needed.

It was also found that bilingualism is associated with a later onset of dementia and AD but not with a lower risk of developing dementia. While our findings are consistent with a previous meta-analysis showing no risk reduction in dementia among bilinguals relative to monolinguals (Mukadam et al., 2017), the meta-analysis included only five studies ($k = 5$) and as such, it might have missed a true effect by virtue of being statistically underpowered. It may be premature to conclude that bilingualism does not reduce the risk of
dementia. Moreover, unlike the previous systematic review (Mukadam et al., 2017), we meta-analyzed cross-sectional studies and found that bilinguals were older than monolinguals at dementia and AD diagnosis. Results from brain mapping (Chapter 4) suggested that bilingualism may help maintain healthy neural circuits and slow the neuropathological processes of AD (Canter et al., 2016; Gold, 2015). This, in turn, could translate into a delay in AD symptom onset and diagnosis.

Finally, the evidence we gathered within this thesis enabled us to develop a study protocol for a prospered RCT testing the effects of increasing PA levels and studying a foreign language among older monolinguals with subjective cognitive decline. Drawing from the evidence gathered herein, we hypothesize that administering a PA intervention of high intensity and frequency for at least six months with an aerobic component coupled with studying a foreign language before participants reach the clinical phase of AD might generate clinically relevant improvements in cognitive performance.
References


